Seminar on Research Data Management – Prepare a

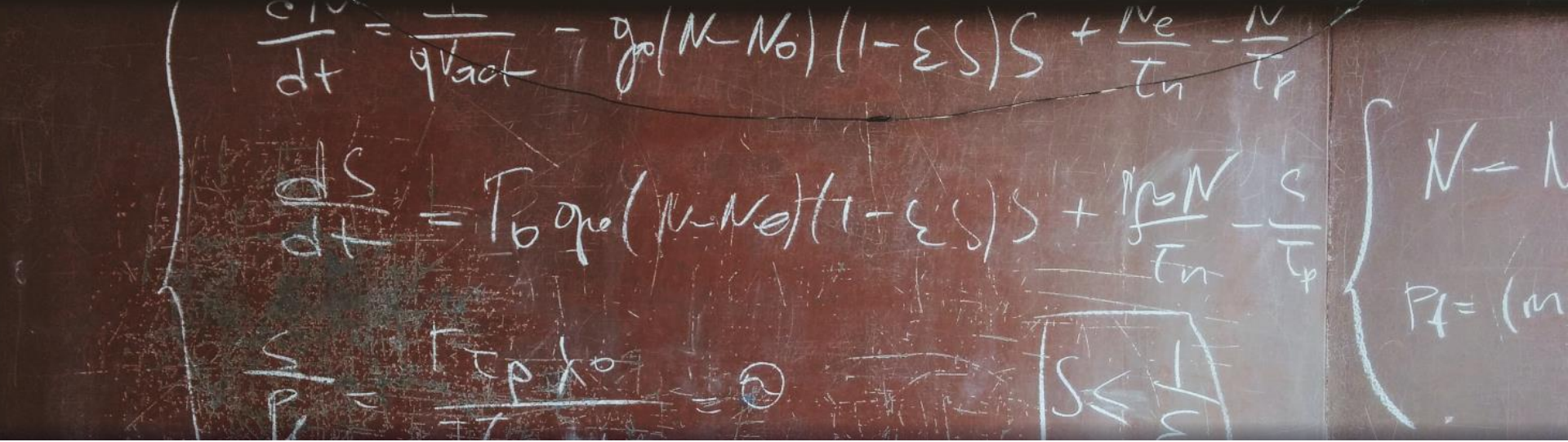
Data Management Plan (DMP)

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



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…


Plan and Design
Collect and Capture
Collaborate and Analyse
Store and Preserve
Share and Publish
Discover, Reuse, Cite
Research data life cycle

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.

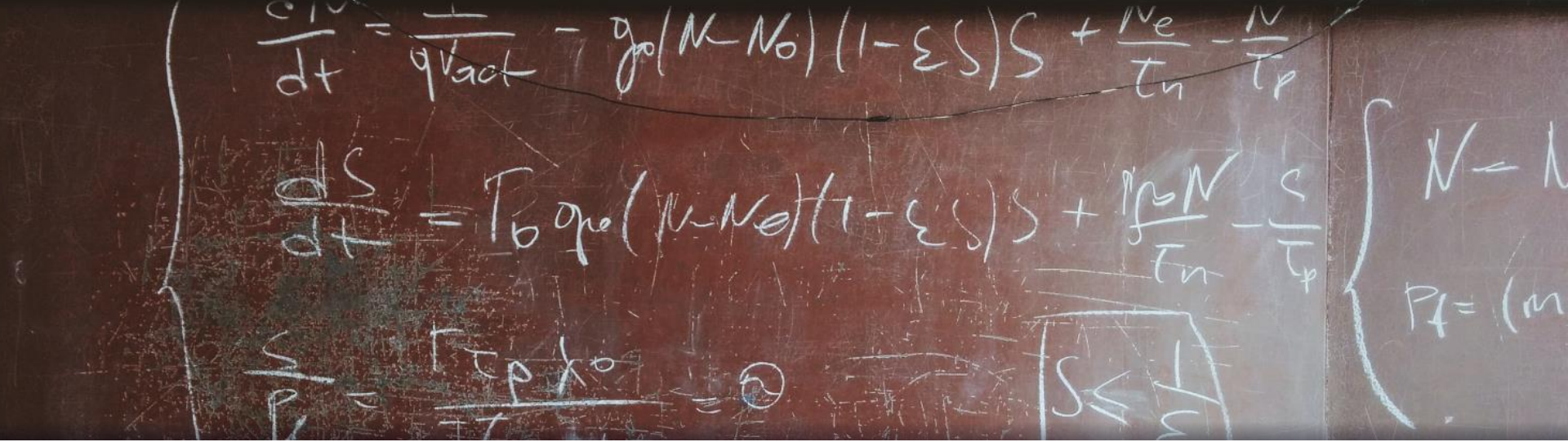
The primary T_TIDE models are all public domain. Published peer-reviewed manuscripts will document the simulations and forcing sufficiently. Recognizing that archiving high-resolution simulations at tidally resolving intervals can result in gigabytes-to-terabytes of data, every effort will be made by 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.

https://library.ucsd.edu/research-and-collections/data-curation/data-management/dmpsample/DMP-Example-Pinkel.pdf

To prepare a DMP

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

Source: JISC



2.
Who has to submit a **DMP**

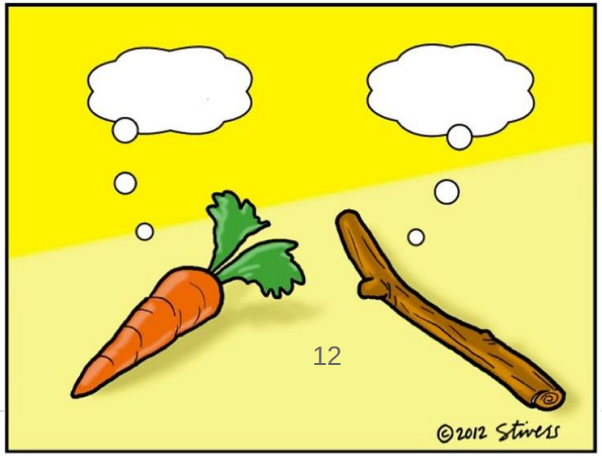
Why research data management?

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

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

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

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

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

HKU Home | Research@HKU | SORTs | Graduate School | Knowledge Exchange Office | Technology Transfer Office

The University of Hong Kong

A+ A A-

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

About Us

Research Funding

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.

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

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

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.

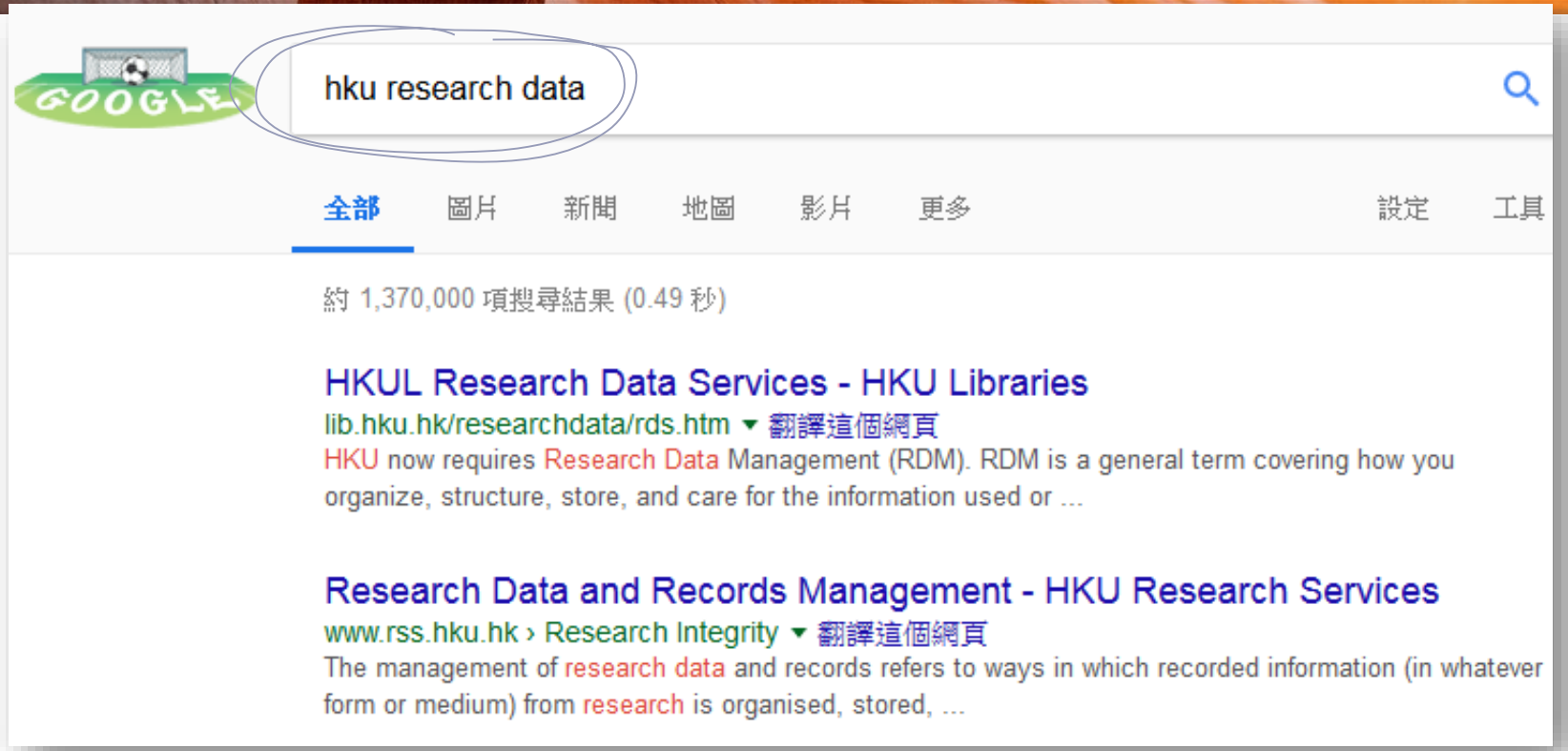
Doctor of Legal Science

Timeline (RPG)

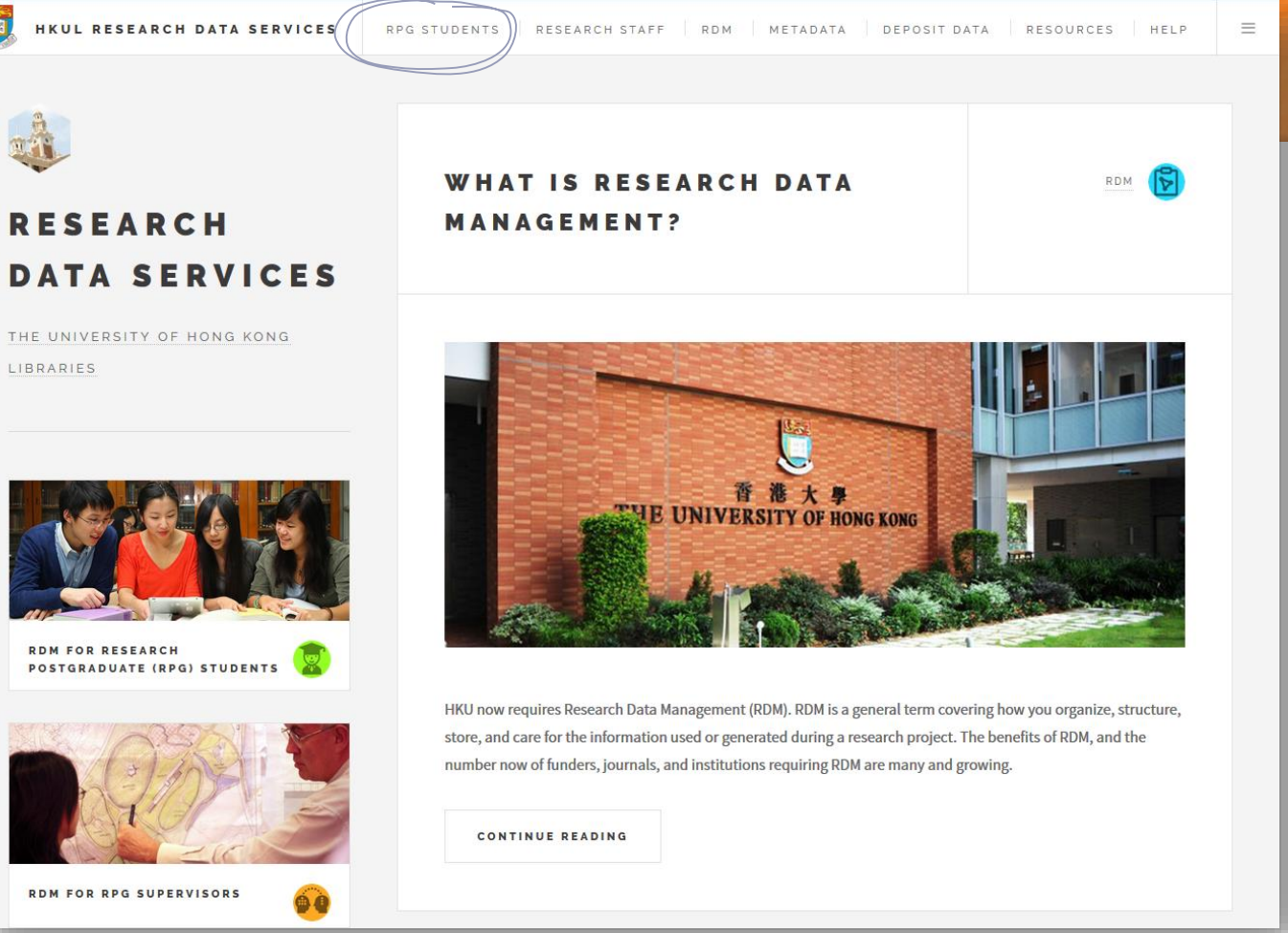
"RPg" includes the degrees of MPhil, PhD, and SJD.

Doctor of Legal Science

HKUL Research Data Services

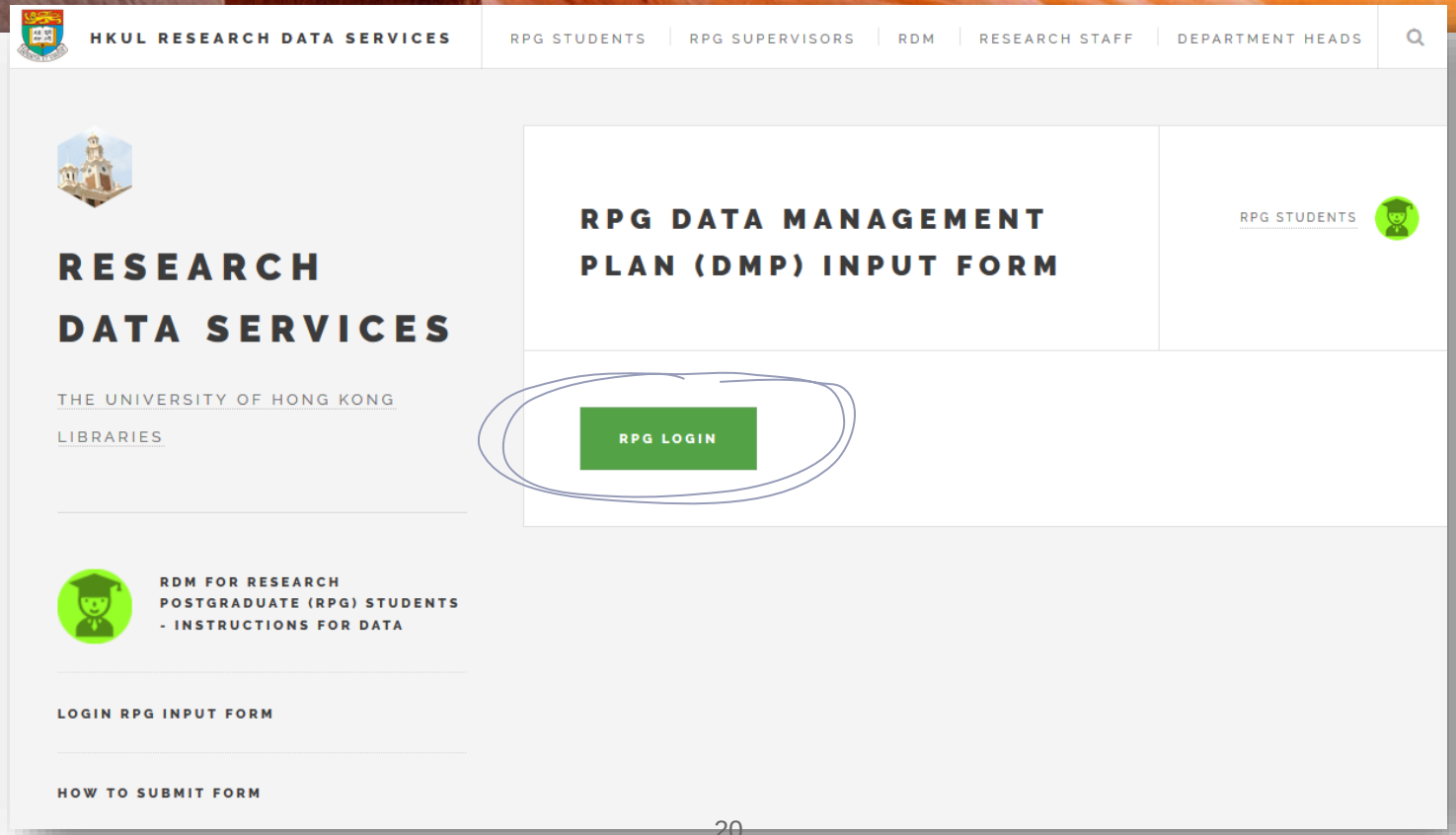


http://lib.hku.hk/researchdata/



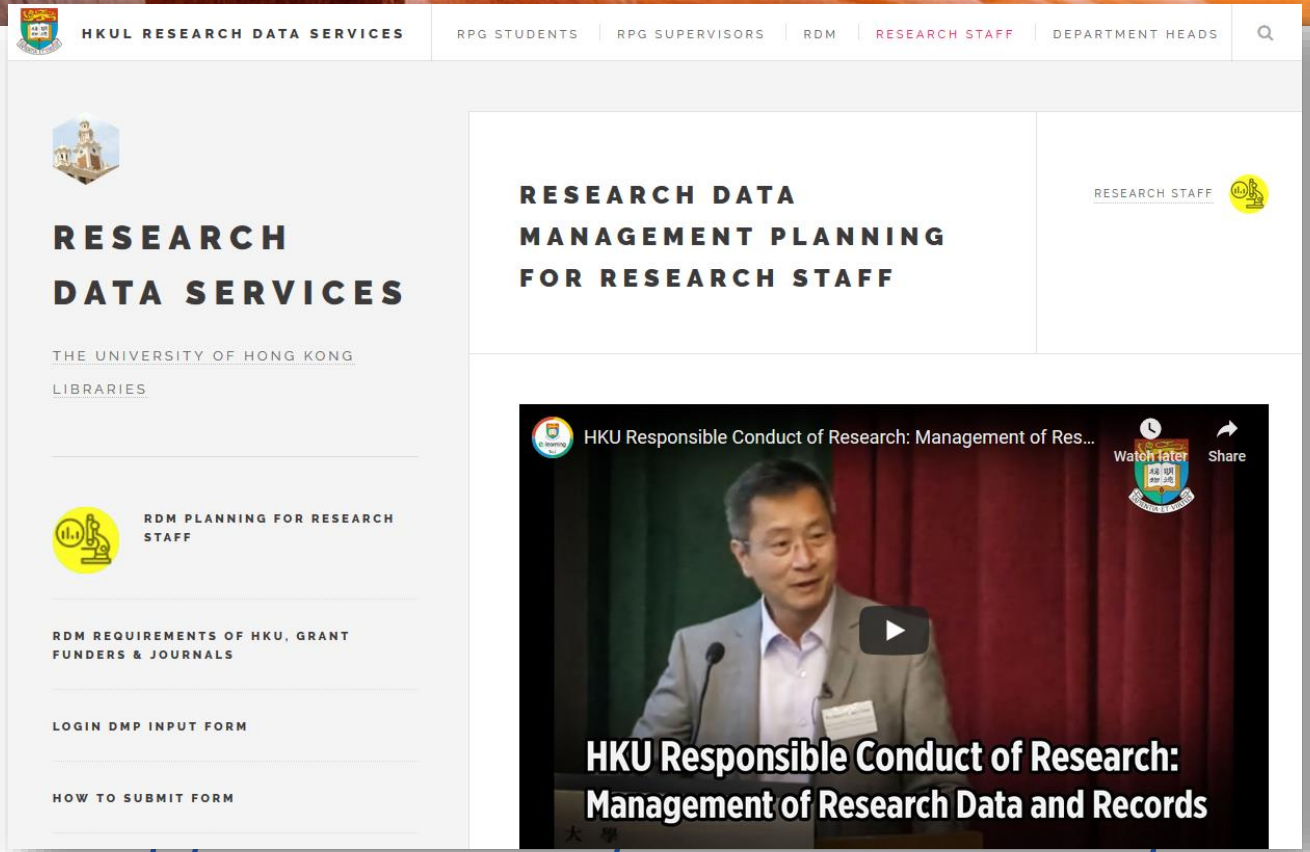
http://lib.hku.hk/researchdata/rds.htm

HKUL Research Data Services



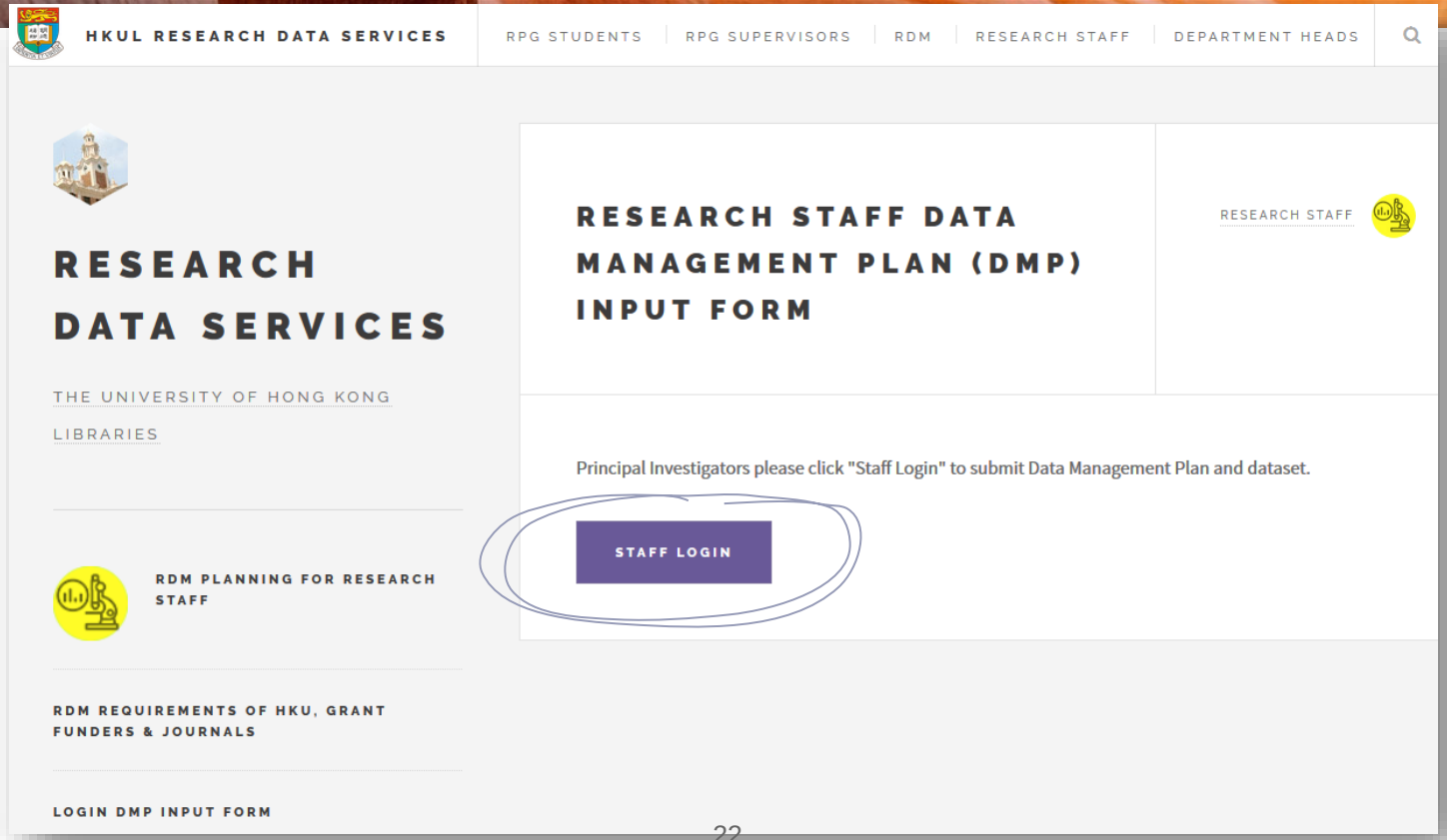
http://lib.hku.hk/researchdata/rpg4.htm

HKUL Research Data Services



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

HKUL Research Data Services

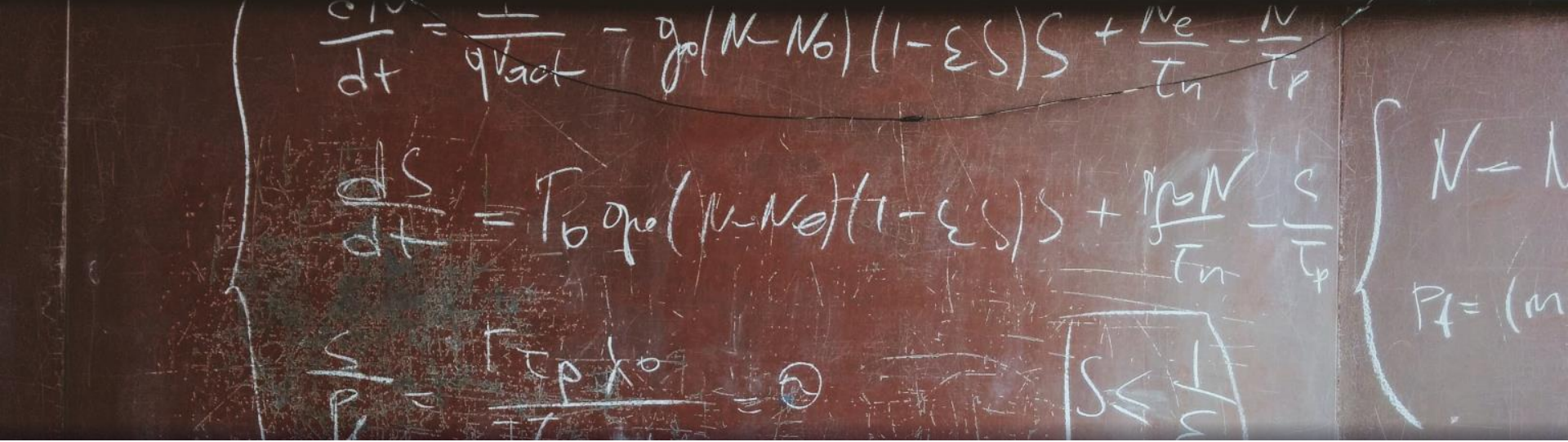


https://hub.hku.hk/researchdata/staff.htm

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.



3.

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

https://dmp.lib.hku.hk

Sign in

HKUL Research Data Services

DMPTool
Build your Data Management Plan

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

DMPTool

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)

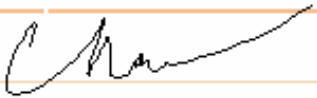
RESPONSIBILITIES AND RESOURCES
Who will be responsible for data management?

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: **Approved by:**

	Postgraduate student		Supervisor
Name:		**Name:**	
Date:	6-11-2017	**Date:**	6-11-2017

Administrative Data (PI)

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

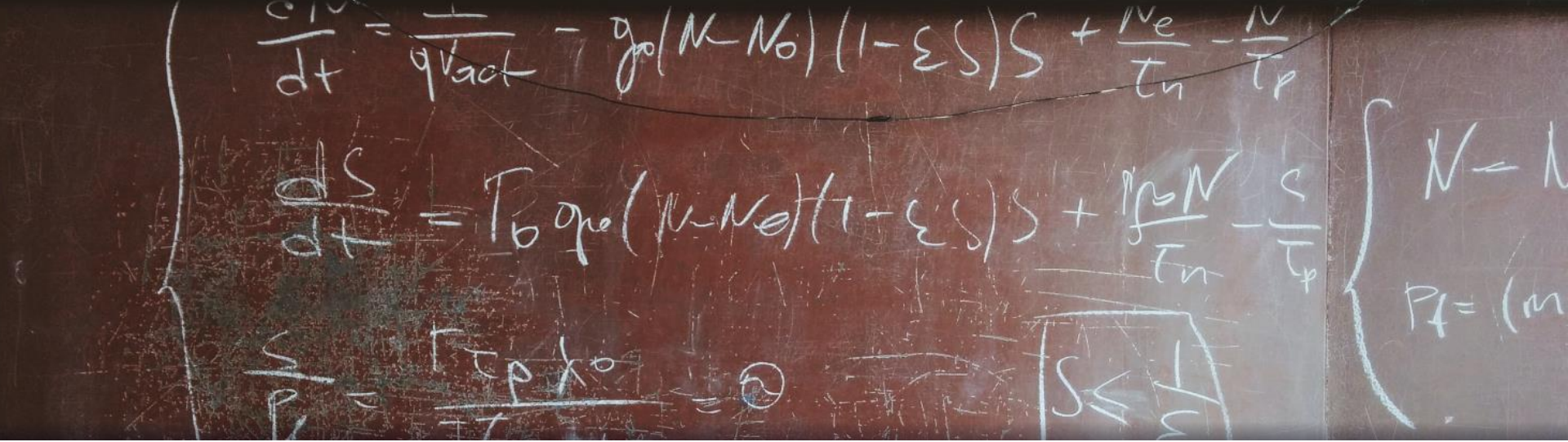
Drag and drop files here, or click in box to choose files.

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



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

Source: DMPTool, Data Curation Centre

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

Source: UvA

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		

File Formats

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

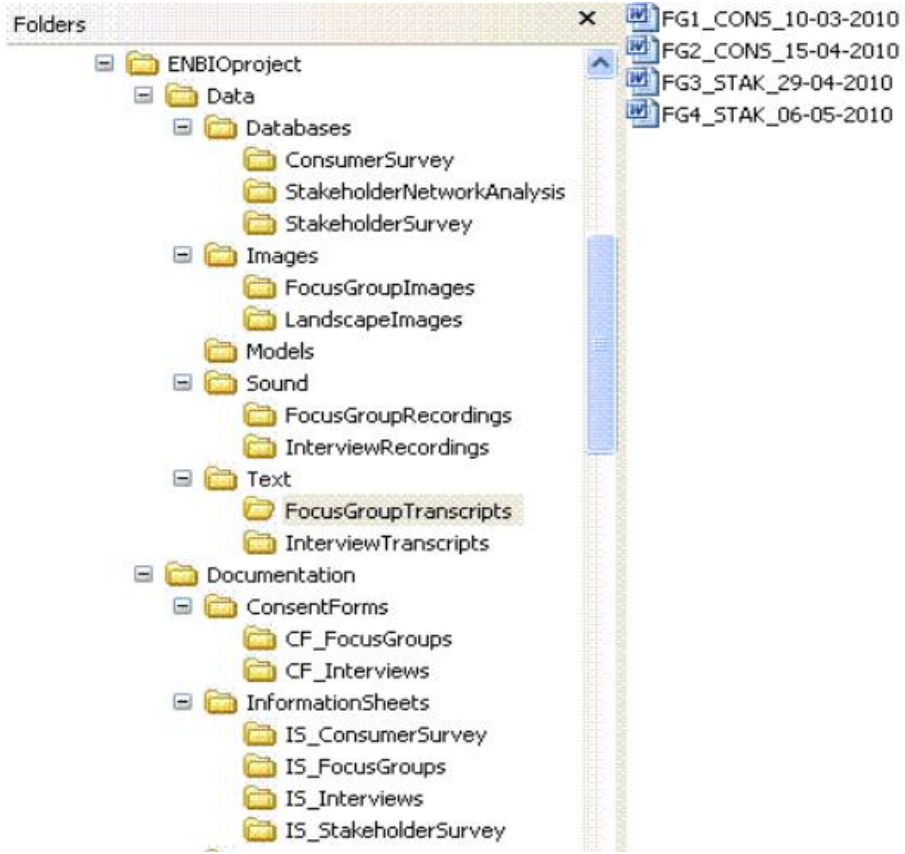
- **Non-proprietary**
- **Open, documented standard**
- **Standard representation (ASCII, Unicode)**
- **Unencrypted**
- **Uncompressed**

Examples of preferred file format choices include:

TXT, or PDF/A,	**not Word**
CSV,	**not Excel**
MPEG-4,	**not Quicktime**
WAV,	**not MP3**
TIFF or JPEG2000,	**not GIF or JPG**

https://libguides.rowan.edu/c.php?g=520327&p=5122622

1. Data Collection



Example folder structure

Meaningful and consistent naming convention for files and folders

https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats

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

https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats

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

https://www.force11.org/fairprinciples

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

https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf

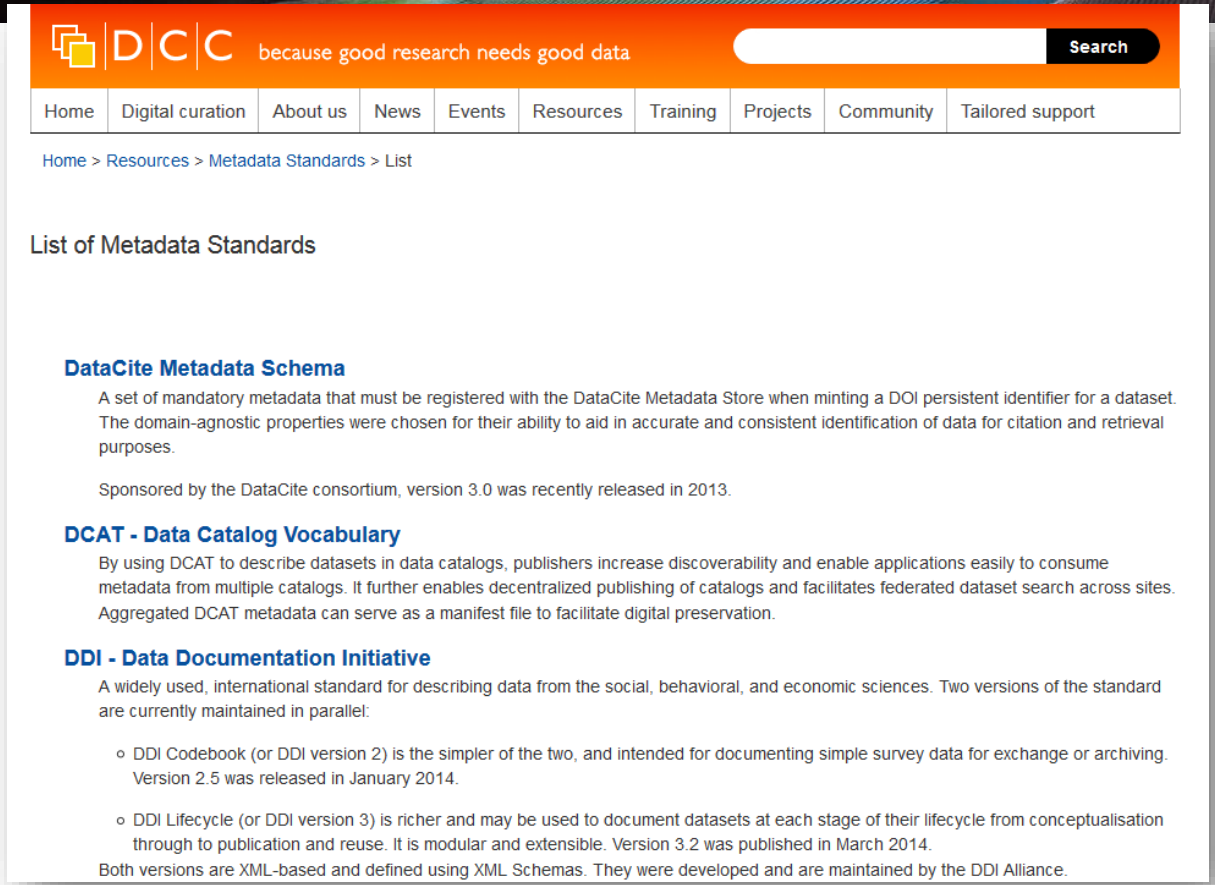
2. Documentation and Metadata

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.

https://libguides.rowan.edu/c.php?g=520327&p=5122622

2. Documentation and Metadata



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

Harvard Biomedical Data Management

Best practices & support services for research data lifecycles

Metadata Overview

"Metadata is structured information th
easier to retrieve, use, or manage an i
about data or information about inform
data was created, analysed and stored
(National Information Standards Organ

Good metadata enables you to understand
helps other researchers discover, access,
facilitates long-term archival preservation

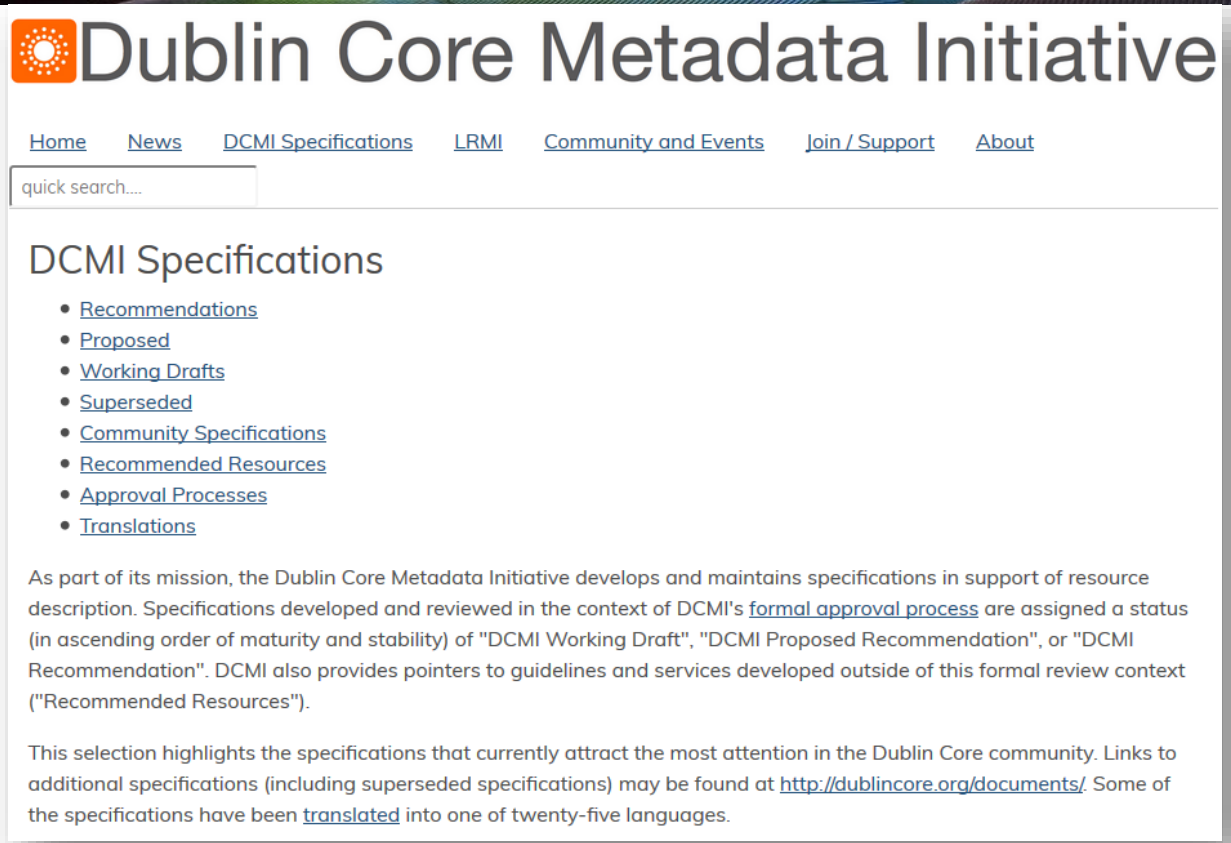
Biomedical metadata may include:

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



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

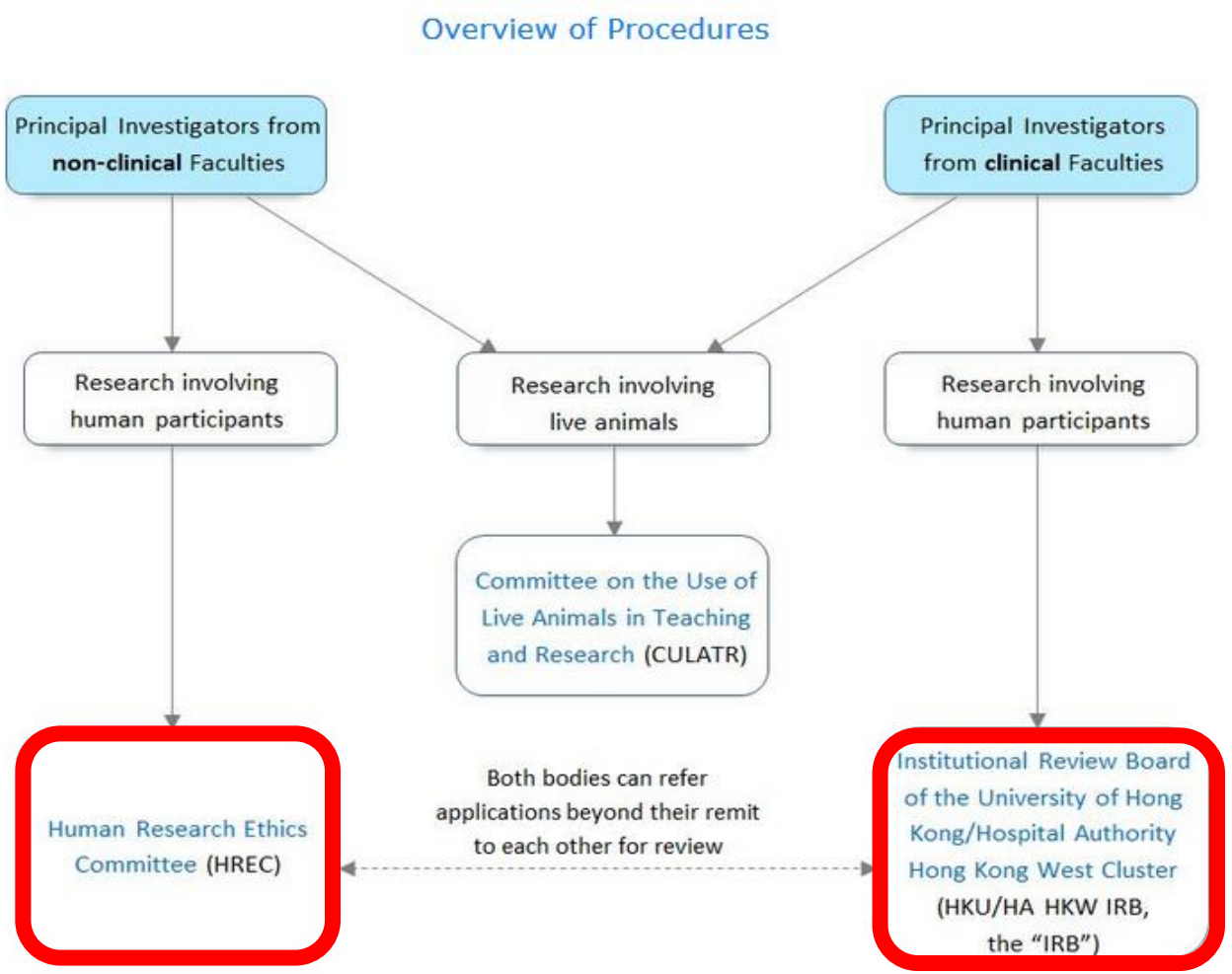
- How will you manage any ethical issues?
- How will you manage copyright and Intellectual Property Rights (IPR) issues?

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?



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

3. Ethics and Legal Compliance

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.

creative commons

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

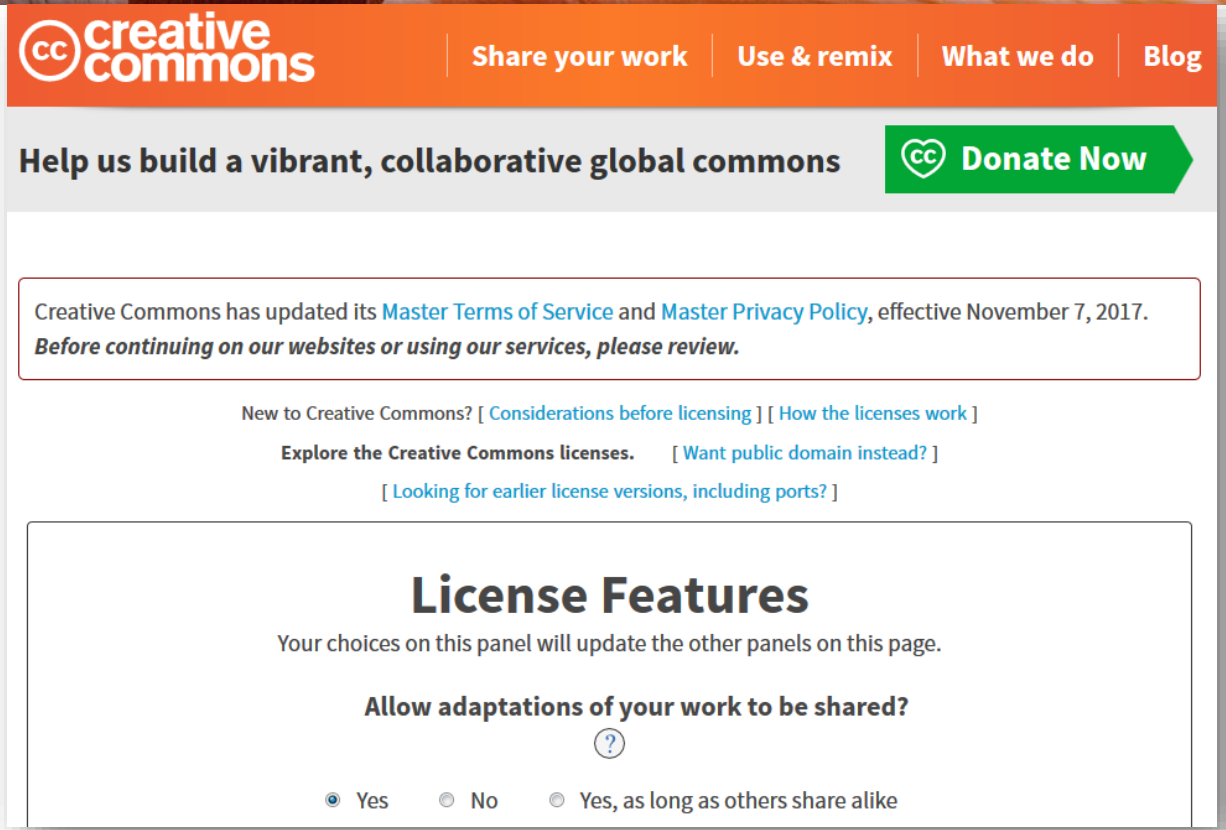
Attribution-NonCommercial
CC BY-NC

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

https://creativecommons.org/licenses/?lang=en

3. Ethics and Legal Compliance



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?

https://library.ucsd.edu/research-and-collections/data-curation/data-management/dmpsample/DMP-Example-Psych.doc

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

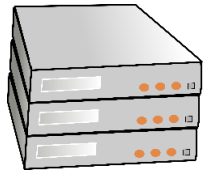
3 copies

2 media

1 offsite

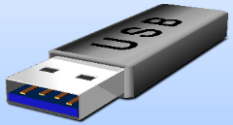
Personal computer

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?

https://www.dataone.org/sites/all/documents/DMP_Copepod_Formatted.pdf

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

Adapted from Whyte, Angus, and Andrew Wilson, 2010. “How to Appraise and Select Research Data for Curation” http://www.dcc.ac.uk/resources/how-guides/appraise-select-data

6. Data Sharing

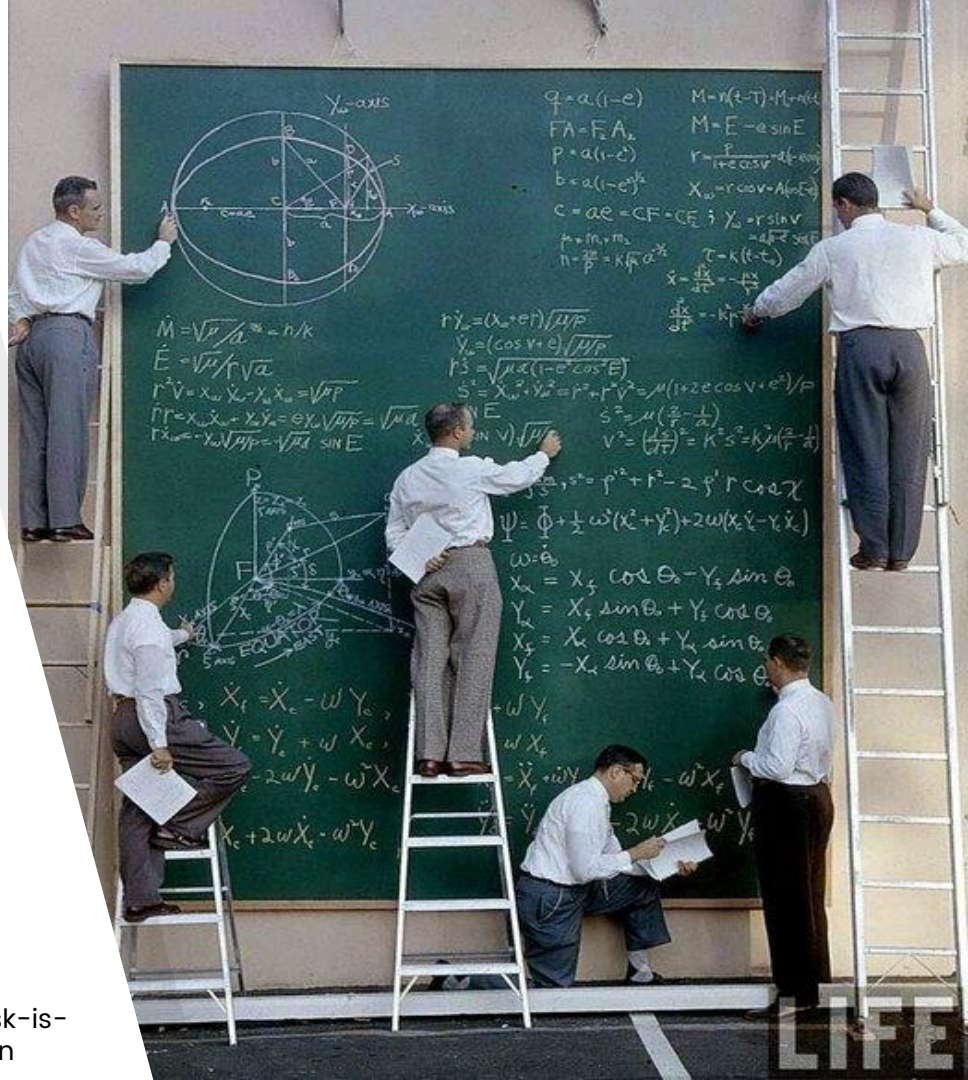


Image courtesy of: https://www.sott.net/article/314865-Cluttered-desk-is-a-sign-of-a-cluttered-mind-of-what-then-is-an-empty-desk-a-sign

6. Data Sharing

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

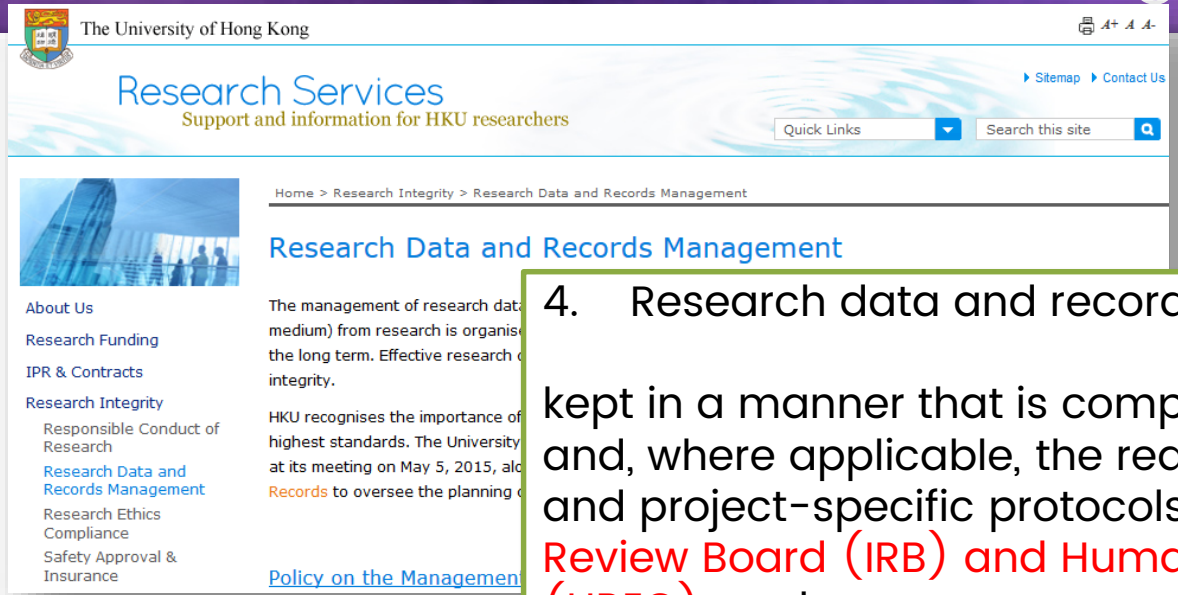
Not a requirement that you share all of your data with anyone who asks

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



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

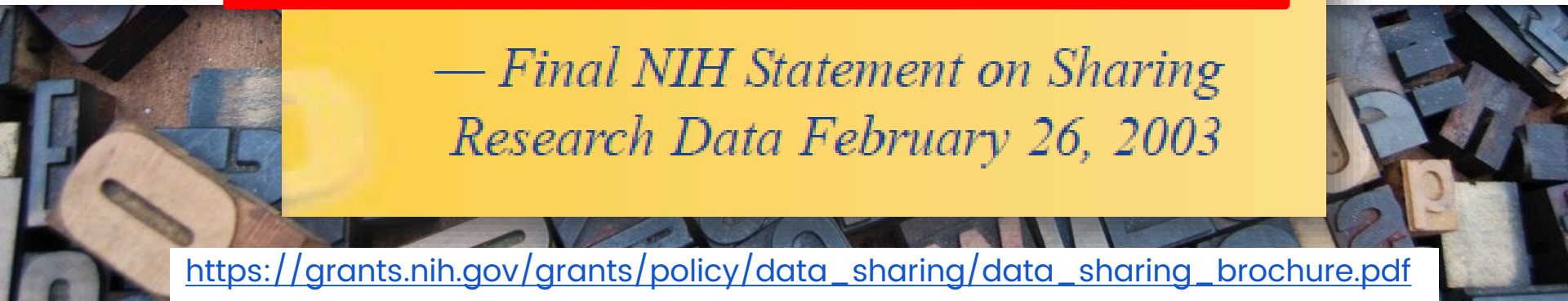
Data supporting a patent

Best practices: UK Data Service, ANDS

“

“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



https://grants.nih.gov/grants/policy/data_sharing/data_sharing_brochure.pdf

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).

http://www.dhcs.ca.gov/dataandstats/data/Pages/ListofHIPAAIdentifiers.aspx

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)

http://www.dhcs.ca.gov/dataandstats/data/Pages/ListofHIPAAIdentifiers.aspx

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

Henderson, M. (2017). *Data management : A practical guide for librarians*

Privacy in Genomics

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/

Issues in Genetics

Coverage and

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/

Sensitive data

Safely sharing sensitive data

Ethics and data sharing

De-identifying your data

Health and medical data

Indigenous data

De-identifying your data

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

https://www.ands.org.au/working-with-data/sensitive-data/de-identifying-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/

Options

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

6. Data Sharing

Institutional repository

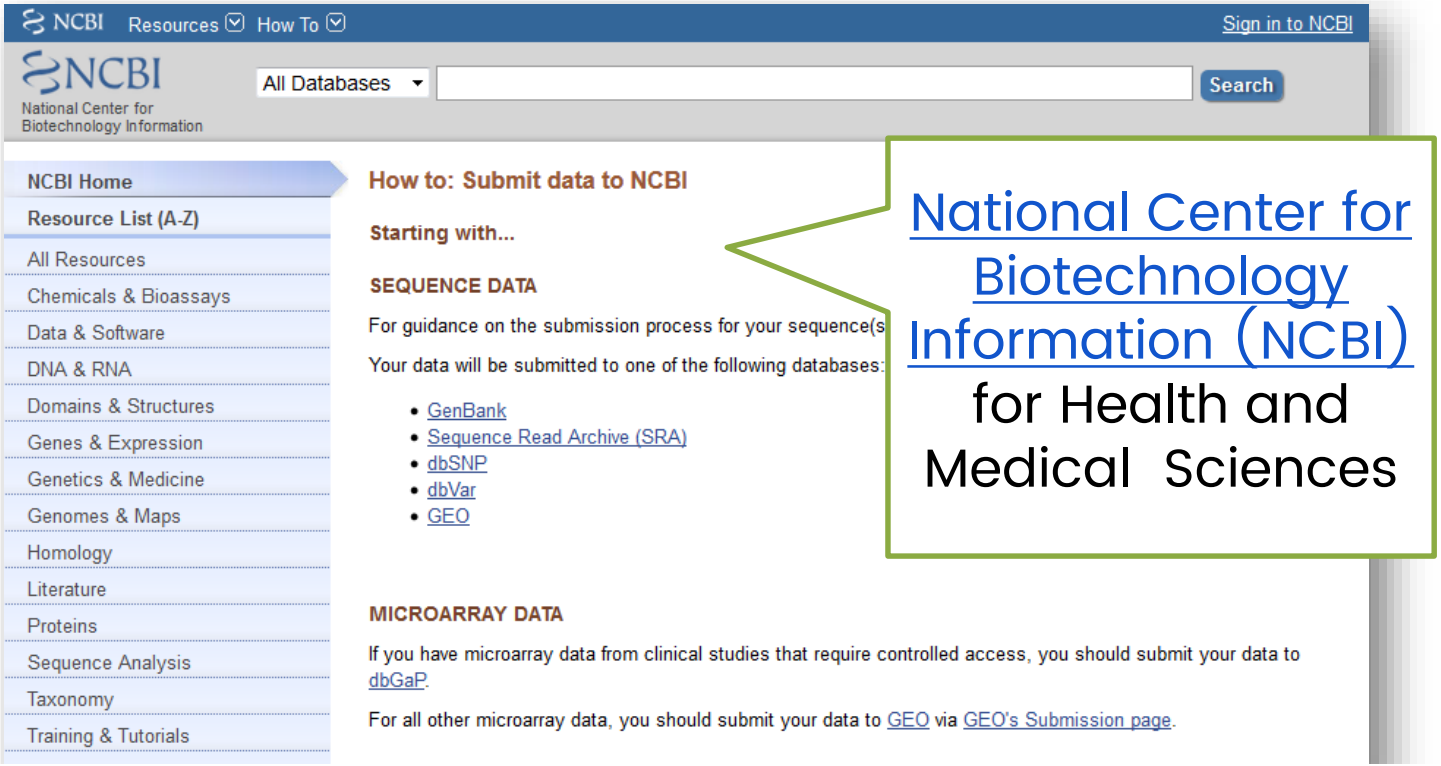
connected to the researcher's institution

Interdisciplinary repository

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

6. Data Sharing

Domain/Subject specific repository

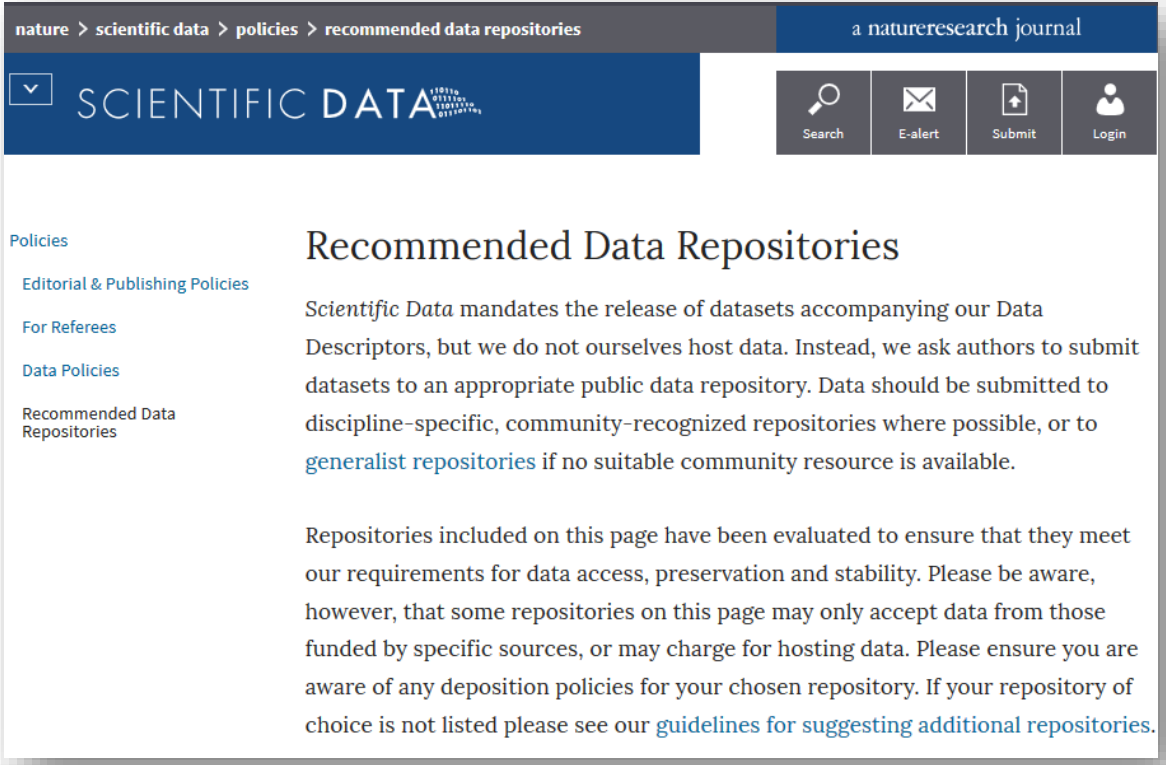


National Center for Biotechnology Information (NCBI) for Health and Medical Sciences

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

6. Data Sharing

Find a Data Repository



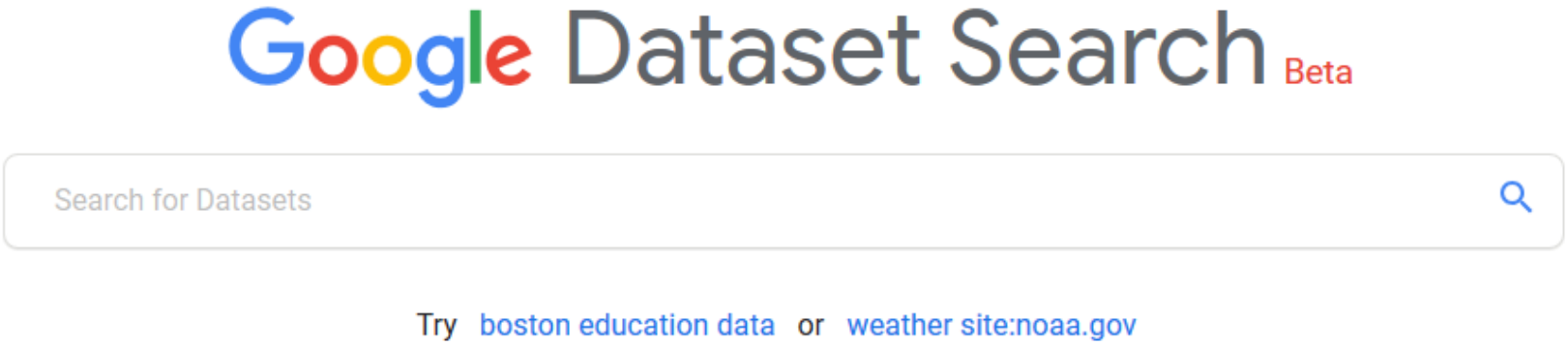
https://www.nature.com/sdata/policies/repositories

6. Data Sharing

Find a Data Repository

https://repositoryfinder.datacite.org/

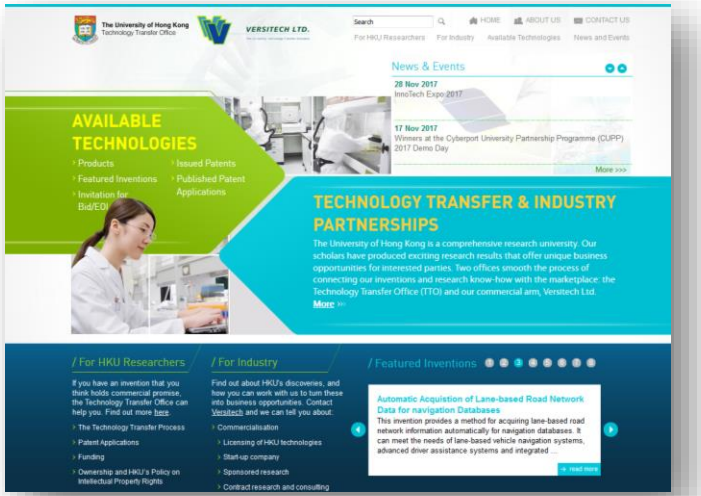
6. Data Sharing



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.



http://www.tto.hku.hk/

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?

https://www.dataone.org/sites/all/documents/DMP_Copepod_Formatted.pdf

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

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

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.

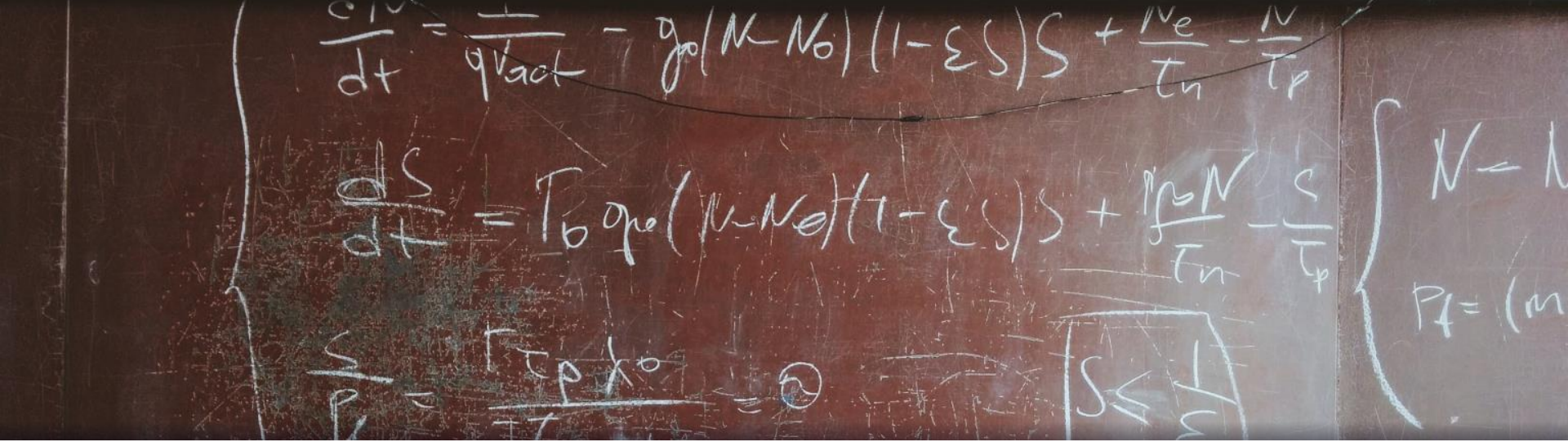
This information will be described in the metadata.

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.

https://www.dataone.org/sites/all/documents/DMP_NutNet_Formatted.pdf



5.
Tools and other Resources

The Best Practices for Biomedical Research Data Management MOCC

canvas NETWORK

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

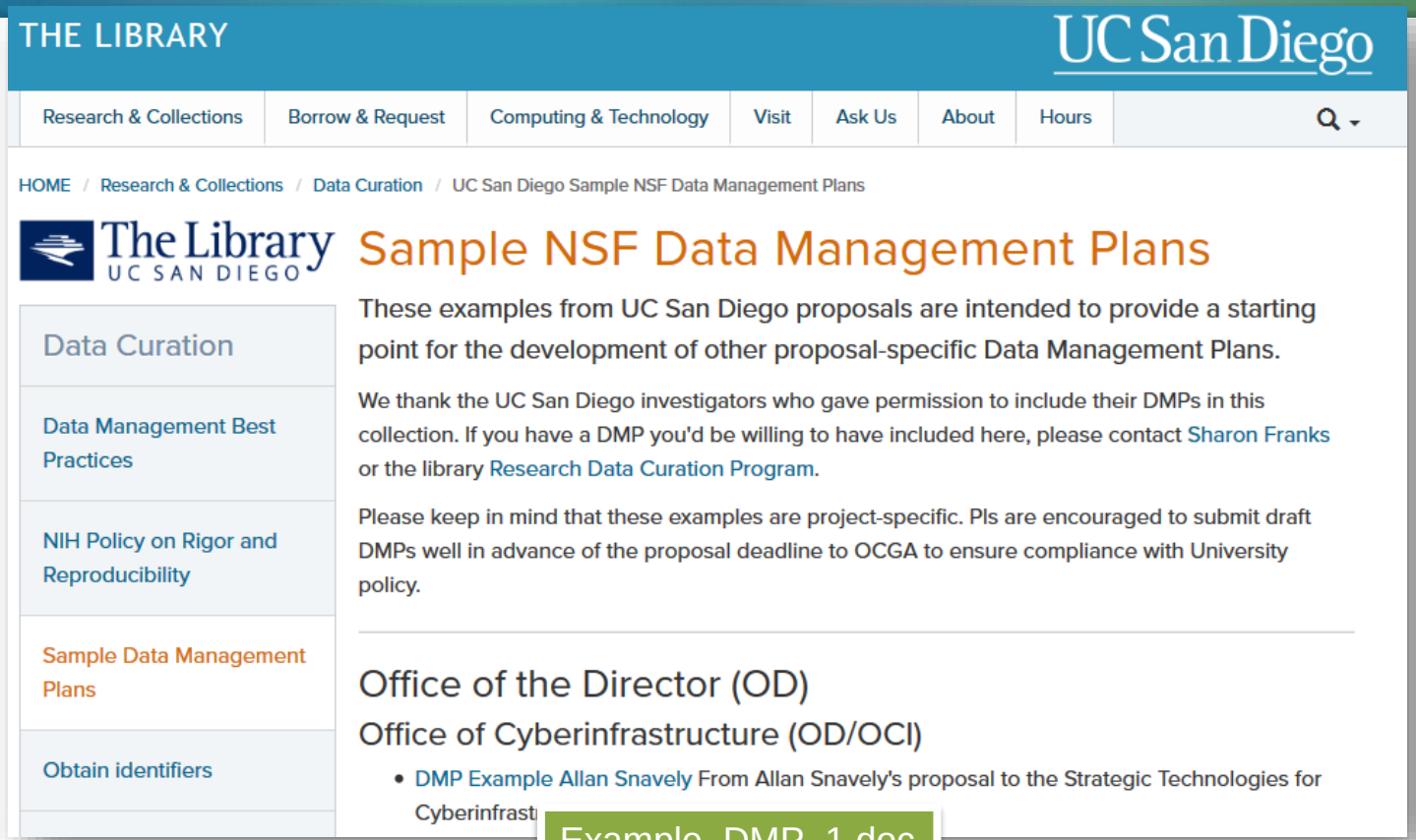
DESCRIPTION

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

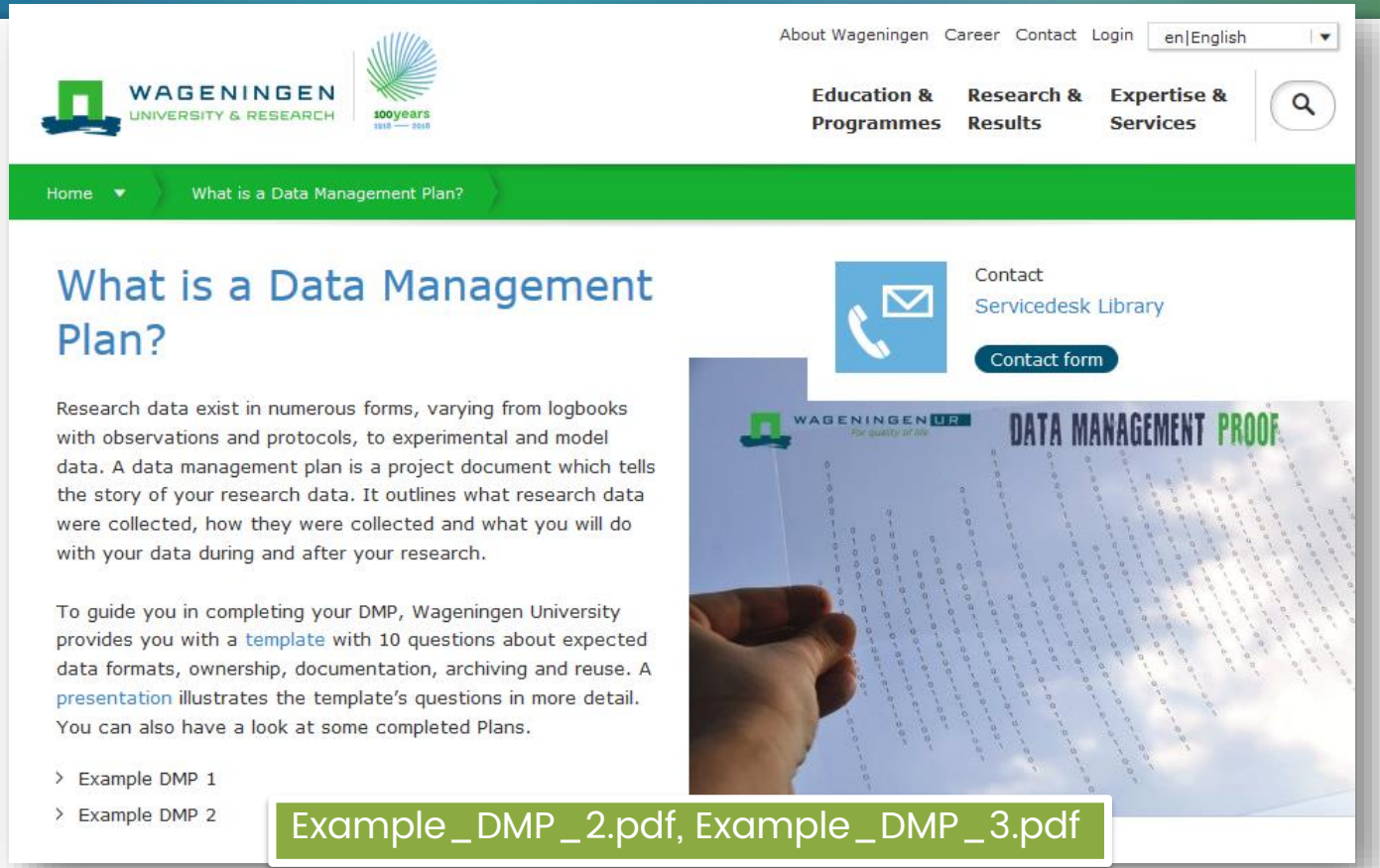
https://www.canvas.net/browse/harvard-medical/courses/biomed-research-data-mgmt

Example DMP



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

Example DMP



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:

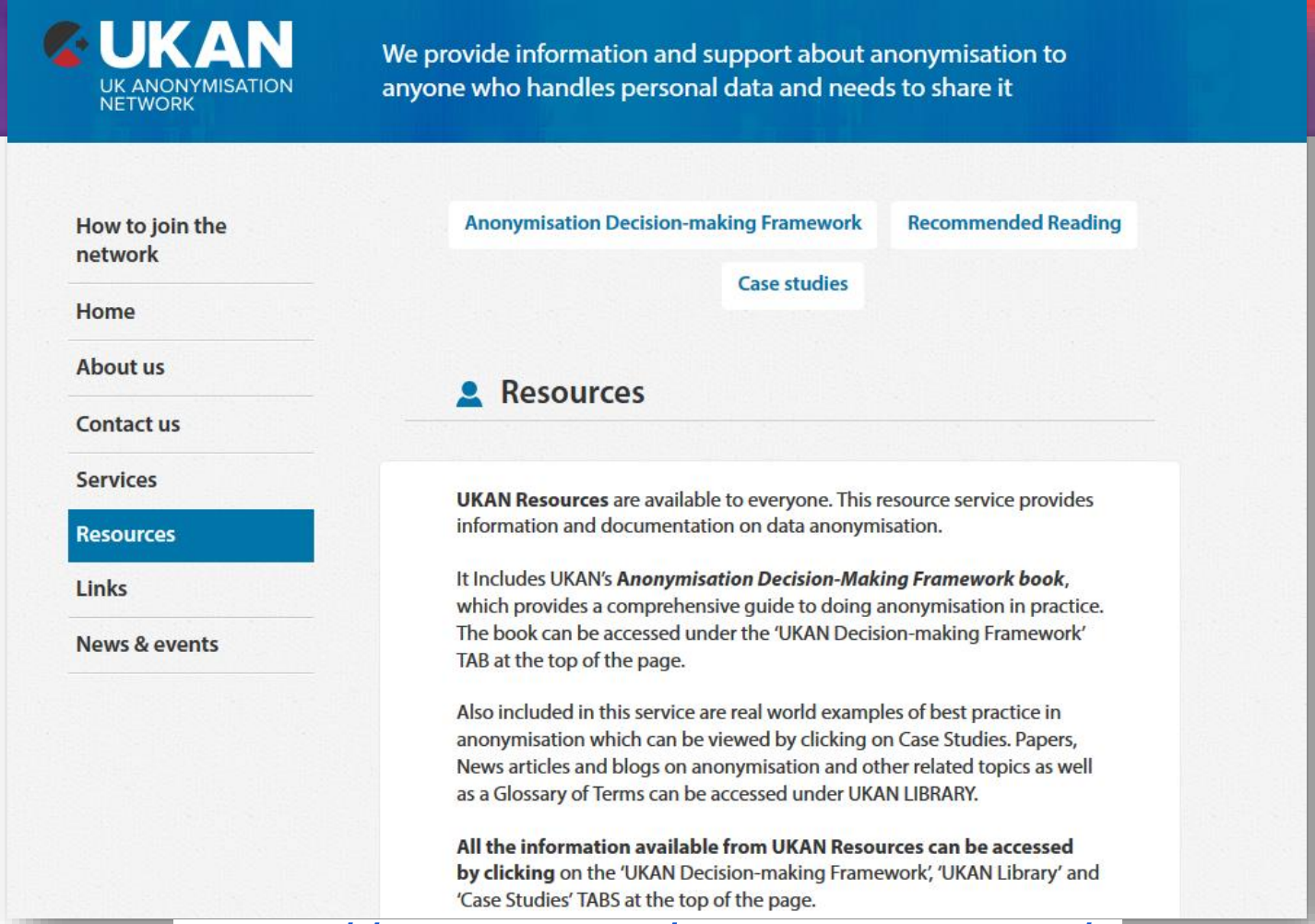
Example_DMP_4.pdf, Example_DMP_5.pdf

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

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



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

Thanks!

Any questions?

Please contact us at:

researchdata@hku.hk

Credits

- Presentation template by SlidesCarnival
- Photographs by Unsplash