

# Strategies And Challenges In Archiving And Sharing Research Data

*\*Dwi Fitriana Cahyaningtyas, Ida Fajar Priyanto  
Graduate School of Library and Information Management  
Universitas Gadjah Mada Indonesia  
[fitrinacahya@gmail.com](mailto:fitrinacahya@gmail.com)*

## **Abstract**

*Research data management is increasing in its importance for research sustainability and the data are best if they are reusable. Collecting research data from previous research is not always easy, especially when there is no regulation concerning with research data management. Research data management is still new in Indonesia and few institutions implemented the regulation. When requesting research data to start the research data repository, the responses varied. This raised a question why researchers refused to submit their data.*

*This study is aimed at (1) understanding what researchers in the field of library and information science know about research data management; (2) whether they are willing to share their research data to be reusable; (3) what are the causes affecting the data submission.*

*This is a qualitative research on the acceptance of researchers' data submission. The informants are researchers in the field of library and information science in Indonesia. Data were collected through an online interview with some researchers in their field.*

*Researchers in Indonesia are still not aware of the importance of research data management. Lack of understanding about the importance of data usability, data citation and the importance of reuse of data, lead researchers to disregard the invitation and persuasion to submit their research data. The results of the study also show that not all researchers are willing to share their data because they are afraid that the data will be used by others and they want to keep their data closed. In addition, some researchers acknowledged that they did not have valid data and they feel inconvenience if others know that their data are not valid.*

**Keywords:** Research data management, researchers, data, data sharing, data archiving

## **1. Introduction**

Technological advances for collecting, storing, and data analyzing have facilitated the collection of more data now than ever before in history — a phenomenon known as data proliferation (Borgman et al. 2007; Quinn and Alexander 2008). The era of data proliferation has brought new opportunities and new challenges in areas as diverse as marketing, homeland security, and molecular biology (Spengler 2000; Shaw et al. 2001; Seifert 2004). In China recently, research data management (RDM) is in great demand of Higher Education Institutes (HEI) internationally since 2005 to 2010 (Moe, 2015). This increase in the breadth and depth of data required has encouraged the development of new data strategies to efficiently manage and share available data. The Swedish government, America and Canada has decided that in all disciplines, all research results that are funded by public

funds are expected to be openly accessible by attaching research data (Borglund and Bogerud, 2020). The Policy which ensuring that research data are available in public archives is increasingly being applied in government, funding agency, and journal level. This policy is based on ideas that authors need to be the ambassadors of our data, to represent it in a way that speaks to why you made the effort to collect it and what you expect to learn from it. The idea behind open data are that it should be possible to accessible, reuse public information for free to create new ideas and innovations.

It is slightly different in Indonesia, mostly the research it is concerned only on the final result, the institutions or the HE requires to submit the contents/subject of the research only, the research data ita only for the complement (1), it is even more concerned with the administrative reports (2), The data which has been collected or processed by reseachers is only stored by themselves (3).

In practice, the data are still a consideration in research both from sponsors and research institutions, the value of the research data are still has not considered. It caused the researcher can not manage well their research data. According to Childs et al (2014) it is often that researchers store research data on media that are unreliable for preservation, easily lost of important data. Most researchers when it came to storing and archiving research data, its frequently selected on hard drive of a personal PC or laptop, or an external hard drive/USB drive or even cloud. There are some considerations and arguments given by researchers when they want to share their research data.

The absence of regulations and data management policy causes researchers to keep their data by themselves. However The Indonesian institutions are still not aware of the importance of research data management. Lack of understanding about the importance of data usability, and the importance of data reuse, lead researchers to disregards the invitation and persuasion to submit their research data. This is because most of the research data are still managed by researchers or research groups, including researchers in the field of librarianship . One of the issues in librarianship is data management in the library. Awareness about Research and data management just emerge quite recently.

Indonesian Institute of Science (LIPI) was the first intitution that started to build national research data repository (Repositori Ilmiah Nasional/RIN) in all diciplines, followed by National Library of Republic of Indonesia that built data repository in the field of librarianship. These two national institutions are leading the country for the research data management.

On the previous research on open data concludes that regulations concerning the collection and preservation of research data are unclear (Grant 2017). In addition, archival aspects are rarely taken into consideration in research projects, and there is a widespread lack of knowledge about how to preserve research data over time. With regards to Indonesia, research data management is still new and needs to let researchers know about this.

## **1. Problem statement**

Good research is the one that has proper management including its data which are reuseable. In Indonesia data have not been managed well. Researchers usually keep their data by themselves. Whether the data management is good or bad, it depends on the individual researchers' ways of keeping them. When researchers are requested to submit their data, their responses vary—some are willing to share, while some others are reluctant to do so. What might be the reasons behind their willingness and reluctance have not been analyzed. This research is to find out whether the researchers know about research data management, the benefits of research data management, and research data sharing

## **2. Literature Review**

### *2.1 The importance of Research data and RDM*

Data are usually collected or produced as part of the research process and is now being generated in ever-increasing volumes and in a variety of digital formats that are often rapidly being replaced. (Berman and Cerf, 2013; Borgman, 2012; Pryor, 2012). Research data can be in any format in which

it is created, for examples: text, numeric, audio-visual, models, computer code, discipline-specific, instrument-specific. Data are divided into two types, namely, (1) primary data (data that are collected for the specific research problem, using procedures that fit the research problem best. (2) secondary data (data that have been collected by others, for another purpose but relevant to your research needs). Complementing this practical need is a changing perception of the value of research data: it has been seen as an asset that must be managed in order to maintain its value (Higgins, 2012; Lavoie, 2012).

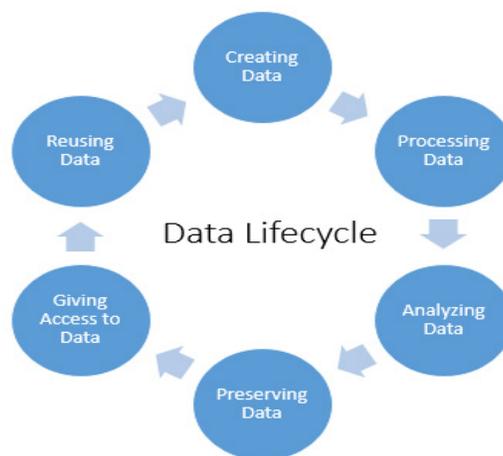
According to the Whyte, A., Tedds, J. (2011) "Research data management is concerned with organizing data, from inputting the data to disseminating and archiving valuable results. It aims to ensure reliable, up-to-date and innovative verification of results to build on existing information." In managing data, we have to consider the process, which are: the creation of data and its plan to use (1) Organization, structure, and data name, should be secure and access is provided, (2), storing data and back them up, (3), and sharing data with research collaborators and more broadly, publishing and get the data cited (4). The focus is on what is needed for validation and re-use.

## 2.2 Why sharing data?

Good management of data should consider research data lifecycle as it allows researchers to analyze and reanalyze data in order to verify the results and replicate studies to train new generations of researchers, to create new ideas and innovations (Bertagnolli et al. 2017). Data sharing has benefit for many reasons: first, without sharing data it is impossible to verify the results of research, this is a good principle of great science (Borgman, 2012). However, it can improve the reuse, visibility of the author, collaboration and research integrity. In addition it also reduces risks of data loss (1), data leak (2), Copyright infringement (3), and Breach of contract (4). Indonesia has some of institutions and some of guidance.

Unfortunately, Indonesia is not quite familiar with research data management. Due to the absence of national or institutional policies, data are generally managed by researchers with limited access indicates the weaknesses of RDM services are lack of skill, lack of research data policy (Marlina, Purwandari. 2019)

The research data lifecycle model describes and identifies the steps to be taken at the different stages of the research cycle to ensure successful data curation and preservation. There are several stages in the research data lifecycle, e.g. data creation, data processing, data analysis, etc.



**Pict 1: research data lifecycle** (source: <https://www.reading.ac.uk/RES/rdm/about/res-rdm-lifecycle.aspx>)

There are a few models that one can make use to plan the data management activities, for example, DCC Curation Lifecycle Model. The diagram above illustrate the lifecycle in six stages. The research data cycle needs to be considered by institutions in implementing RDM policies and services, which include: (1) Create; it is included research design, create the framework collecting data and metadata (2) Process; data entry, validation, description and storage (3) Analyse; interpretation, derivative of data, and publication (4) Preserve; metadata, documentary and archiving (5) Share; metadata back-up, storage, sharing, access control, copyright, and promotion (6) Reuse; Data that are available for discovery and access may be re-used by other researchers

Data are an important economic resource in any of aspects, in any discipline including biomedical research (Downey and Olson 2013). Data can be shared and reused in infinite mode without being “consumed” or reduced in availability (Pronk et al. 2015). Research data management is useful for possible future research but not all researchers are willing to share their data and some take the data for granted or leave the data deteriorating and lost. Regulation concerning with data management in an institution should be implemented in order to make aware of the importance of data repository and data management among researchers. Sadly, in Indonesia most of researchers, the large majority, never receive any training in how to share and communicate about their data, even to their peers. In fact, there are many potential benefits of good research data management, other researchers, and the wider community: Increase impact of research through knowledge transfer (1), Efficiency and ease of data control, reduced easy loss data (2), Research develops through the reuse of data by a wider range of researchers (3), Compliance with funder and institutional policies and expectations (4), Demonstration of research integrity and verify of research results (5) (Markowitz (2015) *Genome Biol* 16, 274).

### **3. Methodology**

This is a qualitative research to find out their understanding of research data management, acceptance of researchers’ data submission, and willingness to share their data. The informants are Indonesian researchers whose backgrounds are librarianship. Data were collected through online interviews with **25** researchers in their field.

### **4. Discussion and Finding**

Researchers in Indonesia are still not aware of the importance of research data management. While the National Science Agency (LIPI) has launched the Research Data Repository for all research in Indonesia, most institutions have not implemented any regulation regarding research data submission in their institutions. This also exists in the subject-based research data management, such as Library and Information Science research data management. All other respondents were also varied from different institutions, such as from higher education, library and other institutions. Due to the limitations of city differences, interviews were conducted by online through video call, phone call and by an email.

Of the 25 researchers contacted, due to busy schedule and other factors, 15 researcher indicated willingness to participate in the survey. Indeed this is in contrast with the data hierarchy in which data reusability is of the highest value. Some researchers state that they use the data for once and they leave the data deteriorating.

In terms of levels of awareness of policies the researchers showed low levels of awareness. Among the researchers who archived their data, it is only 5 of them did it for their own re-use. They can organize, Keep it, make it secure, provide access, store and back it up.

What is also surprising of 3 researchers who refuse to share data because they do not want to share. In addition, some researchers acknowledged that they did not have valid data and they feel inconvenience if others know that their data are not valid.

One informant stated as follows:

*“Aku takut datanya nanti dipake orang dan bisa dimanipulasi”* because they are afraid that the original data can be manipulating by some others.

The other one stated as follows

*“Peneliti selanjutnya bisa mencari data yang lebih baik dari punyaku”* the next researcher can find the better data, than mine. In short, they think that data should be keep by the researcher only.

On the other hand, of the 10 researchers were willing to share the data, but most of them were lost the data or they forgot where they keep it. They are not managing it well enough to make sense of it. They assumed that the lack of attention to storing research data has the impact of not good at manage their personal data management, they just focused on the result of the research. This then causes data loss. Although they are willing to submit their data, but they cannot find their data anymore. Furthermore, the researchers did not consider their research material as public, but as complementary.

Among the researchers who contacted, the rest of 7 they responded the interview but in the end they did not submit the data.

## 5. Conclusion

Data sharing can increase the return of research projects by permitting other researchers to perform secondary collect data, doing further study with their own exploration. Based on the study, Indonesian researchers varied in their view of research data management. From the above discussion, it seems that only some researchers understand the importance of sharing data, while others consider data as private ownership. What is suprising is that some researchers consider that sharing data may enable other researchers manipulate the data.

Regulations and policies are important and need to be socialized first so that researchers understand the importance of RDM for sustainable research. The unclear regulations for sharing data make researchers reluctant to provide data (1) There is no regulations from the research institute (2). It is related to this, the authors provide several recommendations to managers of scientific journals, research institutions, funders, and policy makers. (1) For this reason, there is a need for socialization of the data sharing movement both in research institutions and higher education.

## References

- Article from a journal
  - Berman, F. and Cerf, V. (2013), “Who will pay for public access?”, Science, Vol. 341 No. 6146, pp. 616–7.
  - Borgerud, C., Borglund, E. 2020. Open research data, an archival challenge?. Arch Sci 20, 279–302. <https://doi.org/10.1007/s10502-020-09330-3>
  - Borgman CL. 2012. The conundrum of sharing research data. J Assoc Inf Sci Technol. 63(6):1059–1078.
  - Bertagnolli MM, Khozin S, Murphy MJ. 2017. Data-sharing models. N Engl J Med. 376(23):2305–2306.
  - Childs S, McLeod J, Lomas E, Cook G (2014) Opening research data: issues

and opportunities. *Rec Manag J* 24(2):142–162. <https://doi.org/10.1108/RMJ-01-2014-0005>

- Downey AS, Olson S. 2013. Sharing clinical research data: workshop summary. Washington (DC): National Academies Press
- Grant R. 2017. Recordkeeping and research data management: a review of perspectives. *Rec Manag J* 27(2):159–174. <https://doi.org/10.1108/RMJ-10-2016-0036>
- Higgins, S. (2012), “The lifecycle of data management”, in Pryor, G. (Ed.), *Managing Research Data*, Facet Publishing, London, pp. 17–46.
- Lavoie, B.F. (2012), “Sustainable research data”, in Pryor, G. (Ed.), *Managing Research Data*, Facet Publishing, London, pp. 67–82.
- Markowetz, F. 2015 Five selfish reasons to work reproducibly. *Genome Biol* 16, 274. <https://doi.org/10.1186/s13059-015-0850-7>
- Ministry of Education (MOE) (2015), “The twelfth five-year plan for science and technology of HEI, Ministry of Education of China”, available at: [www.gov.cn/zwgk/2012-03/27/content\\_2100852.htm](http://www.gov.cn/zwgk/2012-03/27/content_2100852.htm)
- Pennock, Maureen. 2007. *Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information*. UK: University of Bath.
- Pronk TE, Wiersma PH, van Weerden A, Schieving F. 2015. A game theoretic analysis of research data sharing. *PeerJ*. 3:e1242
- Pryor, G. (2012), “Why manage research data?”, in Pryor, G. (Ed.), *Managing research data*, Facet Publishing, London, pp. 1–16.

○ Conference paper

- Ekawati Marlina, Betty Purwandari. 2019. Strategy for Research Data Management Services in Indonesia, in “The Fifth Information Systems International Conference 2019” DOI: 10.1016/j.procs.2019.11.184

○ World Wide Web site

- The Research Data Lifecycle.(200) Retrieve on <https://www.reading.ac.uk/RES/rdm/about/res-rdm-lifecycle.aspx>