# RESEARCHERS GUIDELINE TO RESEARCH DATA MANAGEMENT

By: Salmah Abdullah, Roziana Shamsuri, Zubaidah Iberahim, Siti Noorhaniza Hamdan,
Nida Hidayati Ghazali & Liza Ab Llah
Research and Information Services Division, PSAS

## Introduction

Data can be described more broadly in the context of the research from an information science perspective as any information that has been gathered, observed, generated, or developed to validate initial research findings. Research data can be found in non-digital formats like laboratory notebooks and diaries. Data produced by non-research activities such as university administration or teaching, as well as administrative data created during personal activities, desktop or mailbox backups, are not considered as research data. The same information, for instance, may be research data for one researcher but not for another depending on whether it is being used as a key component of a research activity.

## Some benefits of managing research data are:

- locate and comprehend facts when required
- saves time in organizing data
- reduces the danger of data being lost, stolen, or improperly used
- comply with funder and journal requirements
- facilitates easier results validation
- data can be shared, leading to collaboration and greater impact

## Data Management Practices

Here are some good data management practices that researchers can adopt during research activities:

1. **Save your raw data**

   Saving data means keeping track of research materials so that you or others can access and utilize them in the future. Data should be kept in a format that makes it possible to use it later. This may involve storing data in open or easily accessible file formats or just keeping your data on hand with the supporting documentation and other research resources. Don't overwrite your original data with a cleaned copy. Lock or make your original data read-only to protect it.

2. **Backup your data**

   It is important to perform backup because it help prepare for disasters such as accidental deletions, fires or natural disasters, software bugs and hardware failures. Keep three copies of your data on two different devices, and one copy off site. Do not backup or store sensitive data on a cloud services like Google Drive/Dropbox. It is hard to guarantee their data security and cloud services are not reliable for long-term preservation due to syncing-issues.

3. **Describe your data**

   Your data will be accessible to all users if it is described clearly and accurately. It is essential to record contextual details about how and why the data were created. The following questions should be answered in order to describe your data;

   - what it is?
   - how will it be collected? by who?
   - how much data will be generated?
   - what file types do you utilize for data?
   - any equipment or software used?
   - is any personal identifiable information or confidential data?
   - are you using data that someone else produced? if so, where is it come from?
   - what software or programming is required to read or comprehend these files?
   - what was changed? new project members? new methods? when did it happen? why?
   - who will be the data's custodian?

4. **Citing your data**

   Datasets used during the research process should be cited like you would cite an article in the reference or bibliographies sections of your works. The practice of citing research data has evolved as researchers and stakeholders have come to understand the value of including data in the scholarly record between a research output and the supporting evidence that supports it. Citing data will give credit to the responsible researchers and enables those who share the data to assess its impact. You can follow the citation guidelines provided by the publisher when referencing a dataset in a publication. Gather all the essential components and match the reference for textual articles if they don't specify a format for datasets. You can also use DataCite's citation format for a dataset and modify it to fit your own citation format.

5. **Preserve your data**

   Research data preservation is the process of maintaining access to the data so that it can be found, comprehended and re-used in the future. Data preservation strives to maintain content (data) and context (metadata). Preservation must be taken into account at the time research data are created and during the entire life cycle of data. Consider the following steps to ensure your data remain useable and understandable for the future:
   - effectively document your data
   - moving data to new storage media on a regular basis
   - keep many copies of the data on various types of storage media
   - use a format that can be easily imported into different software packages, or move data to new software versions
   - keep data and metadata in a reliable repository that supports long-term data preservation

**Additional Readings**

If you are interested to know what researchers from other university libraries have done on research data management, you can click on this link for further reading:
- Universiti Sains Malaysia
- University of Chicago, United States
- Harvard University, United States
- Massachusetts Institute of Technology (MIT), United States
- University of Oxford
- Stanford University, United States
- ETH Zurich
- California Institute of Technology (Caltech), United States
- University of Pennsylvania
- Yale University, United States
- Princeton University, United States
- UCL, United Kingdom
- National University of Singapore (NUS), Singapore

**Conclusion**

Good data management practice is essential to share research data in an effective way. Among the benefits of good data management practice are: increased data quality; increased research productivity; increased exposure of research data and results through sharing and dissemination; increased reproducibility of experimental procedures; increased validation and verification of results; increased interoperability between data and between data and tools; and improved operation of repositories and infrastructures.