



Seminar on Research Data Management – Prepare a  
**Data Management Plan** (DMP)



**HKU  
Med**

LKS Faculty of Medicine  
School of Biomedical Sciences  
香港大學生物醫學學院



The University of Hong Kong  
**Libraries**

# Today's agenda:



- What is a Data Management Plan (DMP)
- Prepare a DMP with DMPTool
- Major components of a DMP

$$\frac{dN}{dt} = \frac{1}{qV_{act}} - q_0(N-N_0)(1-\epsilon S)S + \frac{N_e}{T_n} - \frac{N}{T_p}$$

$$\frac{dS}{dt} = T_0 q_0(N-N_0)(1-\epsilon S)S + \frac{p_0 N}{T_n} - \frac{S}{T_p}$$

$$\frac{S}{P} = \frac{T_p p_0}{T} = \Theta$$

$$S \leq \frac{1}{\epsilon}$$

$$N = N_0$$

$$P_f = (m$$

1.

What is a

**Data Management Plan (DMP)**

# Data Management Plan (DMP)

- A document in which you describe
- **what data** you will collect during your research project, **Format, types, volume, ownership...**
- **how** you are going to store and manage the data during the project, **File organization, metadata, access...**
- And **what will happen** to the data after the project is finished. **Preservation and sharing...**



Source: [JISC](#)



# Example DMP

## DATA MANAGEMENT PLAN

The project will collect and analyze the following data:

- Conductivity and temperature from glider surveys.
- Horizontal currents from shipboard ADCP and the HDSS Doppler Sonars on the R.V. Revelle.
- LADCP/ CTD profiles from the R.V.Revelle.
- Moored ADCPs.
- CTD-u,v profiles from the McLane profilers.
- CTD profiles from the SIO Fast-CTD.
- Fine and microscale temperature from CHIPODs and moored thermistor chains.

### Quick-Response data management

The T-TIDE PIs have experience with this mix of data types from previous collaborative efforts, such as the ONR IWISE Experiment, 2010-11, in the S. China Sea. To guide both modeling and the Process experiment planning, quick-look Scout data will be centralized on a server at APL, UW.

Scout Quick-look data responsibilities include:

J.Klymak	LADCP-CTD analysis.
S. Johnston	SIO glider analysis
L.Rainville	Co-operative CSIRO glider Tidal analysis
H. Simmons, J.Klymak	Ongoing model output predictions
R.Pinkel, J. Klymak	F-C TD site studies

The centralized data access will be maintained for the Process Experiment, with the McLane and thermistor chain data provided by the relevant PIs.

### Long Term data Archiving

Aside from the LADCP-shipboard CTD profiles, there are currently no established standards for archiving or data from many of the fine-scale sensors used in T-Tide. Archiving standards for glider data are evolving. This is a concern of the Climate Process Team on Ocean Mixing, of which many T-Tide PIs are members. We propose to work with the CPT to evolve formats for data and metadata suitable for archiving both sensor and (critically) model output from the experiment.

All field data collected under this program will be made available as per NSF guidelines within 2 years of collection via published manuscripts, publicly available final reports to NSF, and data archiving with NODC.

Data will be shared in matlab MAT file format and/or as netCDF files. Ultimate archival formats will be determined in consultation with NODC and with the CPT. Adequate archiving is anticipated to be an expensive, time-consuming task. All PIs have included funds for this effort in their budgets.

The primary T-TIDE models are all public domain. Published peer-reviewed manuscripts will document the simulations and forcing sufficiently. Recognizing that archiving high-resolution simulations at tidally resolving intervals can result in gigabytes-to-terabytes of data, every effort will be made by modeling PIs to archive model output and provide data and/or code to interested parties upon request. Model products and output will be available at the end of the grant period.

# To prepare a DMP

- Precise
- Realistic and workable
- Address major issues
- Revisit and revise

$$\left. \begin{aligned}
 \frac{dN}{dt} &= q_{\text{act}} - q_0(N-N_0)(1-\epsilon S)S + \frac{N_e}{T_n} - \frac{N}{T_p} \\
 \frac{dS}{dt} &= T_0 q_0(N-N_0)(1-\epsilon S)S + \frac{p_0 N}{T_n} - \frac{S}{T_p} \\
 \frac{S}{P} &= \frac{T_p \lambda_0}{T} = 0
 \end{aligned} \right\} \begin{aligned}
 N &= N_0 \\
 P &= (m)
 \end{aligned}$$

2.

Who has to submit a **DMP**



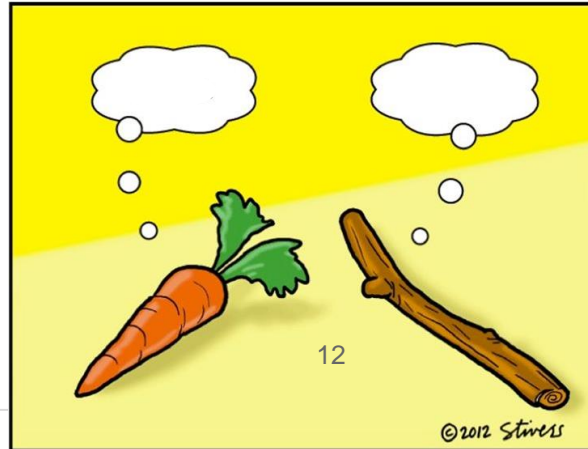
# Why research data management?

## Benefits

- Keep research safe and secure
- Increase research efficiency
- Improve research integrity
- Make research outputs more visible
- Enable collaboration

## Requirements

- Compliance with policies: **HKU** & funders
- Ensure data is accessible and shareable: journals requirement
- Demonstrate responsible practice



Source: [JISC](#)



# Research Services

Support and information for HKU researchers

[▶ Sitemap](#) [▶ Contact Us](#)

Quick Links

Search this site



[About Us](#)

[Research Funding](#)

[IPR & Contracts](#)

[Research Integrity](#)

[Responsible Conduct of Research](#)

[Research Data and Records Management](#)

[Research Ethics Compliance](#)

[Safety Approval & Insurance](#)

[Honours & Awards](#)

[RAE & Resources](#)

[Home](#) > [Research Integrity](#) > [Research Data and Records Management](#)

## Research Data and Records Management

The management of research data and records refers to ways in which recorded information (in whatever form or medium) from research is organised, stored, maintained and accessed both during the lifespan of the research and in the long term. Effective research data and records management supports both high quality research and academic integrity.

HKU recognises the importance of good practice in research data and records management and seeks to promote the highest standards. The University's *Policy on the Management of Research Data and Records* was approved by the Senate at its meeting on May 5, 2015, along with the establishment of a [Task Force on Management of Research Data and Records](#) to oversee the planning of the implementation of the Policy.

### [Policy on the Management of Research Data and Records](#)

1. The University of Hong Kong seeks to promote the highest standards in the management of research data and records (1) as fundamental to both high quality research and academic integrity, and acknowledges its obligations under research funders' data-related policy statements and codes of practice, where available (2), to ensure that sound systems are in place to promote best practice, including through clear policy, guidance, supervision, training and support.

2. The University recognises that accurate and retrievable research data are an essential component of any research



## Research Services

Support and information for HKU researchers

[▶ Sitemap](#) [▶ Contact Us](#)

Quick Links

Search this site



[Home](#) > [Research Integrity](#) > [Research Data and Records Management](#)

### Research Data and Records Management

The management of research data and records refers to ways in which recorded information (in whatever form or

[About Us](#)

[Research F](#)

[IPR & Cont](#)

[Research I](#)

[Respons](#)

[Research](#)

[Research](#)

[Records](#)

[Research](#)

[Complian](#)

[Safety A](#)

[Insuranc](#)

[Honours &](#)

[RAE & Resc](#)

To meet the need to

- **Replicate claimed research results** when called upon to do so
- Ensure **ethical** data collection, storage, and if chosen, re-sharing of data

<http://www.rss.hku.hk/integrity/research-data-records-management>



## Research Services

Support and information for HKU researchers

▶ Sitemap ▶ Contact Us

Quick Links

Search this site



About Us

Research Funding

Home > Research Integrity > Research Data and Records Management

### Research Data and Records Management

The management of research data and records refers to ways in which recorded information (in whatever form or medium) from research is organised, stored, maintained and accessed both during the lifespan of the research and in

5. Research data and records should be retained for as long as they are of continuing value to the researcher and the wider research community, and as long as specified by research funder, patent law, legislative and other regulatory requirements. **The minimum retention period for research data and records is three years after publication or public release of the work of the research.** In many instances, researchers will resolve to retain research data and records for a longer period than the minimum requirement.

RAE & Resources

records (1) as fundamental to both high quality research and academic integrity, and acknowledges its obligations under research funders' data-related policy statements and codes of practice, where available (2), to ensure that sound systems are in place to promote best practice, including through clear policy, guidance, supervision, training and support.

# Guidelines and Procedures (RPG)

Beginning with the September 2017 intake, all **HKU research postgraduate (RPG) students** have responsibility for

1. using a **data management plan (DMP)**, where applicable, to describe the use of data in preparation for, or in the generation of their theses, and
2. depositing, where applicable, a **dataset** in the HKU Scholars Hub

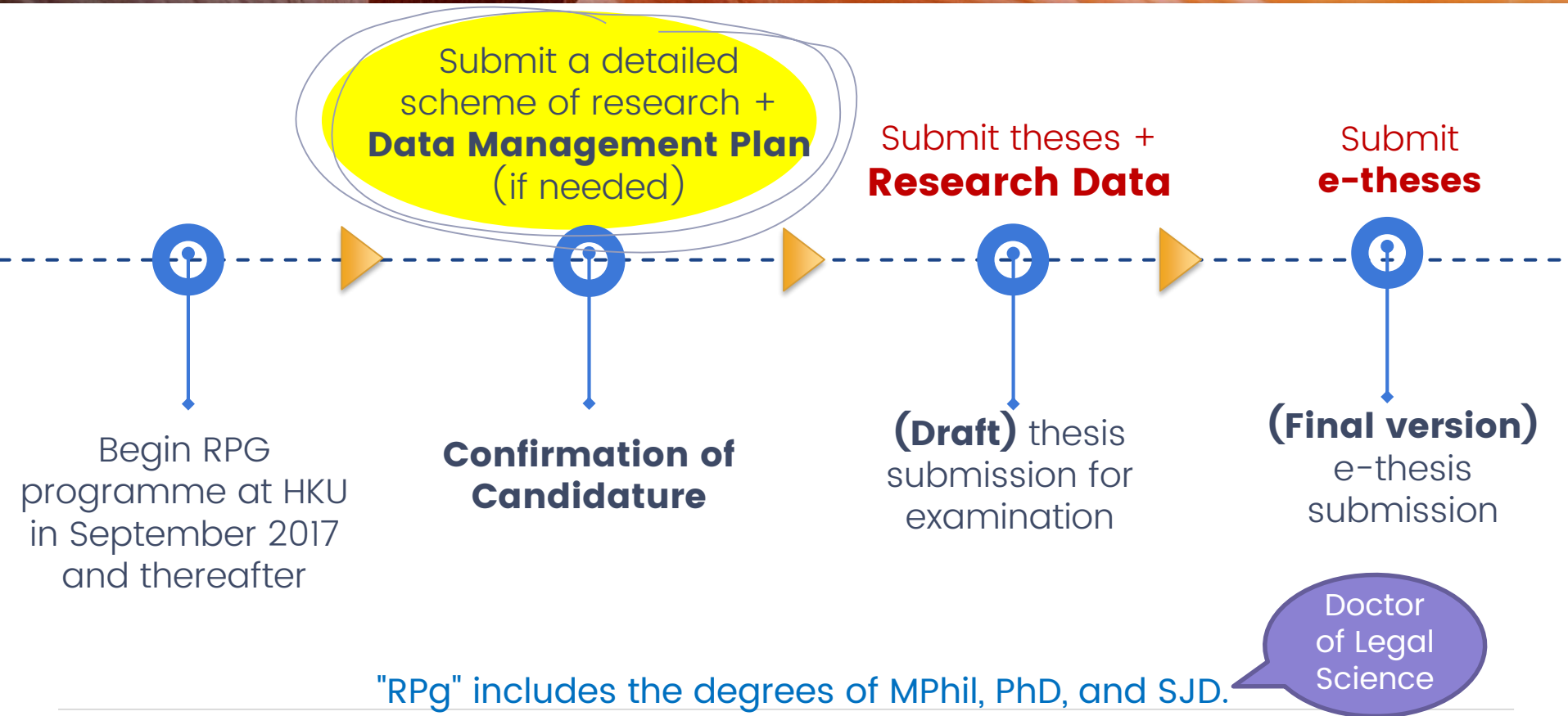
"RPg" includes the degrees of MPhil, PhD, and SJD.



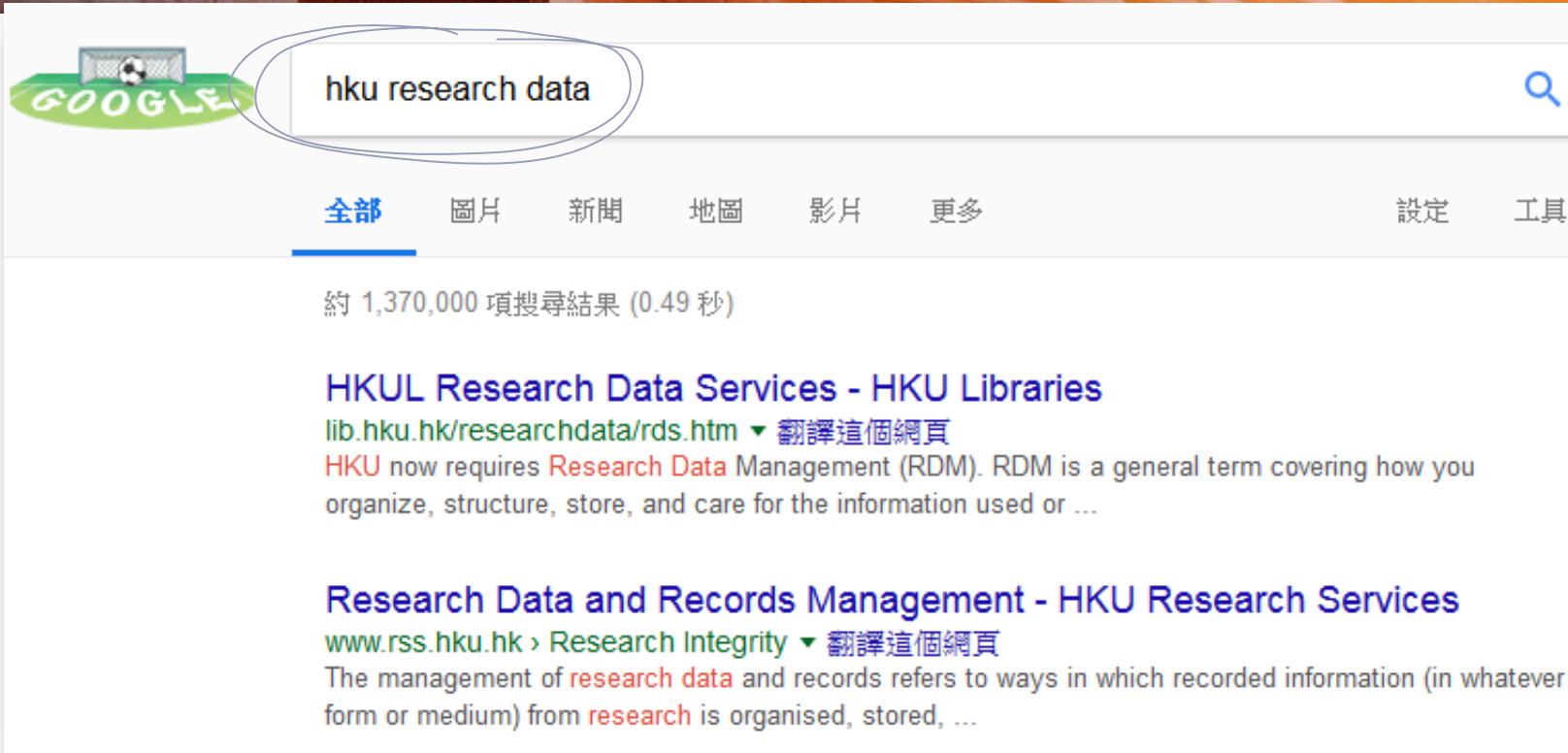
Doctor  
of Legal  
Science



# Timeline (RPG)



# HKUL Research Data Services





## RESEARCH DATA SERVICES

THE UNIVERSITY OF HONG KONG  
LIBRARIES



RDM FOR RESEARCH  
POSTGRADUATE (RPG) STUDENTS

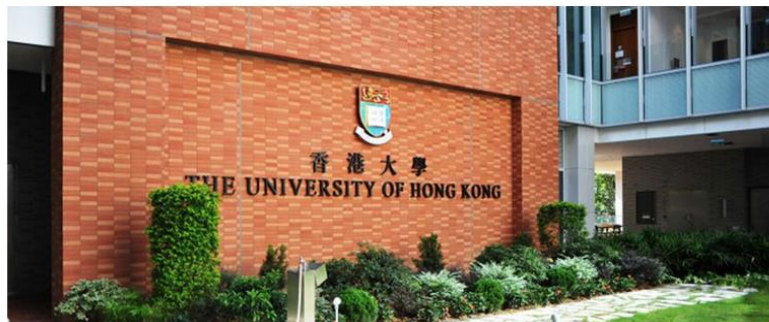


RDM FOR RPG SUPERVISORS



## WHAT IS RESEARCH DATA MANAGEMENT?

RDM




HKU now requires Research Data Management (RDM). RDM is a general term covering how you organize, structure, store, and care for the information used or generated during a research project. The benefits of RDM, and the number now of funders, journals, and institutions requiring RDM are many and growing.


CONTINUE READING

<http://lib.hku.hk/researchdata/rds.htm>

# HKUL Research Data Services


**HKUL RESEARCH DATA SERVICES**

RPG STUDENTS | RPG SUPERVISORS | RDM | RESEARCH STAFF | DEPARTMENT HEADS



## RESEARCH DATA SERVICES


THE UNIVERSITY OF HONG KONG  
LIBRARIES

**RDM FOR RESEARCH POSTGRADUATE (RPG) STUDENTS - INSTRUCTIONS FOR DATA**

LOGIN RPG INPUT FORM


HOW TO SUBMIT FORM

### RPG DATA MANAGEMENT PLAN (DMP) INPUT FORM


RPG STUDENTS 


**RPG LOGIN**

# HKUL Research Data Services

**HKUL RESEARCH DATA SERVICES**

RPG STUDENTS | RPG SUPERVISORS | RDM | **RESEARCH STAFF** | DEPARTMENT HEADS






## RESEARCH DATA SERVICES

THE UNIVERSITY OF HONG KONG  
LIBRARIES

---

RDM PLANNING FOR RESEARCH STAFF

---

RDM REQUIREMENTS OF HKU, GRANT FUNDERS & JOURNALS


---


LOGIN DMP INPUT FORM

---

HOW TO SUBMIT FORM

## RESEARCH DATA MANAGEMENT PLANNING FOR RESEARCH STAFF

RESEARCH STAFF 



HKU Responsible Conduct of Research: Management of Res...


Watch later Share


**HKU Responsible Conduct of Research: Management of Research Data and Records**


<https://hub.hku.hk/researchdata/staff.htm>



# HKUL Research Data Services


**HKUL RESEARCH DATA SERVICES**

RPG STUDENTS | RPG SUPERVISORS | RDM | RESEARCH STAFF | DEPARTMENT HEADS 



## RESEARCH DATA SERVICES


THE UNIVERSITY OF HONG KONG  
LIBRARIES

**RDM PLANNING FOR RESEARCH STAFF**

RDM REQUIREMENTS OF HKU, GRANT FUNDERS & JOURNALS

LOGIN DMP INPUT FORM

### RESEARCH STAFF DATA MANAGEMENT PLAN (DMP) INPUT FORM

RESEARCH STAFF 

Principal Investigators please click "Staff Login" to submit Data Management Plan and dataset.

STAFF LOGIN

# Who has to submit a DMP

PLEASE CHOOSE ONE OF THE FOLLOWING: A, B, C, OR D.

- ☐ A. Data is freely available on the internet, in libraries or archives. DMP and Dataset submission are not needed. Primary supervisor approval will be sought.
- ☐ B. Data has been licensed, contracted for, or purchased with a license that explicitly forbids deposit in storage outside the student's or the primary supervisor's control. Primary supervisor approval will be sought.
- ☐ C. No data was used in my research project for the creation of my thesis. DMP and Dataset submission is not needed. Primary supervisor approval will be sought.
- ☐ D. Submit Data Management Plan (DMP). Dataset will be uploaded later.

$$\frac{dN}{dt} = \frac{1}{qV_{act}} - q_0(N-N_0)(1-\epsilon S)S + \frac{N_e}{T_n} - \frac{N}{T_p}$$

$$\frac{dS}{dt} = T_0 q_0(N-N_0)(1-\epsilon S)S + \frac{q_0 N}{T_n} - \frac{S}{T_p}$$

$$\frac{S}{P_1} = \frac{T_p \lambda_0}{T} = \Theta$$

$$S \leq \frac{1}{\epsilon}$$

$$N = N_0$$

$$P_1 = (m$$

3.

Prepare a **DMP** with **DMPTool**



## Welcome to DMPTool@HKU Libraries

Create data management plans that meet institutional and funder requirements.



### DMPTool by the Numbers



705  
Users



619  
Plans



HKU  
Participating Institution

### Contact Us

For any questions, please send email to HKUL Research Data Services at [researchdata@hku.hk](mailto:researchdata@hku.hk)

# Administrative Data (RPG)

- Your Name
- University ID
- Email
- Degree
- Department/Faculty
- Field of Study
- Supervisor(s)
- Project Title and Description
- Date and Version



## Data Management Plan for Post-Graduate Research Projects

Name:	
Student ID:	
Email:	
Faculty/Department:	
Supervisor:	

### RESEARCH PROJECT TITLE

#### DATA COLLECTION

How will the data be collected or created?

What data will you collect or create?

#### DOCUMENTATION AND METADATA

What documentation and metadata will accompany the data?

#### ETHICS AND LEGAL COMPLIANCE

How will you manage copyright and Intellectual Property Rights (IPR) issues?

How will you manage any ethical issues?

#### STORAGE AND BACKUP

How will you manage access and security?

How will the data be stored and backed up during the research?

#### SELECTION AND PRESERVATION

What is the long-term preservation plan for the dataset?

Which data are of long-term value and should be retained, shared, and/or preserved?

#### DATA SHARING

Are any restrictions on data sharing required?

How will you share the data?

#### RESPONSIBILITIES AND RESOURCES

What resources will you require to deliver your plan?

Who will be responsible for data management?

Prepared by:	Endorsed by:
RPg student:	Supervisor:
Date:	Date:

RPGs only

Research\_Project\_DMP.doc

# Supervisor's Endorsement (RPG)



## RESPONSIBILITIES AND RESOURCES

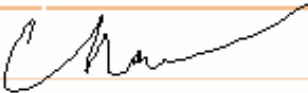
Who will be responsible for data management?

I will be responsible

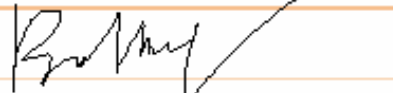
What resources will you require to deliver your plan?

Cost for acquiring external hard disks for off-site copy

**Prepared by:**

	Postgraduate student
Name:	
Date:	6-11-2017

**Approved by:**

	Supervisor
Name:	
Date:	6-11-2017

# Administrative Data (PI)

<b>Project Title:</b>	<b>A cost-effectiveness threshold of body mass patients</b>
<b>HKU Project Code:</b>	0201812
<b>Principal Investigator:</b>	Professor Chan Tai Man
<b>Co-Investigator(s):</b>	Dr Lam Mary Professor Smith Sandra Dr Cheung David
<b>Start Date:</b>	2018-07-05
<b>Completion Date:</b>	2020-07-04
<b>Grant Type:</b>	HMRF Research Fellowship Scheme
<b>Amount:</b>	2000000
<b>Funding Year:</b>	2018/2019

**YOU HAVE SELECTED OPTION D:**

D. Submit Data Management Plan (DMP). Dataset will be uploaded later.

You may proceed to submit your Data Management Plan (DMP) by drag and drop your file to the box below:

Drag and drop files here, or click in box to choose files.

After submitting your DMP, an email will go to your supervisors.

If you would like to update your DMP later, you may revisit this page to upload and replace your previous DMP with an updated version.

Please click "**SUBMIT DMP NOW**" button below to proceed, or click "**LOGOUT**" to exit form without submission.

SUBMIT DMP NOW

LOGOUT

$$\frac{dN}{dt} = \frac{1}{qV_{act}} - g_0(N-N_0)(1-\epsilon S)S + \frac{N_e}{T_n} - \frac{N}{T_p}$$

$$\frac{dS}{dt} = T_0 g_0(N-N_0)(1-\epsilon S)S + \frac{f_0 N}{T_n} - \frac{S}{T_p}$$

$$\frac{S}{P_1} = \frac{T_p \lambda_0}{T} = 0$$

$$S \leq \frac{1}{\epsilon}$$

$$N = N_0$$

$$P_1 = (m$$

4.

Major components of a **DMP**



# 7 Major Components of a DMP

1. Data Collection
2. Documentation and Metadata
3. Ethics and Legal Compliance
4. Storage and Backup
5. Selection and Preservation
6. Data Sharing
7. Responsibilities and Resources

# 1. Data Collection



# 1. Data Collection

- What data will you collect or create?
- How will the data be collected or created?

# 1. Data Collection

- What data will you collect or create?

- Numbers
- Text – survey data, interview transcripts
- Multimedia – image, audio, video
- Software, programming scripts
- Models

*in digital or physical forms*

# 1. Data Collection

- What data will you collect or create?
- How will the data be collected or created?

Type, format, volume of data

Expected rate of increase

How will you structure and name your folders and files?

What standards or methodologies will you use?



# 1. Data Collection

Data stage	Specification of type of research data	Data size/ growth
Raw data		
Processed data		

# 1. Data Collection

## File Formats

It is important to plan for software obsolescence.  
Formats more likely to be accessible in the future are:

- **Non-proprietary**
- **Open, documented standard**
- **Standard representation (ASCII, Unicode)**
- **Unencrypted**
- **Uncompressed**

Examples of preferred file format choices include:

**TXT, or PDF/A,**

**CSV,**

**MPEG-4,**

**WAV,**

**TIFF or JPEG2000,**

**not Word**

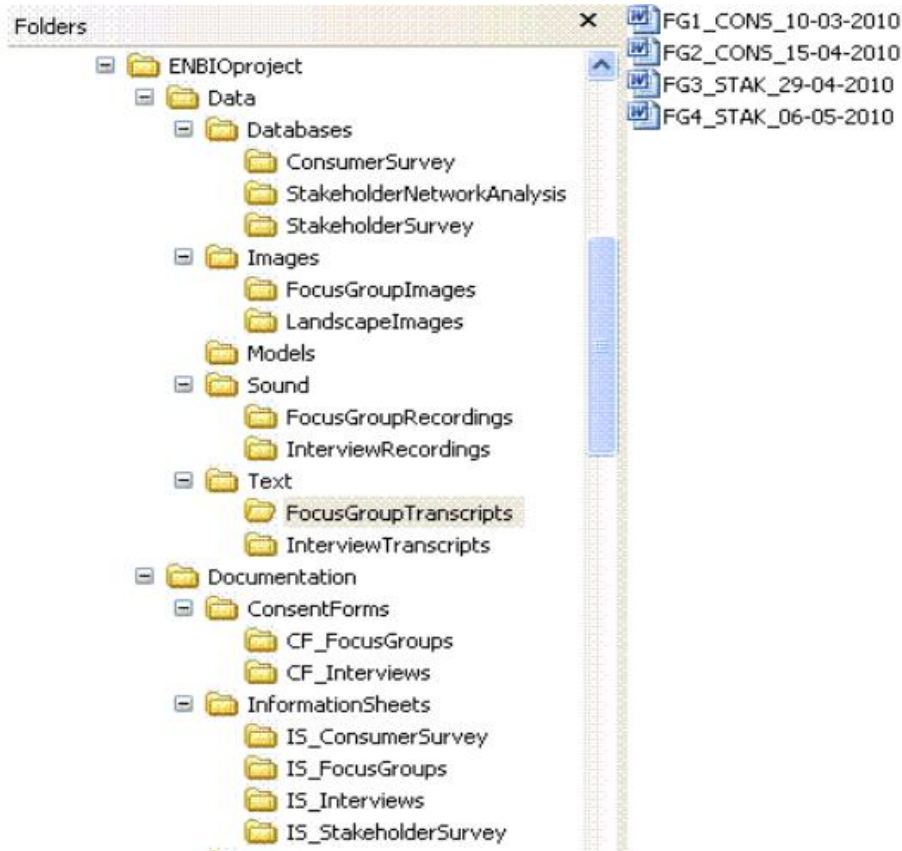
**not Excel**

**not Quicktime**

**not MP3**

**not GIF or JPG**

# 1. Data Collection



Example folder structure

Meaningful and  
consistent naming  
convention for files  
and folders

# 1. Data Collection

## File names – Best practice is to:

- create meaningful but brief names
- use file names to classify types of files
- **avoid using spaces, dots and special characters** (& or ? or !)
- use hyphens (-) or underscores (\_) to separate elements in a file name
- avoid very long file names
- reserve the 3-letter file extension for application-specific codes of file format (e.g. .doc, .xls, .mov, .tif)
- include versioning within file names where appropriate

41

## 2. Documentation and Metadata





## 2. Documentation and Metadata

- What documentation and metadata will accompany the data?

What information is needed for the data to be read and interpreted in the future?

How will you capture/create the documentation and metadata?

What metadata standards will you use?

## 2. Documentation and Metadata

- What documentation and metadata will accompany the data?



**Readme  
file**



**Codebook/  
Data dictionary**



**Electronic Lab  
Notebooks**

## 2. Documentation and Metadata

Basic information needed to make the data reusable.

Title  
Creator  
Persistent identifier  
Subject  
Funders  
Rights  
Access information  
Dates  
File names  
File format  
Versions

Software and version of the  
software required for its  
potential reuse.

Hardware and operation  
system requirements

## 2. Documentation and Metadata

**FAIR** Data Principles: Data should be

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**e-usable

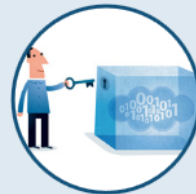
## 2. Documentation and Metadata

### **FAIR** Data Principles



Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

**FINDABLE**



Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

**ACCESSIBLE**



Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

**INTEROPERABLE**



Data and collections have a clear usage licenses and provide accurate information on provenance.

**REUSABLE**

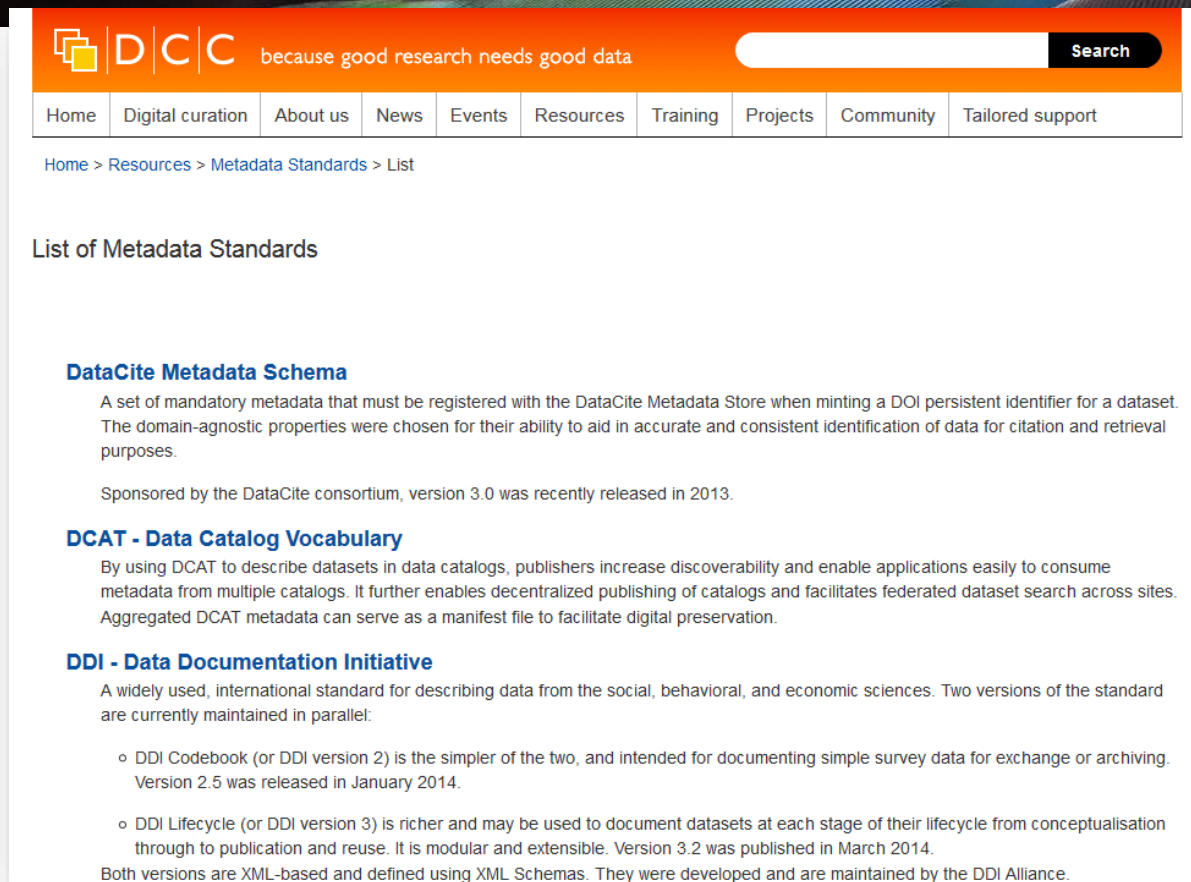


## 2. Documentation and Metadata

### **Metadata Standards**

Metadata (**data about data**) standards help to describe data in a consistent manner. Metadata can include descriptive information, provenance, quality and access/use of data.

## 2. Documentation and Metadata



The screenshot shows the DCC (Digital Curation Centre) website. The header is orange with the DCC logo and the tagline 'because good research needs good data'. A search bar is on the right. Below the header is a navigation menu with links: Home, Digital curation, About us, News, Events, Resources, Training, Projects, Community, and Tailored support. The breadcrumb trail is 'Home > Resources > Metadata Standards > List'. The main heading is 'List of Metadata Standards'. There are three sections listed: 'DataCite Metadata Schema', 'DCAT - Data Catalog Vocabulary', and 'DDI - Data Documentation Initiative'. Each section has a brief description and, in the case of DDI, a bulleted list of details.

**DCC** because good research needs good data  **Search**

[Home](#) [Digital curation](#) [About us](#) [News](#) [Events](#) [Resources](#) [Training](#) [Projects](#) [Community](#) [Tailored support](#)

[Home](#) > [Resources](#) > [Metadata Standards](#) > List

### List of Metadata Standards

#### DataCite Metadata Schema

A set of mandatory metadata that must be registered with the DataCite Metadata Store when minting a DOI persistent identifier for a dataset. The domain-agnostic properties were chosen for their ability to aid in accurate and consistent identification of data for citation and retrieval purposes.

Sponsored by the DataCite consortium, version 3.0 was recently released in 2013.

#### DCAT - Data Catalog Vocabulary

By using DCAT to describe datasets in data catalogs, publishers increase discoverability and enable applications easily to consume metadata from multiple catalogs. It further enables decentralized publishing of catalogs and facilitates federated dataset search across sites. Aggregated DCAT metadata can serve as a manifest file to facilitate digital preservation.

#### DDI - Data Documentation Initiative

A widely used, international standard for describing data from the social, behavioral, and economic sciences. Two versions of the standard are currently maintained in parallel:

- DDI Codebook (or DDI version 2) is the simpler of the two, and intended for documenting simple survey data for exchange or archiving. Version 2.5 was released in January 2014.
- DDI Lifecycle (or DDI version 3) is richer and may be used to document datasets at each stage of their lifecycle from conceptualisation through to publication and reuse. It is modular and extensible. Version 3.2 was published in March 2014.

Both versions are XML-based and defined using XML Schemas. They were developed and are maintained by the DDI Alliance.

<http://www.dcc.ac.uk/resources/metadata-standards/list?page=3>



# Harvard Biomedical Data Management

*Best practices & support services for research data lifecycles*

About ▾ Best Practices ▾ Plan ▾ Store ▾ Share ▾ Resources Support

HOME / PLAN /

## Metadata Overview

“Metadata is structured information that makes it easier to retrieve, use, or manage another resource. It is information about data or information about information. Metadata describes data that was created, analysed and stored. (National Information Standards Organization)”

Good metadata enables you to understand your data, helps other researchers discover, access, and use it, and facilitates long-term archival preservation of your data.

Biomedical metadata may include:

### Metadata schemas:

- [MIBBI – Minimum Information for Biological Investigators](#) (portal to over 40 biomedical data standards)
- [OME-XML – Open Microscopy Environment XML](#) (microscopy data)
- [Protocol Data Elements Definitions](#) (clinical trials data)
- [Digital Curation Centre's list of Disciplinary Metadata Standards](#)
- [Data Documentation Initiative](#) (social sciences data)
- [Dublin Core](#) (general)
- [Darwin Core](#) (Biological Data)

<https://datamanagement.hms.harvard.edu/metadata-overview>

## 2. Documentation and Metadata

### Dublin Core Metadata Initiative

[Home](#) [News](#) [DCMI Specifications](#) [LRMI](#) [Community and Events](#) [Join / Support](#) [About](#)

#### DCMI Specifications

- [Recommendations](#)
- [Proposed](#)
- [Working Drafts](#)
- [Superseded](#)
- [Community Specifications](#)
- [Recommended Resources](#)
- [Approval Processes](#)
- [Translations](#)

As part of its mission, the Dublin Core Metadata Initiative develops and maintains specifications in support of resource description. Specifications developed and reviewed in the context of DCMI's [formal approval process](#) are assigned a status (in ascending order of maturity and stability) of "DCMI Working Draft", "DCMI Proposed Recommendation", or "DCMI Recommendation". DCMI also provides pointers to guidelines and services developed outside of this formal review context ("Recommended Resources").

This selection highlights the specifications that currently attract the most attention in the Dublin Core community. Links to additional specifications (including superseded specifications) may be found at <http://dublincore.org/documents/>. Some of the specifications have been [translated](#) into one of twenty-five languages.

<http://dublincore.org/specifications/>

## 2. Documentation and Metadata

### Dublin Core Metadata Element Set (DCMES)

1. Title
2. Creator – Investigator, Photographer, Author, Composer
3. Subject
4. Description
5. Publisher
6. Contributor
7. Date – Collecting Date, Analysing Date, Create date
8. Type
9. Format
10. Identifier– HKID, Patient ID
11. Source
12. Language
13. Relation
14. Coverage
15. Rights Management

# 3. Ethics and Legal Compliance



### 3. Ethics and Legal Compliance

- How will you manage any ethical issues?
- How will you manage copyright and Intellectual Property Rights (IPR) issues?

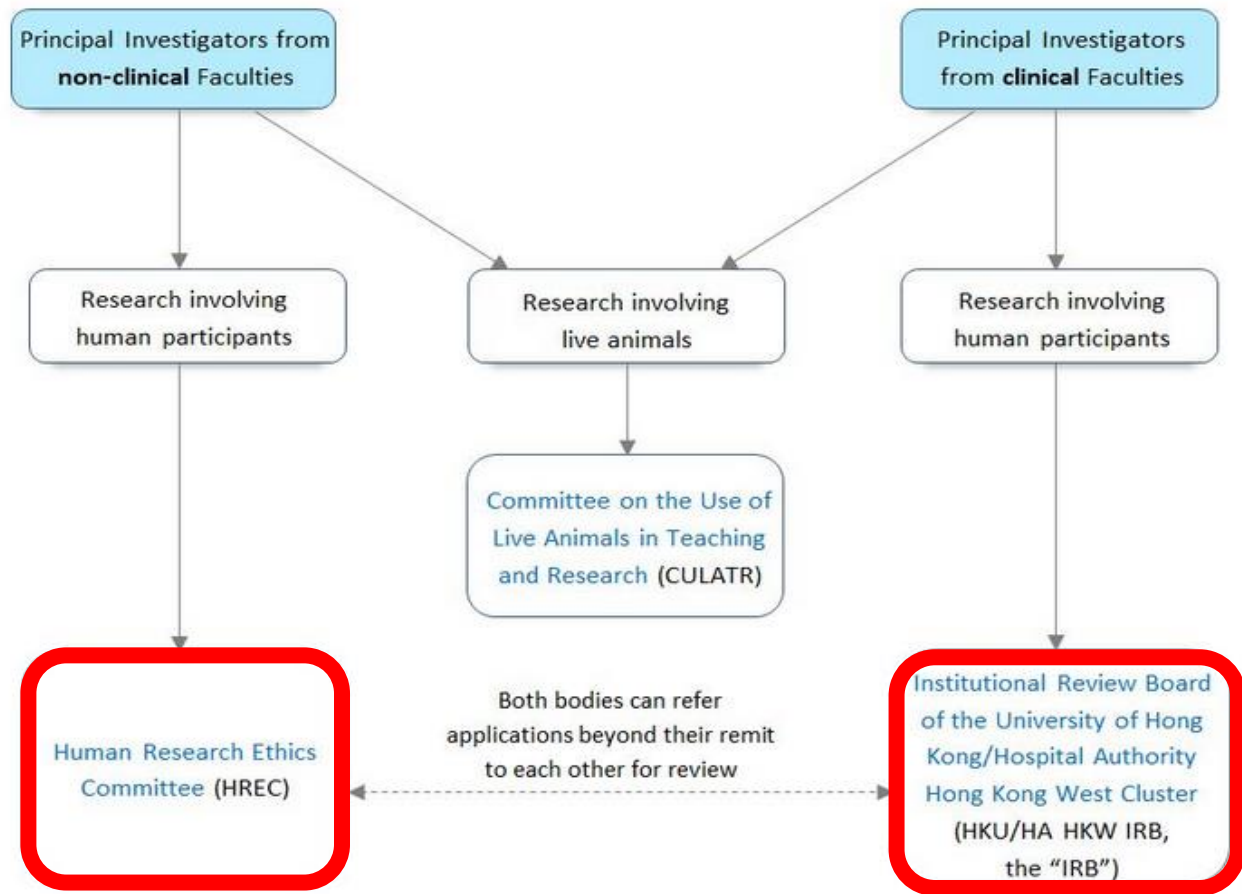
Have you gained consent for data preservation and sharing?

Who owns the data?

How will the data be licensed for reuse?

Any privacy or security issues? How are you dealing with them?

## Overview of Procedures



<http://www.rss.hku.hk/integrity/ethics-compliance>

# 3. Ethics and Legal Compliance

**LSE** THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

## LSE Impact Blog

Home About Latest Our books Series Resources LSE comment Popular

The “long tail” of research impact is engendered by innovative dissemination tools and meaningful community engagement

Research impact often tends not to happen in one emphatic, public moment but rather at more discrete points of the “long tail” of a research project. Achieving this depends largely on the tenacity of the research team but also on key allies such as the community members and service providers who have become energised by the work and inspired to continue it. Taking their own research as a case study, **Kip Jones** and **Lee-Ann Fenge** discuss what it takes to create meaningful community impact, highlighting a commitment to inclusive co-production and public engagement and the use of participatory research to create innovative dissemination tools.

This work is licensed under a **Creative Commons Attribution 3.0 Unported License** unless otherwise stated.

<https://blogs.lse.ac.uk/impactofsocialsciences/2018/02/20/the-long-tail-of-research-impact-is-engendered-by-innovative-dissemination-tools-and-meaningful-community-engagement/>

# 3. Ethics and Legal Compliance

## The Licenses



**Attribution**  
CC BY

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

[View License Deed](#) | [View Legal Code](#)



**Attribution-ShareAlike**  
CC BY-SA

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to “copyleft” free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.



**Attribution-NoDerivs**  
CC BY-ND

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

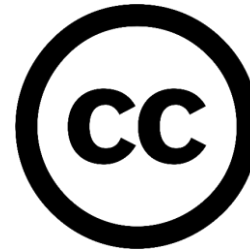
[View License Deed](#) | [View Legal Code](#)



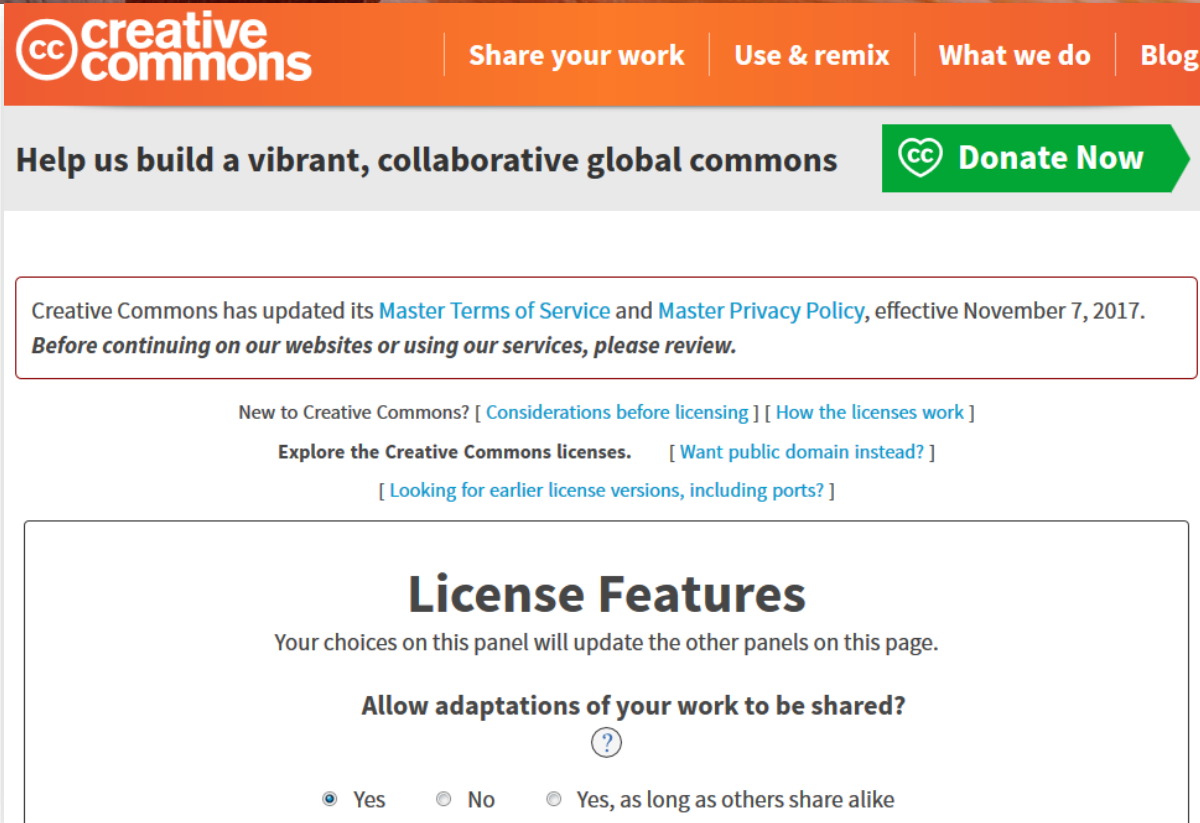
**Attribution-NonCommercial**  
CC BY-NC

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

60



# 3. Ethics and Legal Compliance



The screenshot shows the Creative Commons website interface. At the top is an orange navigation bar with the Creative Commons logo and links for 'Share your work', 'Use & remix', 'What we do', and 'Blog'. Below this is a grey banner with the text 'Help us build a vibrant, collaborative global commons' and a green 'Donate Now' button. A red-bordered box contains a notice about updated terms and privacy policies. Below the notice are several links for new users. The main content area is titled 'License Features' and includes a sub-header 'Allow adaptations of your work to be shared?' with a help icon and three radio button options: 'Yes' (selected), 'No', and 'Yes, as long as others share alike'.

**creative commons** | [Share your work](#) | [Use & remix](#) | [What we do](#) | [Blog](#)

**Help us build a vibrant, collaborative global commons** [Donate Now](#)

Creative Commons has updated its [Master Terms of Service](#) and [Master Privacy Policy](#), effective November 7, 2017.  
*Before continuing on our websites or using our services, please review.*

New to Creative Commons? [ [Considerations before licensing](#) ] [ [How the licenses work](#) ]  
[Explore the Creative Commons licenses.](#) [ [Want public domain instead?](#) ]  
[ [Looking for earlier license versions, including ports?](#) ]

## License Features

Your choices on this panel will update the other panels on this page.

**Allow adaptations of your work to be shared?** ⓘ

☒ Yes   ☐ No   ☐ Yes, as long as others share alike

<https://creativecommons.org/choose/>

## Data Confidentiality

Research records will be kept confidential, and access will be limited to the PI and primary research team members. For each testing session, the recorded data will have any identifying information removed and will be relabeled with study code numbers. A database which relates study code numbers to consent forms and identifying information will be stored separately on password-protected computers in a secured, locked office. These computers are housed in research facilities in the Psychology Building at Indiana University-Bloomington, and in the Psychology Department at UCSD. A list of the names of individuals who have participated in each study will be maintained in order to ensure that no individual is tested more than once on related studies. To maintain the privacy of the participants, any report of individual data will only consist of performance measures without any demographic or identifying information.

### 3. Ethics and Legal Compliance

- How will you manage any ethical issues?
- How will you manage copyright and Intellectual Property Rights (IPR) issues?



# 4. Storage and Backup



## 4. Storage and Backup

- How will the data be stored and backed up during the research?
- How will you manage access and security?

How much storage?

How fast will the data grow?

Need to include costs for storage?

Who will be able to access or use your data?

How will you protect the confidentiality of your subjects?

## 4. Storage and Backup

Data stage	Storage location	Backup procedures (storage medium and location/ how often?)
Raw data		
Processed data		

## 4. Storage and Backup

**3-2-1**

**Rule**

3 copies

2 media

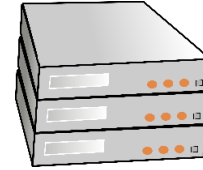
1 offsite



Personal  
computer



External  
hard drive



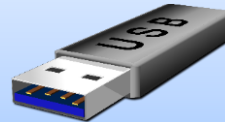
Networked  
drives



Cloud



CD/DVD



## 2. Data Storage and Preservation

Our short-term data storage plan, which will be used during the experiment, will be to save copies of 1) the .txt metadata file and 2) the Excel spreadsheet as .csv files to an external drive, and to take the external drive off site nightly. We will use the Subversion version control system to update our data and metadata files daily on the University of Alberta Mathematics Department server. We will also have the laboratory notebook as a hard copy backup that will be stored in a fire-proof cabinet.

The data set will be submitted to the Knowledge Network for Biocomplexity (KNB) data repository for long-term preservation and storage. The authors will submit metadata in EML format along with the data to facilitate its reuse. The data manager will be responsible for updating metadata and data author contact information in the KNB.

## 4. Storage and Backup

- How will the data be stored and backed up during the research?
- How will you manage access and security?



# 5. Selection and Preservation





## 5. Selection and Preservation

- Which data are of long-term value and should be retained, shared, and/or preserved?
- What is the long-term preservation plan for the dataset?

What data must be retained/destroyed for contractual, legal, or regulatory purposes?

Where or in which repository or data archive will the data be preserved (e.g. institution repository)?

## 5. Selection and Preservation

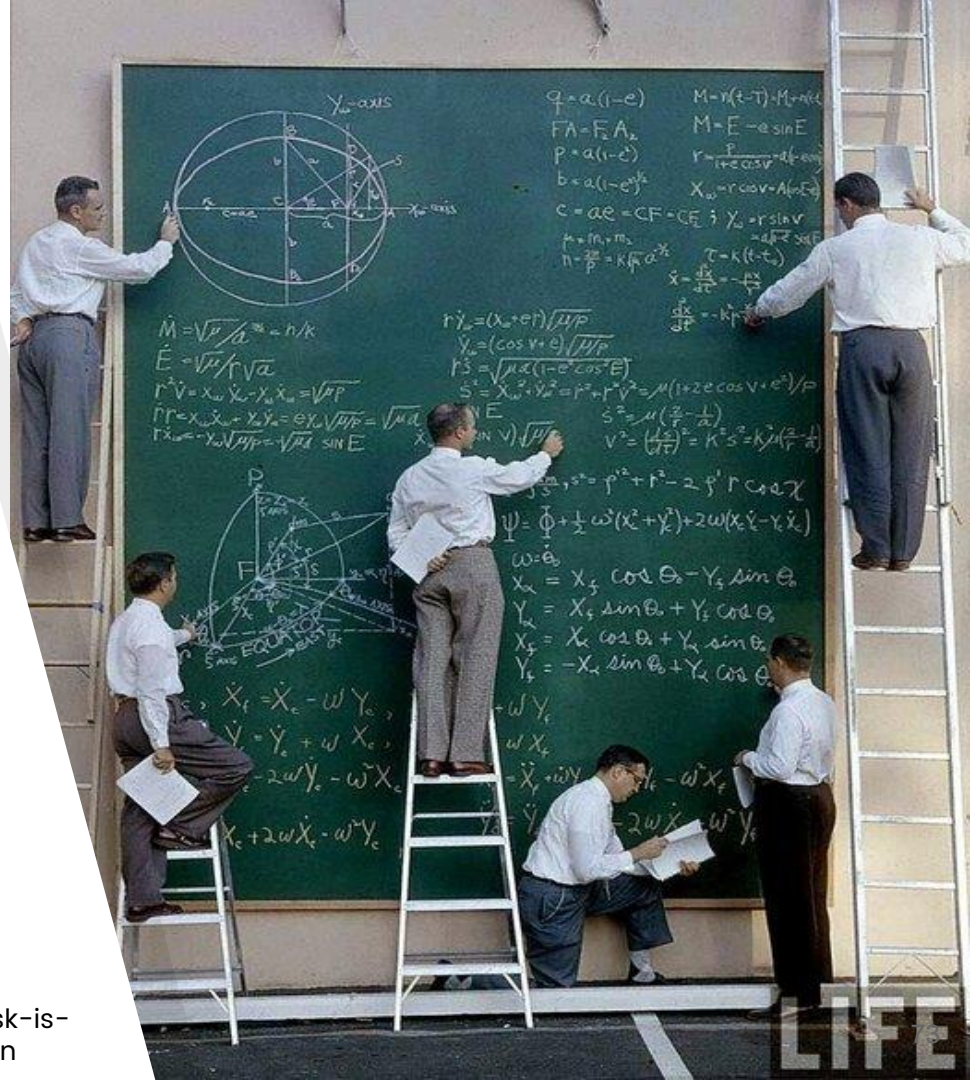
### Criteria for data appraisal

- Relevance to mission
  - Fit with goals and priorities of the institution
  - A legal requirement to keep the data
- Scientific or historical value
  - Data is scientifically, socially or culturally significant
- Uniqueness
  - The dataset is the only available copy
- Potential for redistribution
  - Reliable and usable for future use
- Nonreplicability
  - Costly to replicate
  - Derived from unrepeatable observations

# 6. Data Sharing

© J. R. Eyerman, LIFE magazine

Image courtesy of: <https://www.sott.net/article/314865-Cluttered-desk-is-a-sign-of-a-cluttered-mind-of-what-then-is-an-empty-desk-a-sign>



## 6. Data Sharing

- How will you share the data?
- Are any restrictions on data sharing required?

Not a requirement that you share all of your data with anyone who asks

Access conditions, restrictions, embargoes etc.

How will potential users find out about your data?

Indicate the criteria for deciding who can receive your data and whether or not you will place any conditions on their use

# 6. Data Sharing

The screenshot shows the website of The University of Hong Kong Research Services. The header includes the university's name and logo, a search bar, and links to Sitemap and Contact Us. The main content area is titled "Research Data and Records Management" and includes a breadcrumb trail: Home > Research Integrity > Research Data and Records Management. The page text discusses the management of research data and records, mentioning the importance of research integrity and the University's commitment to high standards. A sidebar on the left lists various research services, including "Research Data and Records Management".

The University of Hong Kong

Research Services  
Support and information for HKU researchers

Home > Research Integrity > Research Data and Records Management

## Research Data and Records Management

The management of research data (medium) from research is organised for the long term. Effective research integrity.

HKU recognises the importance of research integrity and the highest standards. The University at its meeting on May 5, 2015, also approved the **Records** to oversee the planning of research data management.

[Policy on the Management of Research Data and Records](#)

**Research Data and Records Management**

- About Us
- Research Funding
- IPR & Contracts
- Research Integrity
  - Responsible Conduct of Research
  - Research Data and Records Management**
  - Research Ethics Compliance
  - Safety Approval & Insurance

## 4. Research data and records should be:

kept in a manner that is compliant with legal obligations and, where applicable, the requirements of funding bodies and project-specific protocols **approved by the Institutional Review Board (IRB) and Human Research Ethics Committee (HREC)**; and

able to be made available to others in line **with appropriate ethical, data sharing and open access principles** (3).

<http://www.rss.hku.hk/integrity/research-data-records-management>

## 6. Data Sharing

Not all data can be shared



Licensed  
data



Privacy,  
confidential,  
sensitive  
data



Data  
supporting  
a patent



“

“Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data.”

— *Final NIH Statement on Sharing Research Data February 26, 2003*

## 6. Data Sharing

### List of 18 HIPAA Identifiers

**The Health Insurance Portability and Accountability Act (HIPAA)** of 1996 specifies a number of elements in health data that are considered identifiers. If any are present, the health information cannot be released without patient authorization. Such data can be released for research purposes with approval of a waiver of patient authorization from an Institutional Review Board (IRB).

# 6. Data Sharing

## List of 18 HIPAA Identifiers

1. Names;
2. Address,
3. All elements of dates (except year) including birth date, admission date, discharge date, date of death;
4. Phone numbers;
5. Fax numbers;
6. Electronic mail addresses;
7. Social Security numbers;
8. Medical record numbers;
9. Health plan beneficiary numbers;
10. Account numbers;
11. Certificate/license numbers;
12. Vehicle identifiers and serial numbers, including license plate numbers;
13. Device identifiers and serial numbers;
14. Web Universal Resource Locators (URLs);
15. Internet Protocol (IP) address numbers;
16. Biometric identifiers, including finger and voice prints;
17. Full face photographic images; and
18. Any other unique identifying number, characteristic, or code (note this does not mean the unique code assigned by the investigator to code the data)

## 6. Data Sharing

### Indirect Identifiers

- Detailed geographic information (e.g. state, country, province)
- Organizations to which the respondent belongs
- Educational institutions (from which the respondent graduated and year of graduation)
- Detailed occupational titles
- Place where respondent grew up
- Exact dates of events (birth, death, marriage, divorce)
- Detailed income
- Offices or posts held by respondent



### Issues in Genetics

Coverage and  
Reimbursement of Genetic  
Tests

FDA requests comments on  
draft guidance for Precision  
Medicine Initiative

Genetic Discrimination

Genome Editing

Genome Statute and  
Legislation Database

Genomics and Health  
Disparities

Human Subjects Research in  
Genomics

Informed Consent for

## Privacy in Genomics



- 🔍 Overview
- 🔍 Privacy in Research
- 🔍 Privacy in the Clinic
- 🔍 Privacy in Society
- 🔍 Protections
- 🔍 Genetic Privacy Links

### Overview

Each person's DNA sequence includes health and other information about them and their families. Technological advances mean that it is now cheaper and easier than ever to sequence and interpret genomic information. Whether genomic information is being used for research, clinical or other uses, it is important to consider how best to ensure that individuals' privacy is respected. There are laws and policies that serve to protect the privacy of individuals' genomic information, and there is ongoing debate as to whether further measures are needed.

### Privacy in Research

<https://www.genome.gov/27561246/privacy-in-genomics/>



## Privacy in Genomics

### NIH Genomic Data Sharing Policy

The [NIH Genomic Data Sharing Policy](#) sets guidelines on how to protect research participant privacy while still enabling the scientific community access to valuable research data. A key component of the policy is that access to sensitive, individual-level research data held in federal databases is only available to researchers submitting a request. NIH maintains several databases containing such genomic information, such as the [database of genotypes and phenotypes \(dbGaP\)](#), the [National Database for Autism Research \(NDAR\)](#), and [The Cancer Genome Atlas \(TCGA\)](#). To access sensitive data from one of these databases, scientists must request permission for specific uses from [Data Access Committees](#) at the NIH or the database's curating body. It is important to note that not all information in these databases is held under 'controlled access,' and that some data is readily accessible.

<https://www.genome.gov/27561246/privacy-in-genomics/>





## Working with data

### Sensitive data

[Safely sharing  
sensitive data](#)

[Ethics and data  
sharing](#)

[De-identifying your  
data](#)

[Health and medical  
data](#)

[Indigenous data](#)

ANDS | [Working with data](#) | [Sensitive data](#)

## De-identifying your data

[+SHARE](#)

De-identification aims to allow data to be used by others without the possibility of individuals being identified. Data de-identification may be used to:

- protect the privacy of individuals and organisations, such as businesses,
- ensure that the spatial location of mineral or archaeological findings or endangered species is not publicly available.

Data that is still identifiable (i.e. contains personal information) needs to

### Related ANDS Guides

- › [Publishing and sharing  
sensitive data](#)
- › [Data sharing  
considerations for  
Human research Ethics  
Committees \(HRECs\)](#)

### Related information

- › [Sensitive data](#)
- › [Ethics and data sharing](#)
- › [Medical and health data](#)



med.data.edu.au

Data

Contact

Resources

About



## ANONYMISATION

---

### What is Anonymisation?

Anonymisation (or de-identification, confidentialisation) is a process that removes all personal identifying information from data that represents an identifiable individual. One of the main purposes for anonymising personally related information (e.g. health or financial-related records) is to make this information accessible for secondary uses (such as publication or research) without infringing upon an individual's privacy.

---

### Three levels of data identifiability.

The [NHMRC National Statement on Ethical Conduct in Human Research \(2007 – Updated May 2015\)](#) states that with respect to data identifiability, data may be collected, stored or disclosed in three mutually exclusive forms:

- **individually identifiable data**, where the identity of a specific individual can reasonably be ascertained;
- **re-identifiable data**, from which identifiers have been removed and replaced by a code, but it remains possible to re-identify a specific individual by, for example, using the code or linking different data sets;

<https://med.data.edu.au/anonymisation/>

## 6. Data Sharing

### Options

- Institutional repository
- Interdisciplinary repository
- Domain/Subject specific repository
- Self-dissemination through website



# 6. Data Sharing

## Institutional repository

connected to  
the researcher's  
institution



The screenshot displays the homepage of The HKU Scholars Hub, an institutional repository for The University of Hong Kong. The header features the university's crest and name in English and Chinese, along with navigation links for Home, Publications, Researchers, Organizations, Grants, Datasets, Theses, Patents, and Community Service. A central banner with a gear graphic reads "THE HKU SCHOLARS HUB AT THE CENTRE OF HKU". To the right, a text box explains the hub's role in HKU's Knowledge Exchange Initiative. Below this is a search bar with a "Search Q" button. The main content area is divided into three columns: "Featured Scholar" (highlighting Dr. Yiu, Siu Ming, an Associate Professor in Cryptography, Computer Security, and Bioinformatics), "Hub News" (listing recent updates on theses on Amazon and top 1% scientists), and "Relevant Links" (providing links to top scientists, open access, usage stats, and research on the Web of Science).

The University of Hong Kong  
**The HKU Scholars Hub** 香港大學學術庫

Home Publications Researchers Organizations Grants Datasets Theses Patents Community Service

THE HKU SCHOLARS HUB  
AT THE CENTRE OF HKU

The HKU Scholars Hub is the current research information system of The University of Hong Kong. As a key vehicle of HKU's [Knowledge Exchange Initiative](#), The Hub aims to enhance the visibility of HKU authors and their research, and to foster opportunities for collaboration.

Quick Search BETA Research Collaborations Thesis Supervisors Media Commentators

Search for Everything... Search Q

### Featured Scholar

  
**Dr Yiu, Siu Ming**  
• Associate Professor

*Research Interests*

- Cryptography
- Computer Security
- Bioinformatics

[+ MORE](#)

### Hub News

- Dec 2018: [HKU Theses On Amazon and Other Online Retailers](#) (Update).
- Oct 2018: [Report on the Sale of HKU Theses on Amazon and Other Websites](#).
- Oct 2018: [Announcing the 2018 HKU's Top 1% Scientists](#).
- Sep 2018: [HKU Theses On Amazon and Other Online Retailers](#) (Update).

[More](#)

### Relevant Links

- [HKU's Top 1% Scientists](#)
- [Open Access @HKU](#)
- [Usage Stats & Downloads](#)
- [HKU research on Web of Science](#)
- [Research @HKU](#)

## 6. Data Sharing

### Interdisciplinary repository

allows researchers from different disciplines to deposit and make their data available.



# 6. Data Sharing

## Domain/Subject specific repository

The screenshot shows the NCBI website interface. At the top, there's a navigation bar with 'NCBI', 'Resources', and 'How To' links. Below this is a search bar with a dropdown menu set to 'All Databases' and a 'Search' button. On the left side, there's a sidebar with a list of resources including 'NCBI Home', 'Resource List (A-Z)', 'All Resources', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', 'Homology', 'Literature', 'Proteins', 'Sequence Analysis', 'Taxonomy', and 'Training & Tutorials'. The main content area is titled 'How to: Submit data to NCBI' and includes a section for 'SEQUENCE DATA' with a list of databases: GenBank, Sequence Read Archive (SRA), dbSNP, dbVar, and GEO. There's also a section for 'MICROARRAY DATA' with instructions on where to submit data. A green callout box on the right side of the page contains the text: 'National Center for Biotechnology Information (NCBI) for Health and Medical Sciences'.

<https://www.ncbi.nlm.nih.gov/guide/howto/submit-data/>



# 6. Data Sharing

## Find a Data Repository



The screenshot shows the 'Recommended Data Repositories' page on the Scientific Data journal website. The page has a dark blue header with the journal's name and navigation links. A left sidebar contains a list of policy-related links. The main content area features the title 'Recommended Data Repositories' followed by a paragraph explaining the journal's data policy and a subsequent paragraph detailing the evaluation criteria for recommended repositories.

nature > scientific data > policies > recommended data repositories a natureresearch journal

SCIENTIFIC DATA

Search E-alert Submit Login

Policies

- Editorial & Publishing Policies
- For Referees
- Data Policies
- Recommended Data Repositories

### Recommended Data Repositories

*Scientific Data* mandates the release of datasets accompanying our Data Descriptors, but we do not ourselves host data. Instead, we ask authors to submit datasets to an appropriate public data repository. Data should be submitted to discipline-specific, community-recognized repositories where possible, or to [generalist repositories](#) if no suitable community resource is available.

Repositories included on this page have been evaluated to ensure that they meet our requirements for data access, preservation and stability. Please be aware, however, that some repositories on this page may only accept data from those funded by specific sources, or may charge for hosting data. Please ensure you are aware of any deposition policies for your chosen repository. If your repository of choice is not listed please see our [guidelines for suggesting additional repositories](#).

<https://www.nature.com/sdata/policies/repositories>

## 6. Data Sharing

### Find a Data Repository

The image shows the top section of the Repository Finder website. It features a header with a background image of a hallway with several wooden doors. The text 'Repository Finder' is prominently displayed in white, with the subtitle 'Find a repository to upload your data.' below it.

**Repository Finder**  
Find a repository to upload your data.

Repository Finder, a pilot project of the [Enabling FAIR Data Project](#) led by the American Geophysical Union (AGU) in partnership with DataCite and the Earth, space and environment sciences community, can help you find an appropriate repository to deposit your research data. The tool is hosted by DataCite and queries the re3data registry of research data repositories.

Search [re3data](#) for a repository to upload your data

Search

<https://repositoryfinder.datacite.org/>

## 6. Data Sharing

# Google Dataset Search Beta

Search for Datasets

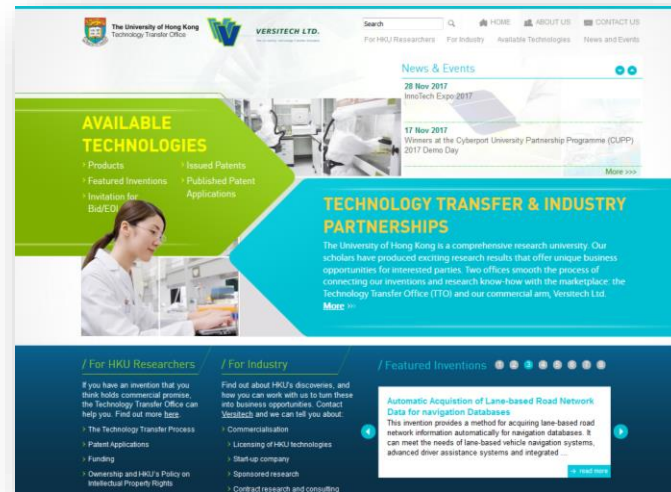


Try [boston education data](#) or [weather site:noaa.gov](#)

## 6. Data Sharing

### Before you share or publish your data

- Review the Depositor's Agreement, and Takedown Policy
- Perhaps you need to anonymize or redact your data before sharing?
- If you have created data which may have commercial value, please consult the **Technology Transfer Office**.



#### 4. Data Dissemination and Policies for Data Sharing and Public Access

We are required to share our data with the CAISN network after all data have been collected and metadata have been generated. This should be no more than 6 months after the experiments are completed. In order to gain access to CAISN data, interested parties must contact the CAISN data manager (data@caisn.ca) or the authors and explain their intended use. Data requests will be approved by the authors after review of the proposed use.

The authors will retain rights to the data until the resulting publication is produced, within two years of data production. After publication (or after two years, whichever is first), the authors will open data to public use. After publication, we will submit our data to the KNB enabling discovery and use by the wider scientific community. Interested parties will be able to download the data directly from KNB without contacting the authors, but will still be encouraged to give credit to the authors for the data used by citing a KNB accession number either in the publication's text or in the references list.

## 6. Data Sharing

- How will you share the data?
- Are any restrictions on data sharing required?

# 7. Responsibilities and Resources





# 7. Responsibilities and Resources

- Who will be responsible for data management?
- What resources will you require to deliver your plan?

Who is **collecting** the data? **analysing** the data?

Who is responsible for implementing the DMP, and ensuring it is reviewed and revised?

Who will be the contact person for questions regarding the research data?

Who can be contacted about the project after it has finished?

Project costs for data storage and costs for making the data accessible.

## 5. Roles and responsibilities

The PI will be responsible for all data management during and after data collection.

[https://www.dataone.org/sites/all/documents/DMP\\_Copepod\\_Formatted.pdf](https://www.dataone.org/sites/all/documents/DMP_Copepod_Formatted.pdf)

## 5. Plans for Archiving and Preservation

All original raw data files and data source processing programs will be versioned over time and maintained in a date-stamped file structure with text files documenting the provenance. The database will be preserved in perpetuity, housed initially at the New Mexico Interstate Stream Commission Central Office in addition to an off-site copy maintained at an NMISC field office and mirrored at the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI). We will also identify appropriate archiving institutions that might serve as a mirror repository. A data policy and stewardship plan will be established. In addition to archiving, each database table will be exported to a delimited text format to ensure accessibility of the data by other software programs. The data manager at the NMISC will be responsible for the management of long-term storage and archived data.

## 7. Responsibilities and Resources

- Who will be responsible for data management?
- What resources will you require to deliver your plan?

[https://www.dataone.org/sites/all/documents/DMP\\_Hydrologic\\_Formatted.pdf](https://www.dataone.org/sites/all/documents/DMP_Hydrologic_Formatted.pdf)

#### 4. Policies for Re-use, Distribution

Access to databases and associated software tools generated under the project will be available for educational, research and non-profit purposes. Such access will be provided using web-based applications, as appropriate.

Materials generated under the project will be disseminated in accordance with University/Participating institutional and NSF policies. Depending on such policies, materials may be transferred to others under the terms of a material transfer agreement.

Those that use the data (as opposed to any resulting manuscripts) should cite it as follows:

Lind, E, E Borer and A Kay. yyyy. Grassland Arthropod abundance and stoichiometry associated with nutrient manipulation. [URL]; accessed on ddmmyyyy.

This information will be described in the metadata.

Intended and foreseeable users of the data are NutNet collaborators and participants, as well as other scientists interested in arthropod-plant relationships. This data set could be used in combination with similar data sets from other NutNet sites or for meta-analysis.

#### 5. Plans for Archiving and Preservation

We will preserve both arthropod datasets generated during this project (abundance and stoichiometry) for the long term in the Digital Conservancy at the U of M. We will include the .csv files, along with the associated metadata files. We will also submit an abstract with the datasets that describe their original context and any potentially relevant project information. Borer will be responsible for preparing data for long-term preservation and for updating contact information for investigators.

$$\frac{dN}{dt} = \frac{1}{qV_{act}} - g_0(N-N_0)(1-\epsilon S)S + \frac{N_e}{T_n} - \frac{N}{T_p}$$

$$\frac{dS}{dt} = T_0 g_0(N-N_0)(1-\epsilon S)S + \frac{g_0 N}{T_n} - \frac{S}{T_p}$$

$$\frac{S}{P_1} = \frac{T_p \lambda_0}{T} = \Theta$$

$$S \leq \frac{1}{\epsilon}$$

$$N = N_0$$

$$P_1 = (m$$

5.

Tools and other Resources

# The Best Practices for Biomedical Research Data Management MOCC



[About Us](#) | [Take a Course](#) | [Login](#)



## BEST PRACTICES FOR BIOMEDICAL RESEARCH DATA MANAGEMENT

Self-paced

**COURSE DATE:**  
Open

**DURATION:**  
On-going

**COMMITMENT:**  
7+ hrs/week

**REQUIREMENT:**  
None

**COURSE TYPE:**  
Self-paced

**CREDENTIAL:**  
Badge, Certificate  
(free)

[Tweet](#)

### DESCRIPTION

Biomedical research today is not only rigorous, innovative and insightful, it also has to be organized and reproducible. With more capacity to create and store data, there is the challenge of making data discoverable, understandable, and reusable. Many funding agencies and journal publishers are requiring publication of

### COURSE INSTRUCTORS



<https://www.canvas.net/browse/harvard-medical/courses/biomed-research-data-mgmt>



# Example DMP

THE LIBRARY

UC San Diego

[Research & Collections](#) [Borrow & Request](#) [Computing & Technology](#) [Visit](#) [Ask Us](#) [About](#) [Hours](#) [Q](#)

[HOME](#) / [Research & Collections](#) / [Data Curation](#) / [UC San Diego Sample NSF Data Management Plans](#)



## Sample NSF Data Management Plans

Data Curation

Data Management Best Practices

NIH Policy on Rigor and Reproducibility

Sample Data Management Plans

Obtain identifiers

These examples from UC San Diego proposals are intended to provide a starting point for the development of other proposal-specific Data Management Plans.

We thank the UC San Diego investigators who gave permission to include their DMPs in this collection. If you have a DMP you'd be willing to have included here, please contact [Sharon Franks](#) or the library [Research Data Curation Program](#).

Please keep in mind that these examples are project-specific. PIs are encouraged to submit draft DMPs well in advance of the proposal deadline to OCGA to ensure compliance with University policy.

---

### Office of the Director (OD)

### Office of Cyberinfrastructure (OD/OCI)



- [DMP Example Allan Snaveley](#) From Allan Snaveley's proposal to the Strategic Technologies for Cyberinfrastructure

Example\_DMP\_1.doc

<https://libraries.ucsd.edu/research-and-collections/data-curation/dmp-samples.html>



# Example DMP



About WageningenCareerContactLoginen|English

Education & ProgrammesResearch & ResultsExpertise & Services


Home > What is a Data Management Plan?

## What is a Data Management Plan?

Research data exist in numerous forms, varying from logbooks with observations and protocols, to experimental and model data. A data management plan is a project document which tells the story of your research data. It outlines what research data were collected, how they were collected and what you will do with your data during and after your research.

To guide you in completing your DMP, Wageningen University provides you with a [template](#) with 10 questions about expected data formats, ownership, documentation, archiving and reuse. A [presentation](#) illustrates the template's questions in more detail. You can also have a look at some completed Plans.

- > Example DMP 1
- > Example DMP 2



Contact  
Servicedesk Library  
[Contact form](#)



Example\_DMP\_2.pdf, Example\_DMP\_3.pdf

<https://www.wur.nl/en/show/What-is-a-Data-Management-Plan.htm>

# Example DMP

**cessda eric**

About ▾ Consortium ▾ Projects ▾ Research Infrastructure ▾ Contact

## Expert Tour Guide on Data Management

- 1. Plan**
  - Benefits of data management
  - Research data
  - Data in social sciences
  - FAIR data
  - European diversity
  - Adapt your DMP: Part 1**
    - Sources and further reading
- 2. Organise & Document**
- 3. Process**
- 4. Store**
- 5. Protect**
- 6. Archive & Publish**

### Examples of DMP questions and answers


For inspiration of filled in DMPs look at some example DMPs we prepared. Both DMPs are based on a fictional research project with a basis in reality. For each topic of the DMP, there are example questions and answers where applicable. The examples are not country specific. Some of the information is generic.



#### Qualitative data

During this project, in-depth interviews with teachers in primary school will be held. The project has just started. Click the link to view and download the DMP:

[DMPQuestionsQualitativeData.pdf](#) (165 KB)



#### Quantitative data

The project concerns a survey which is conducted in order to identify how the evolution of society affects attitudes and behaviour. The project is still running. Click the link to view and download the DMP:

Example\_DMP\_4.pdf, Example\_DMP\_5.pdf

<https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/1.-Plan/Adapt-your-DMP-Part-1>

# Resources

Karl W. Broman & Kara H. Woo (2018) Data Organization in Spreadsheets, The American Statistician, 72:1, 2–10, DOI: 10.1080/00031305.2017.1375989

The screenshot shows the Taylor & Francis Online interface. At the top, there's a blue header with the Taylor & Francis logo, 'Taylor & Francis Online', and links for 'Log in', 'Register', and 'Cart'. Below the header, the journal 'The American Statistician' is featured with its cover image and the text 'Volume 72, 2018 - Issue 1: Special Issue on Data Science'. There are buttons for 'Submit an article' and 'Journal homepage'. A search bar with the placeholder 'Enter keywords, authors, DOI etc.' and a dropdown for 'This Journal' are also present. The main article section displays the title 'Data Organization in Spreadsheets' by Karl W. Broman & Kara H. Woo. It includes the article number 47276, 1 view, 1 CrossRef citation, and 912 Altmetric mentions. The article is marked as 'Open access'. Below the title, there are links for 'Download citation' and the DOI 'https://doi.org/10.1080/00031305.2017.1375989', along with a 'Check for updates' button. A navigation bar at the bottom of the article section offers links for 'Full Article', 'Figures & data', 'References', 'Citations', 'Metrics', 'Licensing', and a PDF icon. The 'ABSTRACT' section begins with a 'Select Language' dropdown and a 'Translator disclaimer'. The abstract text states: 'Spreadsheets are widely used software tools for data entry, storage, analysis, and visualization. Focusing on the data entry and storage aspects, this article offers practical recommendations for organizing spreadsheet data to reduce errors and ease later analyses. The'. To the right of the abstract, there's a 'People also read' section with a card for 'Discussion 50 Years of Data'.

Taylor & Francis Online

Log in | Register | Cart

Journal  
**The American Statistician** >  
Volume 72, 2018 - Issue 1: Special Issue on Data Science

Enter keywords, authors, DOI etc. This Journal

Submit an article Journal homepage

Advanced search

Article

47276  
Views  
1  
CrossRef citations  
912  
Altmetric

**Data Organization in Spreadsheets**

Karl W. Broman & Kara H. Woo

Pages 2-10 | Received 01 Jun 2017, Accepted author version posted online: 29 Sep 2017, Published online: 29 Sep 2017

Open access

Download citation <https://doi.org/10.1080/00031305.2017.1375989> Check for updates

Full Article Figures & data References Citations Metrics Licensing PDF

**ABSTRACT**

Select Language  
Translator disclaimer

Spreadsheets are widely used software tools for data entry, storage, analysis, and visualization. Focusing on the data entry and storage aspects, this article offers practical recommendations for organizing spreadsheet data to reduce errors and ease later analyses. The

People also read

Discussion  
50 Years of  
Data

<https://www.tandfonline.com/doi/abs/10.1080/00031305.2017.1375989>

How to join the  
network

Home

About us

Contact us

Services

**Resources**

Links

News & events

Anonymisation Decision-making Framework

Recommended Reading

Case studies

## Resources

**UKAN Resources** are available to everyone. This resource service provides information and documentation on data anonymisation.

It includes UKAN's **Anonymisation Decision-Making Framework book**, which provides a comprehensive guide to doing anonymisation in practice. The book can be accessed under the 'UKAN Decision-making Framework' TAB at the top of the page.

Also included in this service are real world examples of best practice in anonymisation which can be viewed by clicking on Case Studies. Papers, News articles and blogs on anonymisation and other related topics as well as a Glossary of Terms can be accessed under UKAN LIBRARY.

**All the information available from UKAN Resources can be accessed by clicking** on the 'UKAN Decision-making Framework', 'UKAN Library' and 'Case Studies' TABS at the top of the page.

<https://ukanon.net/ukan-resources/>

# Thanks!



Any questions?

Please contact us at:

[researchdata@hku.hk](mailto:researchdata@hku.hk)





# Credits

- Presentation template by [SlidesCarnival](#)
- Photographs by [Unsplash](#)