

Plan Overview

A Data Management Plan created using DMPTool-Stage

Title: ITB university level RDMP

Creator: Dasapta Erwin Irawan - **ORCID:** [0000-0002-1526-0863](https://orcid.org/0000-0002-1526-0863)

Affiliation: Non Partner Institution

Principal Investigator: Dasapta Erwin Irawan

Data Manager: Dasapta Erwin Irawan

Funder: Digital Curation Centre (dcc.ac.uk)

Funding opportunity number: 33596

Template: Digital Curation Centre

Project abstract:

This RDMP is the official RDMP template for all researches conducted by all ITB academic staffs, regardless the funder. However, ITB strongly suggest that ITB researchers should adhere the specific RDMP requirements from funders.

Last modified: 05-20-2018

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

ITB university level RDMP

Type of data

This RDMP covers the following type of data or document which are considered as data source:

- Raw data that may come in the following forms:
 - any field or laboratory measurements collected during in a research
 - any voice recording and its transcript of an interview or any other forms of data collection campaign
 - any vector and raster based images
 - any video recording and its text caption of an interview or any other forms of data collection campaign
 - survey form responses from participants
 - field notes or laboratory notebook
- Grant Proposals
- Project-level Research Data Management Plans (DMP)
- Shared texts, voice, or video recordings of communication between team member
- Reports
- Preprints
- Maps

Data formats

Although most of researchers use Microsoft-based applications, but we recommend the following file formats:

- Spreadsheets: text format, eg: csv (comma separated value), or txt (using tab separated value). All spreadsheet need to be in database format, with no merged rows or columns
- Documents: text format, eg: txt, markdown, or any other text format that can be created and read using plain text reader like Notepad
- Audio recordings: wav or mp3
- Video recordings: mp4 or mpg
- Images: general image: jpg, png, bmp, tiff
- Maps:
 - raster: geoTiff
 - vector: SHP
- Emails (project communications): txt

Standards of data collection

- All data should be acquired using international and national recognized standards.
- We promote the usage of [SNI \(Indonesian National Standards\)](#).
- Researcher should also refer to the equipment's manual guidelines and also to funder's guidelines.

Folder structure and file naming system

We suggest the following folder structure:

- data
 - raw data
 - process data
- analysis or process

- code
- plots
- tables
- output
 - articles
 - presentations

A readme file should be added to describe the structure and file naming convention.

We suggest the following file naming convention:

- <year-month-date>_<descriptive file name>_<initial of creator or editor>_<ver x>.<extension>
- example:
 - 20171115_IFLSanalysis_CNR_v1.sta
 - 20180417_statsanalysis_ERW_v2.csv
- don't use white spaces in the file name as it's not fully compatible with Linux based operating system and Mac Operating System.
- file name must be descriptive but compact.

Version control

We endorse the team members to use services that offer version control feature, such as GitHub, GitLab, OSF or other dedicated cloud repository services. PIs may request detail information to the admin staff of each service. However if it's not possible, then file naming is very important.

Quality assurance

PIs are fully responsible for data quality assurance. They may employ dedicated staffs or working jointly with data assurance institution to ensure the quality of the data.

All data will be preserved in open formats to ensure that its readability in the future. A metadata should be attached into each data file, or in some instance, a data folder. A readme file should be included in the root folder containing folder structure, general overview and some context of the data.

All deliverables (data, reports, presentations, preprints) should be recorded, listed, and stored in the project repository. A Readme file may be useful to describe the context, time frame, location, structure, and status of the files. A data staff or data steward may be assigned to check the status of the documentation.

We recommend the following minimum metadata schema for general data:

- Title of the dataset (see example)
- Abstract (to give context)
- Creator
- Contributor
- Publisher
- Funder
- Date of publication
- Resource type
- Location
- License/rights

- Data structure
- Data size
- File format

For geospatial dataset, we are referring to the [ISO 19115-1:2014](#) or the older ISO 19115-1:2003 geospatial metadata standard, which is also in used by the Badan Informasi Geospasial of Indonesia (Indonesia Board of Geospatial Information). The following tables contain minimum metadata schema for general dataset and general geodataset ([link to worksheet, open the related sheet](#)).

Ethics form and its evaluation

All researches must comply to ethical regulation. ITB recommends Universitas Padjadjaran Ethics Commision to evaluate ethic forms from ITB researchers.

Access to sensitive data

Users must register to access the data or contact University Data Steward and filling out a sensitive data usage form. The form then will be evaluated by university-level or faculty/school-level data steward (DS), given that the DS should also consult with the data creator or original researcher.

IP rights for the project are held by the university, or it could be a joint IPR management for joint research activity. It should be clearly mentioned in the data agreement.

Storage and its preservation periode

We anticipate less than five gigabyte of data and documents to be generated by the project. As far as possible data will be deposited in long term archives (minimum of 50 years of preservation). Data should be deposited at the start of the project and ended by the time final report submitted to the project funder. An embargo period (maximum of two years) may be assigned if needed. Following the end of the embargo period, an assigned data staff must make the data publicly available until minimum of 50 years.

Data and documents are stored on a three storage levels:

- working offline storage and at least one offline backup using portable hard drive
- online dynamic data repository using: university's available institutional repository and/or open repository services like the [OSF](#) (maintained by [Center for Open Science](#)), [Figshare](#) (maintained by [Digital Science](#)), or [Zenodo](#) (maintained by [CERN](#)).
- online static data repository: [ITB's repository](#) can be used to store the final dataset and other documents.

Backup

We suggest the following backup strategies:

- backup from offline working storage to portable media must be preformed immediately, daily backup is highly recommended.
- backup to cloud storage or repository at least once a week.
- team members are suggested to use backup application such as Apple Time Machine or [Free File Sync](#).

Data accessibility

The research team, relevant members of the research team, and project participants will be granted access to the data repository and to other online services. The access will be set through a unique userid and password system before embargo period ends. The minimum access for the above-mentioned parties will be "read-write" access. While "administrator" role should be given to the PI and at least two other team member one Co-PI and data staff. After exceeding the embargo period, the data repository will be made public.

Selection of material

All final materials as follows will be kept available in the ITB Institutional Repository and OSF dynamic repository:

- data:
 - raw data
 - final processed data
- reports:
 - preliminary report
 - mid term report and
 - final report

All intermediate and ongoing files, including data and other documents will be made available in the OSF dynamic repository.

Preservation

Long term preservation of publicly available data will be through appropriate repositories including institutional repository. More than one archive may be selected using the [LOCKSS principle](#) or [FAIR principle for data sharing](#) as the main criteria. In this case, ITB recommends OSF dynamic repository and ITB static institutional repository.

Sensitive data and information

The data creator or DS should ensure the anonymization/de-identification upon sensitive data. ITB refers to [EU Horizon 2020](#) for sensitive data handling.

Data should be deposited immediately following project kickoff and ended by the time final report submitted to the project funder. An embargo period (maximum of two years) may be assigned if needed. Following the end of the embargo period, an assigned data staff must ensure the availability of the data. A minimum of 10 years of preservation should be in consideration. However, there are open repositories that provide longer preservation, eg: up to 50 years or more.

Shared data and its form

In general sense, we recommend sharing raw, processed, analyzed, and final dataset. However, given the nature of the project, PIs may appeal for another form of data sharing. They could fill in data assessment form in order to come up with appropriate data sharing mechanism. PIs may have to:

- choose which type of data that they think could be safely shared without breaching data release agreement with other parties.
- separate primary from another institution from the primary new data acquired by team members.

End user license

We recommend to use moderate licenses, eg: [CC-BY license](#), [MIT license](#), and [Academic Free License](#), as the default license for data and also for all resulting documents. However for PIs may propose another license, such as [CC0 waiver](#), [CC-BY-SA](#). For sensitive data, PIs may suggest a restrictive license.

Data indexing

All data and data repository should be able to be found by at least one indexing services, eg: [Google Scholar](#). Common repositories are now accessible via [BASE](#) and [ONESearch](#) (a feature from Indonesia National Library and Archive). To be formally cited, we also recommend the usage of persistent link, eg: DOI from [CrossRef](#) or [Datacite](#).

Restriction is given to sensitive data (refer to [Horizon 2020 Guidelines](#)). Any request to this kind of data must be sent to Official Data Staff (eg: University-level or project-level Data Steward).

PI and an assigned data steward (DS) responsible for research data management. This includes file conversion, classifying and managing the various research outputs identified in this RDMP, throughout the research cycle and during the lifetime of the data.

In the case of a change of PI or data steward (DS), responsibility will transfer to the one of the Co-PI or to a data staff assigned by the PI or institution.

Aside to the data collection phase, the major costs of data management for the project are for management and storage components. The management components should be funded by research project, while the storage is the responsible of university, or a PI may select a free-open repositories.
