# **Hakai Institute Research Catalogue Form and Data Packaging Guide**

**Please make a copy of this document before filling in the following information for your dataset and return this form, along with the data package to** [catalogue-team@hakai.org](mailto:catalogue-team@hakai.org)

Hakai Institute is a rapidly growing organization with a data collection as varied as it is large. A truly comprehensive research catalogue should include sometimes overlooked but vital supporting documentation and detail. Our goal is to populate the [Hakai Metadata Catalogue](http://hecate.hakai.org/geonetwork) with ancillary information, which will help enable us to inform Hakai researchers, affiliate collaborators, community partners, and the general public.

Data and results produced by Hakai employees will be made public and licensed under [Creative Commons Attributions 4.0 International](https://creativecommons.org/licenses/by/4.0/).

Recommendations for the contents of your data package:

1. If you would like to communicate additional information to the user that does not fit under the metadata schema, create a README.txt file to 1) orient users with the contents of your data package; 2) Describe thoroughly the field, lab, and data processing protocols used to produce your data; 3) Make dependencies and software requirements for opening your data or scripts explicit; 4) Provide any additional information that you deem useful such as links to literature referenced, equipment manuals, or spatial extents of sampling.
2. Create a ‘DATA DICTIONARY’. This describes each variable in every table of your data package. Include variable name, units, description. Recommended format: .**csv** or .**txt**
3. Please add a license.txt file, found [here](https://drive.google.com/drive/u/0/folders/1y5w2fciw7UfL4cfWGD-RnR5h-HT7ea3C?ths=true).
4. If updating an existing data package (because you are adding new data, or correcting errors), create a CHANGELOG.txt to keep track of what changes or additions have occurred since the last version. Follow [this guide](https://keepachangelog.com/en/1.0.0/) to keeping a changelog.
5. Save all your data tables as **.csv** files (other open formats such as .**txt** are fine too.).
6. Put all files into a folder and name the folder after your ‘Dataset-Title’ (should be the same name you enter for ‘Dataset Title’ below).
7. To create a new version of a data package, zip all existing files and store them in a sub-folder of the ‘Dataset-Title’ folder and name the folder after the previous version of the data package. All the new files can be put directly in the ‘Dataset-Title’ folder.

Optional:

* Include scripts that were used to clean or process data from the raw data, calculate values in the final data package, or that join data tables for common views in a subfolder.
* Include in your readme a link to the GitHub repository that you might be using to track changes of your project, including specific commits or releases that match the versioning used in a changelog.
* If your data package has numerous tables in a relational database structure, include a diagram describing relationships of tables such as an Entity Relationship Diagram.

**Catalogue Entry Form**

[Click here to view an example completed form.](https://docs.google.com/document/d/1ALsUvwzWzNImg_9w4jX9mp0e6eeuHZcbaR5HtZsgBDw/edit?usp=sharing)

**Dataset Title:**

**Topic – Location – Date(s) if applicable). Example: Jellyfish-Monitoring-Study–Pruth Bay, Calvert Island–September-2016. Try to avoid acronyms.**

|  |
| --- |
| **Seagrass Fish & Macroinvertebra** |

**Point of contact information:**

**Name - research institution - email**

|  |
| --- |
|  |

**How the data should be cited:**

Example: (Evans, W., A. Hare, and K. Pocock. (2017). High-resolution record of surface seawater CO2 content from December 2014 to April 2016 collected in Hyacinthe Bay, British Columbia, Canada. Version 1.0. Hakai Institute. Dataset. [access date].)

|  |
| --- |
|  |

**Description / Abstract:**

**A simple, easy to read description of what the dataset is.**

|  |
| --- |
|  |

**Data Type:**

**Format: TXT / SHP / CSV / XLSX / WORD / LAS / etc.**

|  |
| --- |
|  |

**Spatial Coverage Description:**

**Geographic description of the area covered in this study? Example: Coastal area between Triquet Island and Spider Island – BC – Central Coast.**

|  |
| --- |
|  |

**If it is a geospatial dataset please state coordinate system used:**

|  |
| --- |
|  |

**Topic Category:**

**For geospatial records, please select a topic category from this lis** [**https://www2.usgs.gov/science/about/thesaurus-full.php?thcode=15**](https://www2.usgs.gov/science/about/thesaurus-full.php?thcode=15)

|  |
| --- |
|  |

**Temporal Coverage:**

**When was the data collected? Example: July 21st 2016 to August 10th 2016.**

|  |
| --- |
|  |

**External links or resources to include in the record**

**Metadata records / applicable documents / google drive folders.**

|  |
| --- |
|  |

**Keywords**

**List keywords that describe your dataset (4-7 is ideal). Example: Oceanography, CTD, conductivity, temperature, salinity, depth, Calvert Island, coast, ocean, water, sample, plankton, chlorophyll.**

|  |
| --- |
|  |

**Record Thumbnail Image**

**If you have a square thumbnail image that well represents this dataset please provide it along with this metadata sheet.**

**Thank you for contributing to our metadata folder. Please submit this form along with any supporting documents to catalogue-team@hakai.org**

After submitting the form, the Hakai technology team will follow up with you to create a metadata record in our online library at [hecate.hakai.org/geonetwork](https://hecate.hakai.org/geonetwork/srv/eng/catalog.search#/home)