RPG procedures for Research Data Management (RDM)
# RDM: Past Workshop Materials

<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
<th>Description</th>
<th>PDF Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>February 2018</td>
<td>RPG Procedures for Research Data Management (RDM) - For Faculties of Business &amp; Economics, Education, Law &amp; Social Sciences</td>
<td>-</td>
<td>PDF of presentation</td>
</tr>
<tr>
<td>February 2018</td>
<td>RPG Procedures for Research Data Management (RDM) - For Faculty of Engineering</td>
<td>-</td>
<td>PDF of presentation</td>
</tr>
<tr>
<td>January 2018</td>
<td>Research Data Management Workshop - Manage Qualitative Research Data Using NVivo 11 for Windows</td>
<td>-</td>
<td>PDF of presentation</td>
</tr>
</tbody>
</table>

[http://lib.hku.hk/researchdata/help.htm](http://lib.hku.hk/researchdata/help.htm)
<table>
<thead>
<tr>
<th>REQUIREMENTS</th>
<th>BENEFITS</th>
</tr>
</thead>
<tbody>
<tr>
<td>✗ Compliance with policies: HKU &amp; funders</td>
<td>✗ Keep research safe and secure</td>
</tr>
<tr>
<td>✗ Ensure data is accessible and shareable: journals requirement</td>
<td>✗ Increase research efficiency</td>
</tr>
<tr>
<td>✗ Demonstrate responsible practice</td>
<td>✗ Improve research integrity</td>
</tr>
<tr>
<td></td>
<td>✗ Make research outputs more visible</td>
</tr>
<tr>
<td></td>
<td>✗ Enable collaboration</td>
</tr>
</tbody>
</table>

Source: JISC
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 project and necessary to verify and defend, when required, the process and outcomes of research. Research data are

http://www.rss.hku.hk/integrity/research-data-records-management
"PI should assess data archive potential and opportunities for data sharing. Due additional weight will be given to an application where the applicants are willing to make research data available to others."

Availability of data, material and methods

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature journal is that authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must also be disclosed in the submitted manuscript.

After publication, readers who encounter refusal by the authors to comply with these policies should contact the chief editor of the journal. In cases where editors are unable to resolve a complaint, the journal may refer the matter to the authors' funding institution and/or publish a formal statement of correction, attached online to the publication, stating that readers have been unable to obtain necessary materials to replicate the findings.

http://www.nature.com/authors/policies/availability.html
After publication, all data and materials necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of Science. All computer codes involved in the creation or analysis of data must also be available to any reader of Science. After publication, all reasonable requests for data or materials must be fulfilled. Any restrictions on the availability of data, codes, or materials, including fees and restrictions on original data obtained from other sources must be disclosed to the editors as must any Material Transfer Agreements (MTAs) pertaining to data or materials used or produced in this research, that place constraints on providing these data or materials. Patents (whether applications or awards to the authors or home institutions) related to the work should also be declared. Fossils or other rare specimens must be deposited in a public museum or repository and available for research. Unreasonable restrictions on data or material availability may preclude publication.
PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.

When submitting a manuscript online, authors must provide a Data Availability Statement describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the final article.

Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection. PLOS journal editors encourage researchers to contact them if they encounter difficulties in obtaining data from articles published in PLOS journals. If restrictions on access to data come to light after publication, we reserve the right to post a correction, to contact the authors' institutions and funders, or in extreme cases to retract the publication.

Methods acceptable to PLOS journals with respect to data sharing are listed below, accompanied by guidance for authors as to what must be indicated in their data availability statement and how to follow best practices in reporting. If authors did not collect data themselves but used another source, this source must be credited as appropriate. Authors who have questions or difficulties with the policy, or readers who have difficulty accessing data, are encouraged to contact the journal office (plosone@plos.org). If you have broader questions about the PLOS data availability policy, contact data@plos.org.
Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Introduction

It is often necessary to genotype biological samples to select individuals from a large population with a desired genetic variant. Genetic variants generated by mutagenesis or natural variation can take the form of single nucleotide polymorphisms (SNPs) or insertions/deletions (indels). Sufficiently variable indels can be distinguished during PCR followed by restriction digestion using a cleaved amplified polymorphic sequence (CAPS) technique. If a restriction site is created or altered by the mutation such that only one allele contains the restriction site, a polymerase chain reaction (PCR) followed by a restriction digest can be used to distinguish the two alleles. However, in the case of most CRISPR-induced alleles, no such restriction sites are present in the target sequences. In this case, a derived CAPS (dCAPS) approach can be used in which mismatches are purposefully introduced in the oligonucleotide primers to create a restriction site in one, but not both, of the amplified templates. Web-based tools exist to aid dCAPS primer design, but when supplied sequences that include indels, the current tools often fail to suggest appropriate primers. Here, we report the development of a Python-based, species-agnostic web tool, called indCAPS, suitable for the design of PCR primers used in dCAPS assays that is compatible with indels. This tool should have wide utility for screening editing events following CRISPR/Cas9 mutagenesis as well as for identifying specific editing events in a pool of CRISPR-mediated mutagenesis events.

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0188406#
Data Availability Statement: The full dataset is available for download at: https://osf.io/5z89v/?view_only=145ae36726f146b5bb5ea6762a0d4f6.

Conclusions
Similar benefits were found when observing the traditional expert model or the error-strewn model, suggesting that viewing poor performance may be as beneficial as viewing expertise in the early acquisition of robotic surgical skills. Further work is required to understand, then
**Data Availability:** The authors confirm that, for approved reasons, some access restrictions apply to the data underlying the findings. Data underlying this study cannot be made publicly available in order to safeguard participant anonymity and that of their organisations. Ethical approval for the project was granted on the basis that only extracts of interviews would be shared (with appropriate anonymisation) as part of publications and other research outputs. In order to share data with other researchers, the participants must be contacted and consent to this data release. Other researchers will be able to request the dataset by contacting the corresponding author or Chair of the University of Sheffield Information School Research Ethics Committee ([ischool_ethics@sheffield.ac.uk](mailto:ischool_ethics@sheffield.ac.uk)).

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.
1. Guidelines and Procedures For 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**.
Guidelines and Procedures

- MPH5 & PHD5 Probation and Confirmation of Candidature – for description of a data management plan (DMP)
- MPH7 & PHD7 Period of Study – for describing when in the period of study, a dataset, where applicable, is to be submitted
- MPH14 & PHD14 Submission of Thesis for Examination – for description of dataset submission
- MPH15 & PHD15 Thesis Examination – for consideration of DMP Entry results and dataset if applicable, and if desired by the examiners

https://www.gradsch.hku.hk/gradsch/current-students/handbooks
Submit a detailed scheme of research +
**Data Management Plan**
(if needed)

Guidelines and Procedures

Begin RPG programme at HKU

Confirmation of Candidature

Submit theses + research data

Thesis submission for examination
2. RPG Input Form
# RPG Data Management Plan (DMP) Input Form

## My Profile

<table>
<thead>
<tr>
<th>Field</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>University No.</td>
<td>3012345678</td>
</tr>
<tr>
<td>Name</td>
<td>Mr. CHAN Tai Man</td>
</tr>
<tr>
<td>Email</td>
<td><a href="mailto:u1234567@hku.hk">u1234567@hku.hk</a></td>
</tr>
<tr>
<td>Degree</td>
<td>Master of Philosophy</td>
</tr>
<tr>
<td>Department/Faculty</td>
<td>School of Biomedical Sciences</td>
</tr>
<tr>
<td>Year of Admission</td>
<td>2017</td>
</tr>
</tbody>
</table>

## Supervisors

<table>
<thead>
<tr>
<th>Role</th>
<th>Supervisor</th>
<th>Date Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Supervisor</td>
<td>Professor Chan Siu Man</td>
<td>(from 2017-09-01 to 2019-08-31)</td>
</tr>
<tr>
<td>Co-Supervisor</td>
<td>Professor Wong Mei Yi</td>
<td>(from 2017-09-01 to 2019-08-31)</td>
</tr>
</tbody>
</table>
**MY PROFILE**

<table>
<thead>
<tr>
<th>University No.</th>
<th>3012345678</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Mr CHAN Tai Man</td>
</tr>
<tr>
<td>Email</td>
<td><a href="mailto:u1234567@hku.hk">u1234567@hku.hk</a></td>
</tr>
<tr>
<td>Degree</td>
<td>Master of Philosophy</td>
</tr>
<tr>
<td>Department/Faculty</td>
<td>School of Biomedical Sciences</td>
</tr>
<tr>
<td>Year of Admission</td>
<td>2017</td>
</tr>
<tr>
<td>Hub RPG URL</td>
<td><a href="http://hub.hku.hk/cris/rpg/rpg123456">http://hub.hku.hk/cris/rpg/rpg123456</a></td>
</tr>
</tbody>
</table>

**SUPERVISORS**

<table>
<thead>
<tr>
<th>Primary Supervisor</th>
<th>Professor Chan Siu Man</th>
<th>(from 2017-09-01 to 2019-08-31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-Supervisor</td>
<td>Professor Wong Mei Yi</td>
<td>(from 2017-09-01 to 2019-08-31)</td>
</tr>
</tbody>
</table>

Please check your supervisor(s) names shown above. If your profile information is correct, click "INFORMATION CORRECT" button below to proceed.

If you found any incorrect information in your profile record, please click the "LOGOUT" button below to exit form and contact your department/faculty to correct your record.
Please check your supervisor(s) names shown above. If your profile information is correct, click "INFORMATION CORRECT" button below to proceed.

If you found any incorrect information in your profile record, please click the "LOGOUT" button below to exit form and contact your department/faculty to correct your record.
Option 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.
YOU HAVE SELECTED OPTION A:

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.

The URL of your dataset retrieved online:
https://figshare.com/articles/_Genetic_Predisposition_to_Increased_Blood

Citation, including address of library or archive used (if not online):

Proitsi, P, Lupton, MK, Velayudhan, L, Newhouse, S, Fogh, I, Tsolaki, M, Daniilidou, M, Pritchard, M, Kloszewska, I, Soininen, H, Meccoci, P, Vellas, B, Williams, J, Stewart, R, Sham, P, Lovestone, S, Powell, JF. (2014). Data from: Genetic Predisposition to Increased Blood Cholesterol and Triglyceride Lipid Levels and Risk of Alzheimer Disease: A Mendelian Randomization Analysis. [Data File]. The authors confirm that, for approved reasons, some access restrictions apply to the data underlying the findings. Data are from the Genetic and Environmental Risk for Alzheimer’s Disease 1 (GERAD1) Consortium, the AddNeuroMed, the Dementia Case Register (DCR) and the Alzheimer’s Disease Neuroimaging Initiative (ADNI) studies. Data access is available on request from the following: GERAD1, williamsj@cardiff.ac.uk. AddNeuroMed, simon.lovestone@psych.ox.ac.uk. DCR data, john.powell@kcl.ac.uk. ADNI data is publicly available (adni.loni.usc.edu).

Approval status:
Approved by Professor Chan Siu Man

Reference No.:
2018011304

Print this page:
Option B

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.
YOU HAVE SELECTED OPTION B:

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.

The name of licensor, seller, or owner of your dataset:

Proitsi, P, Lupton. Data are purchased from the Genetic and Environmental Risk for Alzheimer's Disease 1 (GERAD1) Consortium.

Approval status: Approved by Professor Chan Siu Man

Reference No.: 2018011304
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.
YOU HAVE SELECTED OPTION C:

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.

Approval status: Approved by Professor Chan Siu Man
Reference No.: 2018011304
Option D

Submit Data Management Plan (DMP). Dataset will be uploaded later.
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.
YOU HAVE SELECTED OPTION D:

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

DATA MANAGEMENT PLAN SUBMISSION STATUS

<table>
<thead>
<tr>
<th>Version 1</th>
<th>DMP Submitted on 20180113 15:04:34</th>
<th>Download DMP File</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approval status:</td>
<td>Approved by Professor Chan Ying Shing</td>
<td></td>
</tr>
<tr>
<td>Reference No.:</td>
<td>2018011304</td>
<td></td>
</tr>
</tbody>
</table>

If you would like to further update your DMP, you may revisit this page to upload and replace your previous DMP by clicking "Update DMP" button below.

You have not yet submitted your DATASET, click "Submit Dataset" to submit!
Online tool for creating a Data Management Plan (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-uv profiles from the McLean 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 IWSE Experiment, 2010 1st, 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-CTD site studies

The centralized data access will be maintained for the Process Experiment, with the McLean 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.

Welcome.

DMPRoadmap has been jointly developed by the Curation Center to help you write data management plans.

Screencast on how to use DMPRoadmap

Sign in

[Form fields for email address and password]

Forgot your password?

[Options to remember me]

Sign in

Or, sign in with your institutional credentials (UK users only)

Create account

New to DMPRoadmap? Create an account today.
Please fill in the basic project details below

<table>
<thead>
<tr>
<th>Plan name</th>
<th>Thesis Title 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>33333333333</td>
</tr>
<tr>
<td>Grant number</td>
<td></td>
</tr>
<tr>
<td>Principal Investigator/Researcher</td>
<td>Chan Tai Man</td>
</tr>
<tr>
<td>Principal Investigator/Researcher ID</td>
<td>Dr Chan Bo Hung</td>
</tr>
<tr>
<td>Plan data contact</td>
<td><a href="mailto:chanbohung@hku.hk">chanbohung@hku.hk</a></td>
</tr>
</tbody>
</table>

**Biology**

**Description**

[Save] [Cancel]
<table>
<thead>
<tr>
<th>Sections</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Collection</td>
<td>What data will you collect or create?</td>
</tr>
<tr>
<td></td>
<td>How will the data be collected or created?</td>
</tr>
<tr>
<td>Documentation and Metadata</td>
<td>What documentation and metadata will accompany the data?</td>
</tr>
<tr>
<td>Ethics and Legal Compliance</td>
<td>How will you manage any ethical issues?</td>
</tr>
<tr>
<td></td>
<td>How will you manage copyright and Intellectual Property Rights (IPR) issues?</td>
</tr>
<tr>
<td>Storage and Backup</td>
<td>How will the data be stored and backed up during the research?</td>
</tr>
<tr>
<td></td>
<td>How will you manage access and security?</td>
</tr>
<tr>
<td>Selection and Preservation</td>
<td>Which data are of long-term value and should be retained, shared, and/or preserved?</td>
</tr>
<tr>
<td></td>
<td>What is the long-term preservation plan for the dataset?</td>
</tr>
<tr>
<td>Data Sharing</td>
<td>How will you share the data?</td>
</tr>
<tr>
<td></td>
<td>Are any restrictions on data sharing required?</td>
</tr>
<tr>
<td>Responsibilities and Resources</td>
<td>Who will be responsible for data management?</td>
</tr>
<tr>
<td></td>
<td>What resources will you require to deliver your plan?</td>
</tr>
</tbody>
</table>
What data will you collect or create?

This project will generate three main types of raw data.
1. Images from transmitted-light microscopy of Giemsa-stained squashed larval brains.
2. Images from confocal microscopy of immunostained whole-mounted larval brains.
3. Western blot data.

Answered 5 minutes ago by eunice98@hku.hk

How will the data be collected or created?

Measurements and quantification of the images will then be recorded in spreadsheets.
Micrograph data is expected to total between 100GB and 1TB over the course of the project.
Scanned images of western blots are expected to total around 1GB over the course of the project.
Other derived data (measurements and quantifications) are not expected to exceed 10MB.
THESIS TITLE 1

MY CURATION CENTER’S DEFAULT TEMPLATE

ADMIN DETAILS
Plan Name: My Curation Center’s Default Template
Plan ID: 3333333333
Grant number: -
Principal Investigator / Researcher: Chan Tai Man
Plan Data Contact: chanbohung@hku.hk
Plan Description: Biology
Funder: -
Institution: HKU
Your ORCID: -

DATA COLLECTION
What data will you collect or create?

This project will generate three main types of raw data.
1. Images from transmitted-light microscopy of giemsa-stained squashed larval brains.
2. Images from confocal microscopy of immunostained whole-mounted larval brains.
3. Western blot data.

How will the data be collected or created?

Measurements and quantification of the images will then be recorded in spreadsheets. Micrograph data is expected to total between 100GB and 1TB over the course of the project. Scanned images of western blots are expected to total around 1GB over the course of the project. Other derived data (measurements and quantifications) are not expected to exceed 10MB.

DOCUMENTATION AND METADATA
What documentation and metadata will accompany the data?

All samples on which data are collected will be prepared according to published standard protocols in the field. All microscopes used for sample examination are serviced and recalibrated regularly. All Drosophila lines used in experiments are checked periodically for phenotypic markers. Drosophila are maintained in live culture according to standard methods in the field.
Files will be named according to a pre-agreed convention. The dataset will be accompanied by a README file which will describe the directory hierarchy and filenaming convention. Each directory will contain an INFO.txt file describing the experimental protocol used in that experiment. It will also record any deviations from the protocol and other useful contextual information. Microscope images capture and store a range of metadata (field size, magnification, lens phase, zoom, gain, pinhole diameter etc) with each image. This should allow the data to be understood by other members of our research group and add contextual value to the dataset should it be reused in the future.
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

<table>
<thead>
<tr>
<th>Prepared by:</th>
<th>Approved by:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postgraduate student</td>
<td>Supervisor</td>
</tr>
<tr>
<td><strong>Name:</strong></td>
<td><strong>Name:</strong></td>
</tr>
<tr>
<td><strong>Date:</strong> 6-11-2017</td>
<td>Date:</td>
</tr>
</tbody>
</table>
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.
4. Examples
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 model process, and collaborative efforts, such as the ONR IWIS experiment in the Sargasso Sea. To guide both modeling and the Process Experiment, the data will be centralized on a server at APL. Scout Quick-look data responsibilities:

- J.Klymak: LADCP-CTD analysis
- S. Johnston: SIO glider analysis
- L.Rainville: Co-operative Ocean Data Program (CORDP)
- 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.

2. Documentation and Metadata

What documentation and metadata will accompany the data?
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?
Intellectual Property Rights (IPR)

- Intellectual property rights (IPR) include copyright, patents, trademarks and design rights.
- The HKU Intellectual Property Rights Policy sets out ownership of intellectual property rights of staff and students and explains the University position on both copyright and patents.

http://www.rss.hku.hk/contracts/ipr
"If research data included in a thesis are obtained by a collaborative effort (including collaboration between the Student and a supervisor or other researcher at the University), such data may be the joint property of the Student and the collaborating party.

It is strongly advised that Students and supervisors/researchers make clear agreements in advance concerning the ownership and use of Intellectual Property Rights created in connection with a Student thesis."

http://www.rss.hku.hk/contracts/ipr
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. **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.

---

**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?
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
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 Versitech, or the Technology Transfer Office.
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?

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.
5. DATA SUBMISSION
**WHAT TO DEPOSIT**

**Essential,**

- Data Management Plan (DMP)
- Dataset(s) **quantitative** and/or **qualitative**, **raw** and/or **processed**,
- **Metadata** about all the data files including file formats (please use open formats wherever possible), Code book (i.e. description of variables), etc.
- **Readme file**, giving particulars of data

https://lib.hku.hk/researchdata/data.htm
If data includes **personal data**, the data should be put under **restricted access**, 

- **Personal data from clinical research** (i.e. Institutional Review Board (IRB) approved) [https://www.med.hku.hk/research/research-ethics/human-ethics]  
  - provide approval code, consent forms, ethical application form when available, please state the risk of re-identification from the different datafiles and how the risk has been minimised for any dataset intended for sharing.

- **Personal data from non-clinical research** (i.e. Human Research Ethics Committee (HREC) approved) [http://www.rss.hku.hk/integrity/ethics-compliance/hrec]  
  - provide **approval code, consent forms, ethical application form**, please state the risk of re-identification from the different datafiles and how the risk has been minimised for any dataset intended for sharing.

https://lib.hku.hk/researchdata/data.htm
What to deposit

If data includes interviews,
• Interview transcripts
• Blank questionnaire & interviewer guidelines

If field research data,
• provide copy of file research notebook in digital format, preferably machine readable

If lab research data,
• of working papers and/or lab research notebooks in digital format, preferably machine readable

https://lib.hku.hk/researchdata/data.htm
What to Deposit

For **simulated data**,  
• how was it generated? Please either explain or provide a link.

For **other types** of data, such as **Image or video data, Creative or Design data**,  
• please explain what type of data and how was it collected/generated.

If **software** is needed to read or analyze any of the datafiles,  
• please provide full details of software name, version needed, and any instructions necessary to obtain the software. **If you have written your own script for analyzing the data, please include this script also in final deposit.**

https://lib.hku.hk/researchdata/data.htm
Example DMPs and guidance

**UCSD Example Data Management Plans**
Over 20 example plans submitted to the National Science Foundation (NSF) in the United States by academics at UC San Diego

**Colorado School of Mines examples**
A variety of US example DMPs from Mines and elsewhere

**NSF data management plans**
5 DMPs submitted to the NSF, shared by the DataOne initiative

**Biology and chemistry DMPs**
Three example DMPs from the USA shared by NECDMC, an instructional tool for teaching RDM to undergraduates, graduate students, and researchers in the health sciences, sciences and engineering.
UC San Diego Sample NSF Data Management Plans

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 Snively From Allan Snively's proposal to the Strategic Technologies for Cyberinfrastructure (STCI) program.

Office of Integrative Activities (OD/OIA)
DMP Example Todd Martz SIO.ppt From Professor Todd Martz's proposal entitled "MRI Development of an instrument for testing and calibration of autonomous sensors for..."
Arts and Humanities

**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 conceptualization 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.

**MIDAS-Heritage**  
A British cultural heritage standard for recording information on buildings, archaeological sites, shipwrecks, parks and gardens, battlefields, areas of interest and artefacts.

Sponsored by the Forum on Information Standards in Heritage, MIDAS Version 1.1 was released in October 2012.

**OAI-ORE (Open Archives Initiative Object Reuse and Exchange)**  
The goal of these standards is to expose the rich content in aggregations of Web resources to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. The standards support the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship with flexible tools, shared data, and services across all web-based...
SUBJECT-SPECIFIC METADATA AND TOOLS

Research Data Alliance

Social and Behavioral Sciences

**CARARE metadata schema** [Edit]
An application profile of the MIDAS Heritage standard intended for delivering metadata to the CARARE service environment about an organisation's online collections, monument inventory database and digital objects.

**CESSDA MLI - Council of European Social Science Data Archives Minimum Level of Information** [Edit]
A common base profile of DDI for use by the member archives of CESSDA.

**GESMES/TS (Generic Statistical Message for Time Series)** [Edit]
An extension of SDMX used to exchange statistical data and metadata.

General Research Data

**AGLS Metadata Profile** [Edit]
An application of Dublin Core designed to improve visibility and availability of online resources, originally adapted from the Australian Government Locator Service metadata standard for use in government agencies.

**Asset Description Metadata Schema (ADMS)** [Edit]
Used to describe semantic assets, defined as highly reusable metadata (for example: XML schema, generic data models) and reference data (for example: code lists, taxonomies, dictionaries, vocabularies) that are used for eGovernment system development.

**Dryad Metadata Application Profile** [Edit]
An application profile based on the Dublin Core Metadata Initiative Abstract Model, used to describe multi-disciplinary data underlying peer-reviewed scientific and medical literature.

**GSIM (Generic Statistical Information Model)** [Edit]
A reference framework that provides a common terminology across and between statistical organisations; aligns with DDI and SDMX.

**OpenAIRE Guidelines for publication repositories, data archives and CRIS systems** [Edit]
The OpenAIRE Guidelines are a suite of application profiles designed to allow research institutions to make their scholarly outputs visible through the OpenAIRE infrastructure. The profiles are based on established standards and designed to be used in conjunction with the OAI-PMH metadata harvesting.

Recommended Repositories

PLOS requires that authors comply with field-specific standards for preparation and recording of data and select repositories appropriate to their field, for example deposition of microarray data in ArrayExpress or GEO; deposition of gene sequences in GenBank, EMBL or DDBJ; and deposition of ecological data in Dryad. Authors are encouraged to select repositories that meet accepted criteria as trustworthy digital repositories.

PLOS has identified a set of established repositories below, which are recognized and trusted within their respective communities. For further information on environmental and biomedical science repositories and field standards, we suggest utilizing FAIRsharing; we have also created a FAIRsharing page of PLOS recommended data repositories. Additionally, the Registry of Research Data Repositories (Re3Data) is a full scale resource of registered repositories across subject areas. Both FAIRsharing and Re3Data provide information on an array of criteria to help researchers identify the repositories most suitable for their needs (licensing, certificates and standards, policy, etc.).

Authors are encouraged to select the repository most appropriate for their research. PLOS does not dictate repository selection for the data access policy. If authors use repositories with stated licensing policies, the policies should not be more restrictive than the Creative Commons Attribution (CC BY) license. More information about the content license can be found in our licenses and copyright policy.

If no specialized community-endorsed open repository exists, institutional repositories that use open licenses permitting free and unrestricted use or public domain, and that adhere to best practices pertaining to responsible data sharing, sustainable digital preservation, proper citation, and openness are also suitable for data deposition.

Cross-disciplinary repositories

- Dryad Digital Repository
- figshare
- Harvard Dataverse Network
- Open Science Framework
- Zenodo

http://journals.plos.org/plosone/s/data-availability