

## STRATEGIES AND CHALLENGES IN ARCHIVING AND SHARING RESEARCH DATA

**Dewi Maharani R<sup>1</sup>, Surya Adi Sasmita<sup>2</sup>**

*<sup>1, 2</sup>Universitas Terbuka (INDONESIA)*

### **Abstract**

Research data management is increasingly important for research sustainability and the data is best when it is reusable. Collecting research data from previous research is not always easy, especially if there are no regulations governing the management of research data. Management of research data is still new in Indonesia and only a few institutions have implemented these regulations. When requesting research data to initiate research data storage, responses vary. This raises the question of why the researchers refused to submit their data. This study aims to (1) find out what researchers in the field of library and information science do about research data management; (2) whether they are willing to share their research data for reuse; (3) what factors influence the submission of data. This research is a qualitative research about the acceptability of submitting research data. The informants are researchers in the field of library and information science in Indonesia. Data was collected through online interviews with several researchers in their fields. Researchers in Indonesia are still not aware of the importance of managing research data. While the National Science Agency (LIPI) has launched a Research Data Repository for all research in Indonesia, most institutions have not implemented regulations regarding the submission of research data within their institutions. This is also found in the management of subject-based research data, such as the management of library and information science research data. Lack of understanding about the importance of data usage, data citation and the importance of data reuse, causes researchers to ignore invitations and inducements to send their research data. The results of the study also show that (1) Not all researchers are willing to share their data because they are afraid that the data will be used by other people and want the data to be kept confidential. Indeed, this is different from data hierarchies where data reusability has the highest value. Some researchers claim that they used the data for once and they let the data deteriorate. This then causes data loss. Even though they are willing to submit their data, but they can't find their data anymore. In addition, some researchers admit that they do not have valid data and they feel uncomfortable if others find out that their data is not valid.

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

### **1 INTRODUCTION**

Advances in technology for collecting, storing, and analyzing data 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 challenges in areas such as marketing, homeland security, and molecular biology (Spengler 2000; Shaw et al. 2001; Seifert 2004). In China recently, research data management (RDM) was highly demanded by international Higher Education Institutes (HEI) from 2005 to 2010 (Moe, 2015). The increased breadth and depth of required data has driven the development of new data strategies to efficiently manage and share available data. The governments of Sweden, America and Canada have decided that America and Canada have decided that in all disciplines, all research results funded by public funds are expected to be openly accessed by attaching research data (Borglund and Bogerud, 2020). Policies ensuring research data is available in public archives are increasingly being implemented at the government, funding

agency, and journal levels. This policy is based on the idea that authors need to be ambassadors of our data, to represent them in a way that demonstrates why you are seeking to collect it and what you hope to learn from it. The idea behind open data is that it should make it possible to access, reuse public information for free to create new ideas and innovations.

Slightly different in Indonesia, most research is only concerned with final results, institutions or tertiary institutions are required to convey the content/subject of research only, research data is only for complement (1), namely with administrative reports (2), data that has been collected or processed by researchers are only kept alone (3). In practice, data is still not considered in research either from sponsors or research institutions, the value of research data is still not considered. This causes researchers to be unable to manage their research data properly. According to Childs et al (2014) researchers often store research data on unreliable media for preservation, important data is easily lost. Most researchers when it comes to storing and archiving research data, often preferably on personal PC or laptop hard drives, or external hard drives/USB drives or even the cloud. There are many considerations and arguments given by researchers when they want to share their research data.

The absence of data management regulations and policies makes researchers unable to understand the concepts and processes of RDM. However, the Government of Indonesia is still not aware of the importance of managing research data. Lack of understanding about the importance of data usage, data citation and the importance of data reuse, causes researchers to ignore invitations and inducements to send their research data. This is because most of the research data is still managed by researchers or research groups. One of the problems in librarianship is data management in the library. Awareness about Research and data management has emerged recently.

The Indonesian Institute of Sciences (LIPI) was the first institution to start building work and data management and national data research, namely the National Scientific Repository (RIN) in all disciplines. After that National Library followed to build a repository in the librarianship only. These two national agencies are currently starting to build RDM and repository.

Previous research on open data concluded that the regulations regarding the collection and preservation of research data were unclear (Grants 2017). In addition, archival aspects are rarely considered in research projects, and there is a lack of knowledge about how to preserve research data over time.

Hence, there is a need to better understand the challenges of open research data from an archival perspective. The purpose of this research is to find out to what extent researchers in Indonesia can understand data sharing and its benefits, as well as their willingness to share data.

## 2 METHODOLOGY

In this paper is a qualitative research on acceptance of submission of data researchers. The informants are researchers in Indonesia. Data collection was carried out through online interviews with several researchers in their fields, via video calls and telephone.

Good research is research that has good management including data that can be reused. In Indonesia the data is not well managed. Researchers usually keep their data alone. Good or bad data management depends on how each researcher stores it. When researchers were asked to submit their data, their responses varied—some were willing to share, while others were reluctant to do so. What may have been the reasons behind their willingness and reluctance has yet to be analysed. This research is to find out the reasons for how far researchers know about data management, benefits and also understanding for conveying research data.

## 3 FINDINGS AND DISCUSSION

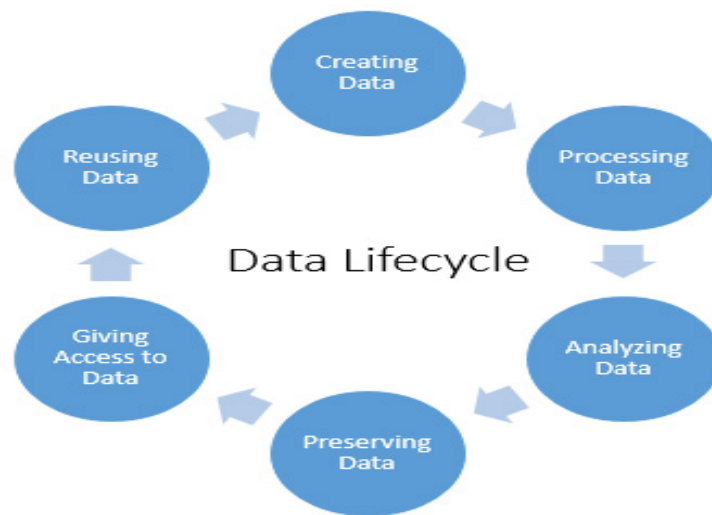
Data used to be collected or produced as part of an academic research process and are now being generated in ever-increasing volumes and in a variety of digital formats that are often rapidly being replaced. (Berman and Cerf, 2013; Borgman, 2012; Pryor, 2012). Research data Can exist in any format in which it is generated, for example: text, numeric, audio-visual, model, computer code, discipline specific, instrument specific. Data is divided into two types, including; (1) primary data (data collected for a particular research problem, using the procedures best suited to the research problem. (2) secondary data (data that has been collected by others, for other purposes, but has some relevance to your research) this practical need is a change in the perception of the value of research data: it has been seen as an asset that must be managed to maintain its value (Higgins, 2012; Lavoie, 2012).

According to Whyte, A., Tedds, J. (2011) "Research data management is concerned with organizing data, from entry into research cycle to disseminating and archiving valuable results. It aims to ensure reliable, up-to-date and innovative verification of results to build existing information." In data management research concerns how you to: Generate data and plan its use (1) Organize, compile and name data, Store it – make it secure, provide access (2), store and back it up, Find sources of information (3), and share with collaborators and more broadly, publish and get citations (4). it focuses on what is needed for validation and reuse.

Good management of data management must consider the life cycle of research data, enabling scientists to reliably re-analyze data and verify results and replicate studies to train new generations of researchers, to create new ideas and innovations (Bertagnolli et al. 2017). Data

sharing has been promoted for various reasons: first, without data sharing it is impossible to verify research results, a key principle of good science (Borgman, 2012). However, it can increase reuse, author visibility, collaboration and research integrity. In addition, it reduces the risk of data loss (1), data leakage (2), copyright infringement (3), and breach of contract (4).

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



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

There are several models that can be used to plan data management activities, for example the DCC Curation Lifecycle Model. The diagram below and the following legend illustrate this life cycle in six stages. The research data cycle needs to be considered by institutions in implementing RDM policies and services, which include: (1) Create; it includes the research design, creating a data collection framework and metadata (2) Process; data entry, validation, description and storage (3) Analysis; interpretation, data derivation, and publication (4) Preservation; metadata, documentary and archive (5) Share; metadata backup, storage, sharing, access control, copyright and promotion (6) Reuse; Data available for discovery and access can be reused by other researchers

Digital Curation: a life cycle approach to managing and preserving usable digital information so a life cycle approach is necessary because: Reliable reuse of digital materials is only possible when materials are curated in such a way that their authenticity and integrity are maintained (Pennock, 2007).

Data is an important economic resource in all aspects, in any discipline including biomedical research (Downey and Olson 2013). Data can be shared and reused in an unrestricted fashion 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 data for granted or allow data to deteriorate and disappear. Regulations regarding data management in an institution need to be implemented to raise awareness of the importance of data storage and data management among researchers. Unfortunately, in Indonesia most researchers, for the most part, have never received training on how to share and communicate about their data, even to their colleagues. In fact, there are many potential benefits from good research data management, other researchers, and the wider community: Increasing research impact through knowledge transfer (1), Efficiency and ease of data control, reduced data loss (2), Research progress through data reuse by international researchers (3), Adherence to policies and expectations of funders and institutions (4), Demonstration of research integrity and validation of research results (5) (Markowitz (2015) *Genome Biol* 16, 274).

Of the 40 researchers contacted, 15 indicated a willingness to participate in the survey. Many informants did not answer all survey questions. Out of 40 researchers, 5 are interested in sharing data. Out of 40, 10 gave fully into their research but without sharing data. They think that the lack of attention in storing research data has a negative impact on managing the management of their personal data, they only focus on research results.

In terms of the level of policy awareness, the researcher shows a low level of awareness. Of the researchers who archived their data, only 5 did so for reuse. They can manage it, Store it, make it secure, provide access, store it, and back it up.

What was also surprising was that several informants admitted that they did not want to share because they were afraid that the original data could be manipulated by others. This is quite challenging, because they don't want other people to know that some researchers make mistakes when conducting research. And the rest have no concern for keeping their data. They cannot properly manage the management of their personal data, but they have no worries when sharing research results. Furthermore, researchers do not regard their research materials as public, but as complementary.

#### **4 CONCLUSION**

Sharing data can increase the returns on research projects by allowing other researchers to carry out secondary data collection, furthering their own exploratory studies. The lack of clarity on

regulations for sharing data makes researchers reluctant to provide data (1) There are no regulations from research institutions (2).

Related to this, the authors provide several recommendations to scientific journal managers, research institutions, funders, and policy makers. (1) For this reason, it is necessary to socialize the data sharing movement both in research institutions and in tertiary institutions.

## REFERENCES

- Borgman CL. 2012. The conundrum of sharing research data. *J Assoc Inf Sci Technol.* 63(6):1059–1078.
- Borgerud, C., Borglund, E. Open research data, an archival challenge?. *ArchSci* 20, 279–302(2020). <https://doi.org/10.1007/s10502-020-09330-3>
- 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>
- Research Data Lifecycle. Retrieve on <https://www.reading.ac.uk/RES/rdm/about/res-rdm-lifecycle.aspx>
- 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>
- Markowitz, F. Five selfish reasons to work reproducibly. *Genome Biol* 16, 274 (2015). <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/zwggk/2012-03/27/content\\_2100852.htm](http://www.gov.cn/zwggk/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
- Whyte, A., Tedds, J. (2011). ‘Making the Case for Research Data Management’. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Available online: [/resources/briefing-papers](http://resources/briefing-papers)
- Pronk TE, Wiersma PH, van Weerden A, Schieving F. 2015. A game theoretic analysis of research data sharing. *PeerJ.* 3:e1242
- Whyte, A., Tedds, J. (2011). ‘Making the Case for Research Data Management’. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Available online: [/resources/briefing-papers](http://resources/briefing-papers)

