

Promoting data sharing among Indonesian scientists: A proposal of generic university-level Research Data Management Plan (RDMP)

Dasapta Erwin Irawan[‡], Cut Novianti Rachmi[§]

[‡] Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Bandung, Indonesia

[§] Faculty of Medicine, Universitas Padjadjaran, Bandung, Indonesia

Corresponding author: Dasapta Erwin Irawan (dasaptaerwin@outlook.co.id)

Reviewable v1

Received: 04 Jul 2018 | Published: 06 Jul 2018

Citation: Irawan D, Rachmi C (2018) Promoting data sharing among Indonesian scientists: A proposal of generic university-level Research Data Management Plan (RDMP). Research Ideas and Outcomes 4: e28163.

<https://doi.org/10.3897/rio.4.e28163>

Abstract

Every researcher needs data in their working ecosystem, but despite of the resources (funding, time, and energy), that they have spent to get the data, only a few are putting more real attention to data management. This paper is mainly describing our recommendation of RDMP document at university level. This paper would be a form of our initiative to be developed at university or national level, which also in-line with current development in scientific practices mandating data sharing and data re-use.

Researchers can use this article as an assessment form to describe the setting of their research and data management. Researcher can also develop more detail RDMP to cater specific project's environment. In this Research Data Management Plan (RDMP), we propose three levels of storage: offline working storage, offline backup storage and online-cloud backup storage, located on a shared-repository. We also propose two kinds of cloud repository: a dynamic repository to store live data and a static repository to keep a copy of final data.

Hopefully, this RDMP could solve problems on data sharing and preservation, and additionally could increase researchers' awareness about data management to increase the value and impact of their researches.

Keywords

Research Data Management Plan, open data, data sharing, data repository, reproducible research

Introduction

Good data management will support scientific discovery (Wilkinson et al. 2016), but at the other end we have been observing cultural barrier in data sharing (Davidson et al. 2014). More insights about data sharing and the diverse perceptions among scientists in various fields have been endlessly discussed (Tenopir et al. 2011, Tenopir et al. 2015, van Panhuis et al. 2014, Wallis et al. 2013).

Every researcher needs data in their working ecosystem, but despite of the resources (funding, time, and energy), that they have spent to get the data, only a few are putting extra attention to data management (Irawan 2018, Irawan 2017). Data management strategy is not just a administrative document, but by far, it plays an important role to guide researchers to store, backup, preserve, and share their research data in a proper and sustainable manner.

This paper describes a guideline to build a university-level Research Data Management Plan (RDMP) and how it can promote data sharing among scientists. This RDMP would be the first one to be developed at university level in Indonesia. This project is in-line with current development in scientific practices mandating data sharing and data re-use. The goal of this RDMP project are to build awareness about data sharing and preservation to scientists, especially academic staffs, and to build a practical and simple tool to help them manage their research data. The goal of an RDMP project are to guide researchers to manage their data, including curating, storing, sharing, and preserving it for immediate and future use.

This RDMP proposal is largely extracted from our experience in developing RDMP for an international research collaboration funded by RCUK (Research Council UK) (Irawan and Rachmi 2018).

Description

General overview

The concern to having a proper RDMP was triggered by difficulties faced by researchers to find data from another researcher or previous research and to extract data from reports. The other problem is to find guidelines, especially in Indonesia, on how to appropriately manage your research data, to store them, and to keep them available in the long run. Clearly scientists have issues on how to re-use dataset from prior research, how to cite them into their own work (re-use), and how do they know the limitation of such action.

Due to the large effort to get data in terms of funding, time, and energy, the life time of data should be more than one or two years, as we find to be the general nature in Indonesia research ecosystem (Irawan 2018, Neylon 2017a, Neylon 2017b, Borghi et al. 2018) (Fig. 1). Another important point to address is the barrier of data sharing that involves the fear of getting scooped, the lack of knowledge to Intellectual Property Rights (IPR) and data ownership. Therefore, by developing this document, we could solve the barriers and at the same time we could come up with another way to increase the value of research data, instead of only looking at mainstream metrics.

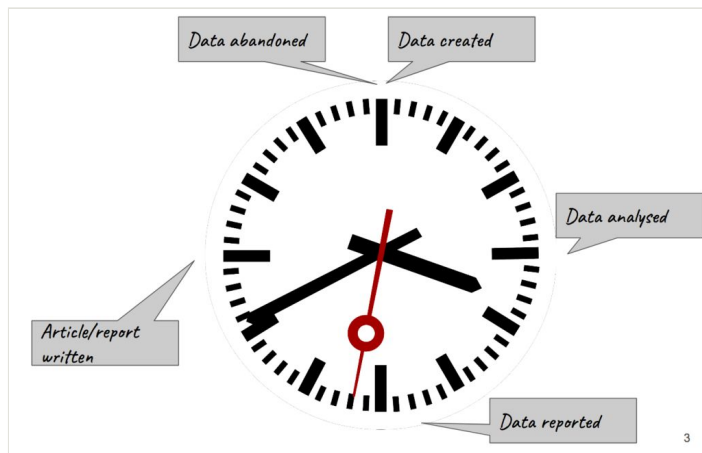


Figure 1. [doi](#)

Current situation of data lifecycle (Irawan 2018).

How to use this article as guidelines

Researchers can use this article as an assessment form to describe the setting of their research and data management requirements from potential funder. Researchers can develop more detail RDMP to cater specific project's environment. They should justify the setting of their research and requirement of the funder regarding data sharing and data preservation.

Seven components in RDMP

The proposed RDMP is divided into seven components:

1. Data collection
2. Documentation and metadata
3. Storage and backup
4. Preservation
5. Sharing and re-use
6. Responsibilities and resources
7. Ethics and legal compliance

References

Given the different nature of research, funders, and DMP standards, we refer to the following sources in developing this RDMP:

- Data sharing culture (Neylon 2017a, Neylon 2017b)
- Open data principles and reproducible research (Irawan et al. 2017)
- RDMP check lists or rubric (Digital Curation Center 2016, Teperek et al. 2017, University of California Curation Center 2018)
- RDMP case study from various fields of sciences (Neylon 2017c, Traynor 2017, Wael 2017, Woolfrey 2017)

Component 1: Data collection

What types of data will you collect, create, link to, acquire and/or record?

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 phase
 - any vector and raster based images
 - any video recording and its text caption of an interview or any other forms of data collection phase
 - survey form responses from participants
 - field notes or laboratory records
- Grant Proposals: funders may request researcher to submit their research plan as pre-registration document in several platform such as [OSF](#) or [Curate Science](#).
- Project-level RDMP: some funders, such as RCUK, mandates the submission of final RDMP before the project begins.

- Shared texts, voice, or video recordings of communication between team member.
- Reports: this documents may appear as preliminary report, mid-term report, final report, or short communications.
- Preprints: preprint has been admitted as part of research output by [several funders](#) (Bourne et al. 2017).
- Maps

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

Although most of researchers use Microsoft-based applications and Most open repositories accept and provide native viewer for many formats, but the following are our choice of formats. You may refer to [University of Sydney RDMP](#) file formats or [Cornell University's preservation file formats](#) for more information.

Spreadsheets

It should be written in text format, eg: csv (comma separated value), or txt (using tab separated value). Data creators should format the spreadsheet in a "database" format:

- start the data immediately in cell (1,1),
- avoid merging rows or columns, and
- clearly use the correct and consistent cell format, eg: number, string, date, time, category.

Documents

We recommend text-based (ASCII) file, eg: txt, [Markdown](#), or any other text format that can be created and read using plain text reader like Notepad

Audio/video recordings

- Audio recordings: wav or mp3
- Video recordings: mp4 or mpg

Images and maps

- general image: jpg, png, bmp, tiff
- raster: geoTiff
- vector: SHP

Emails (project communications)

Although most researchers are now using proprietary email clients like Ms Outlook or Apple Mail, but they need to store selected emails in to plain text as well.

What conventions and procedures will you use to structure, name and version control your files to help you and others better understand how your data are organized?

Files are uploaded to online repository and organized into folders by phase or by working package. If the file organization get too complicated to accommodate into a set of folder structure, then it should be separated and linked together. We recommend the following set of folders to organize the files.

root folder:

- data:
 - raw
 - processed
- analysis
 - code (or script)
 - tables
 - figure or image)
- output
 - report
 - presentation
 - article (or manuscript)

Some field of research may have other specific folder arrangement, but generally they should have the components in the figure. If some team members choose to maintain a [Google Drive](#), [DropBox](#), [Onedrive](#) or other cloud services, then they should make an accessible link to the drives or folders and register the links to the data repository. To accommodate limited storage, Principal Investigator (PI), Co-PI, and team members may also maintain an open repository, such as: [OSF](#), [Figshare](#), [Zenodo](#), [GitHub](#), [GitLab](#), and other similar services, given that such services offer version control and access option features. All services should be linked together to a central repository. The team may also maintain a dedicated project website to store the data and related research documents, to keep track of the activities, and to store the project's repository or storage structure.

Component 2: Data documentation and metadata

What documentation will be needed for the data to be read and interpreted correctly in the future?

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.

How will you make sure that documentation is created or captured consistently throughout your project?

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 (DS) may be assigned to check the status of the documentation.

What metadata standard will be needed to describe your data

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 refer to the [ISO 19115-1:2003](#) geospatial metadata standard, which is also 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](#)).

Component 3: Storage and backup

What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

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. A minimum of 10 years of preservation should be in consideration, but there are open repositories that provide longer preservation, eg: up to 50 years or more. 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 10 years.

How and where will your data be stored and backed up during your research project?

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: institutional repository can be used to store the final dataset and other documents.

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](#).

How will the research team and other collaborators access, modify, and contribute data throughout the project?

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.

Component 4: Preservation

Where will you deposit your data for long-term preservation and access at the end of your research project?

Selection of material

All final materials as follows will be kept available in the 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 [LOCKS S principle](#) or [FAIR principle for data sharing](#) as the main criteria. In this case, we recommend OSF dynamic repository and static institutional repository.

Indicate how you will ensure your data is preservation-ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

For all data generated from a research, we may ask the data creator to convert it from any proprietary file formats to open formats for long term preservation. Other option would be to have a data staff (DS) may be assigned to work on file conversion. The data creator or DS should ensure the anonymization/de-identification upon sensitive data.

Component 5: Sharing and Re-use

What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final)

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.

Have you considered what type of end-user license to include with your data?

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.

What steps will be taken to help the research community know that your data exists?

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](#).

Component 6: Responsibilities and Resources

Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible of.

PI and an assigned DS are 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.

How will responsibilities for managing data activities be handled in case substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

In the case of a change of PI or DS, responsibility will be transferred to the one of the Co-PIs or to a DS assigned by the PI or institution.

What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

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.

Component 7: Ethics and Legal Compliance

An Intellectual Property Rights (IPR) officer at university level is very much needed in this case, but researchers should also have enough basic knowledge regarding this subject.

If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

A university-level or several faculty-level Data Steward (DS) should be assigned to ensure the management of sensitive data and general data management in general. The access to the such data may be restricted to PI, one of the Co-PIs, and the DS. The DS will have a checklist form to help them assess the situation.

If applicable, what strategies will you undertake to address secondary uses of sensitive data?

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

How will you manage legal, ethical, and intellectual property issues?

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.

Acknowledgements

We thank the following persons for their feedback and corrections to this article: the repository team of Institut Teknologi Bandung, Willem Vervoort and Gene Melzack (from University of Sydney), Driajana, Akhmad Riqqi, and Yudi Darma (from UDARA team), and also Sarah Lindley (from University of Manchester), open science community and INArxiv preprint server users

Hosting institution

The university solely, or in case of a joint research, the hosting institution should be clearly stated in the data sharing and ownership agreement.

Author contributions

Both authors contribute evenly in this article.

Conflicts of interest

Both authors declare no competing interest upon the publishing of this paper.

References

- Borghi J, Abrams S, Lowenberg D, Simms S, Chodacki J (2018) Support Your Data: A Research Data Management Guide for Researchers. *Research Ideas and Outcomes* 4: e26439. <https://doi.org/10.3897/rio.4.e26439>
- Bourne P, Polka J, Vale R, Kiley R (2017) Ten simple rules to consider regarding preprint submission. *PLOS Computational Biology* 13 (5): e1005473. <https://doi.org/10.1371/journal.pcbi.1005473>
- Davidson J, Jones S, Molloy L, Keiser UB (2014) Emerging Good Practice in Managing Research Data and Research Information within UK Universities. *Procedia Computer Science* 33: 215-222. <https://doi.org/10.1016/j.procs.2014.06.035>
- Digital Curation Center (2016) Checklist for a Data Management Plan. <http://www.dcc.ac.uk/resources/data-management-plans/checklist>. Accessed on: 2018-5-08.
- Irawan D, Vervoort R, Melzack G (2017) Open Data Workshop SSEAC USYD - ITB. Open Science Framework <https://doi.org/10.17605/OSF.IO/S76GU>
- Irawan DE (2017) A light introduction to research data management. Figshare <https://doi.org/10.6084/M9.FIGSHARE.5418694.V1>
- Irawan DE (2018) RDM policy and data archiving at university level -- technical bits -- an example from ITB. Figshare <https://doi.org/10.6084/M9.FIGSHARE.6179084.V1>
- Irawan DE, Rachmi CN (2018) Institut Teknologi Bandung Repository and RDM. 1. OSF repository. Release date: 2018-5-16. URL: <https://osf.io/59vcn/>
- Neylon C (2017a) Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers. *Research Ideas and Outcomes* 3: e21705. <https://doi.org/10.3897/rio.3.e21705>
- Neylon C (2017b) Building a Culture of Data Sharing: Policy Design and Implementation for Research Data Management in Development Research. *Research Ideas and Outcomes* 3: e21773. <https://doi.org/10.3897/rio.3.e21773>
- Neylon C (2017c) Data Management Plan: IDRC Data Sharing Pilot Project. *Research Ideas and Outcomes* 3: e14672. <https://doi.org/10.3897/rio.3.e14672>
- Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M (2011) Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6 (6): e21101. <https://doi.org/10.1371/journal.pone.0021101>
- Tenopir C, Dalton E, Allard S, Frame M, Pjesivac I, Birch B, Pollock D, Dorsett K (2015) Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE* 10 (8): e0134826. <https://doi.org/10.1371/journal.pone.0134826>
- Teperek M, Mollitt B, Southall J, Donaldson M (2017) Wellcome DMP Assessment Rubric V2.0. Zenodo <https://doi.org/10.5281/ZENODO.257650>
- Traynor C (2017) Data Management Plan: Empowering Indigenous Peoples and Knowledge Systems Related to Climate Change and Intellectual Property Rights. *Research Ideas and Outcomes* 3: e15111. <https://doi.org/10.3897/rio.3.e15111>
- University of California Curation Center (2018) Data Management Plan Tool. <https://dmptool.org/>. Accessed on: 2018-5-08.
- van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, Heymann D, Burke DS (2014) A systematic review of barriers to data sharing in public health. *BMC Public Health* 14 (1): . <https://doi.org/10.1186/1471-2458-14-1144>

- Wael R (2017) Data Management Plan: HarassMap. Research Ideas and Outcomes 3: e15133. <https://doi.org/10.3897/rio.3.e15133>
- Wallis J, Rolando E, Borgman C (2013) If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. PLoS ONE 8 (7): e67332. <https://doi.org/10.1371/journal.pone.0067332>
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, Silva Santos LBd, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, Schaik Rv, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, der Lei Jv, Mulligen Ev, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018. <https://doi.org/10.1038/sdata.2016.18>
- Woolfrey L (2017) Data Management Plan: Opening access to economic data to prevent tobacco related diseases in Africa. Research Ideas and Outcomes 3: e14837. <https://doi.org/10.3897/rio.3.e14837>

© 2018. This work is published under <http://creativecommons.org/licenses/by/4.0/> (the “License”). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.