

A Preliminary Examination on the Development of Research Data Management (RDM) Policy

Suhaimi Napis¹, Valerie Anak Michael², Yusmadi Yah Jusoh³, Rusli Abdullah³, Fatimah Sidi³, Iskandar Ishak³, Mohammad Hamiruce Marhaban⁴, Yusnita Tugiran⁵

¹Faculty Biotechnology and Bimolecular Sciences, Universiti Putra Malaysia

²Faculty of Design and Architecture, Universiti Putra Malaysia

³Faculty of Computer Science and Information Technology, Universiti Putra Malaysia

⁴Faculty of Engineering, Universiti Putra Malaysia,

⁵Research Management Centre, Universiti Putra Malaysia

(suhaimi@upm.edu.my, valerie19mic@yahoo.com.my, yusmadi@upm.edu.my, rusli@upm.edu.my, fatimah@upm.edu.my, iskandar_i@upm.edu.my, mhm@upm.edu.my, yusnita@upm.edu.my)

Article Info

Volume 81

Page Number: 3114- 3120

Publication Issue:

November-December 2019

Abstract:

Data sharing is an important pre-requisite towards enhancing knowledge in science and technology in recent years. The need to systematically collect, collate and manage research data will further facilitate knowledge archiving for future reference and re-use. This paper describes the development of a policy framework for data repository in one of Malaysian public university with a focus on the awareness of data sharing among researchers. Questionnaires were distributed to 164 respondents focusing on the benefit of data sharing, lack of data sharing requirement and types of data to be included in data repository management. From the finding, the respondents are willing to share and provide data to the other researchers provided there is a proper policy document. This research also discusses the recommended guidelines for the data repository policy and also on how to promote data sharing among researchers.

Article History

Article Received: 5 March 2019

Revised: 18 May 2019

Accepted: 24 September 2019

Publication: 14 December 2019

Keywords: Data sharing, policy and governance, data repository, Research Data Management (RDM)

1. INTRODUCTION

Generally, in the university surrounding, there are diversity of data and information has been contributed and practices by the researchers for the purpose of research project. The purpose of this research is to propose a policy of data repository in one of the public university. The purpose is to study the awareness of data sharing among the researcher. Basically, the data is specialized in statistics, numbers of digital images, sound recordings, musical scores, transcripts of interviews, computer source code, simulations, survey data and classified work observation with proper notes. In addition, the researcher usually stored their data in hard disk, email and Cloud storage which are

provided by Google and Amazon. Once the researcher wants to share their data, they usually exchange data through email and shuffle the data among the project teammate.

2. LITERATURE REVIEW

The accessibility of research data has a huge prospective for scientific evolvement. Based on The University of Salford (2019) stated that research data are involved in quantitative information or qualitative data collected by researchers where the researcher work included experimentation, observation, interview or other techniques. According to ICSSR Data Service: Indian Social Science Data Repository, India (2015) had listed

social & statistical data of numerous nationwide assessments on industries, education, employment and unemployment, domestic consumer expenses, enterprise landholdings, management data, submitted data in CSV, EXCELformat and etc. ICSSR Data Service also offered quantitative and qualitative scientific research data created by many social science institutes, research resources; the data including resources may be interviewer guides, questionnaires, data collection method, codebooks, database dictionaries, project summaries or description, and bibliographic of publications relating to the data. Nanyang Technological University (NTU), Singapore (2019) highlighted all research proposals must include data management plan (DMP) that records the intention on how research data arising from the research project will be managed, used and shared.

In the context of data sharing, Sayogo and Pardo (2013) stated that the ability of scholars to organize their documents and contract with the matters of shared data will encourage their data value, criteria and security, moral and supervise use of shared data will affect their tendency to share their data. Meanwhile, Kim and Zhang (2015) stated that the interpretation toward data sharing by the researcher can be characterized as a beneficial or unbeneficial assessment of data sharing practice. In addition, researchers are more expected to reflect sharing their data with others if they are in benefit of sharing data. At this point, there is a researcher who willing to share the research data by considering the service in distributing knowledge with other researchers. And so when it comes to open data, then it would be a good benefit to the other researcher who interested to make references about the new knowledge.

Meanwhile, some researcher chooses to make their data as confidential. According to Kim and Adler (2015) stated that the data sharing from the social science is not actively applied by the researcher because of

limitation from the framework by the organization, responsibilities and individual reasons. At this point, the cluster of the data really affected the data sharing and at the same time, the awareness from an institution is crucial especially on the trustworthiness toward institution and duties. Meanwhile, Higgins, Taylor, Lisboa and Arshad (2013) stated that it is crucial to share among public sector organizations by applied efficient method in order to acknowledge the prospective of data sharing and to create a practical method to data sharing that is lawfully acquiescent. Aside from the public sector, the academic institution also needs a strict law in order to protect the research data if the institution implements the data sharing. For example, Rice, Ekmekcioglu, Haywood, Jones, Lewis, Macdonald and Weir (2013) mentioned that at the University of Edinburgh has creating responsiveness plan by means of collaborating data which is continuing research data management work with the higher institution and funding policies. Aside from sharing the data, other elements such as to build a system is also important so that the value of data sharing could give benefit to other researchers. Based on MacMillan (2014) stated that while researchers surely value access to others' data and share data more easily, at the same time there is still work to do to change engrained system, and construct access to others' data, and convey a request to share data more easily. So, it is important to established workflows and builds value for data sharing.

At this point, a university must construct a plan to the university's researcher so that the researcher could aware about the policies offered by the university. Fecher, Friesike and Hebing (2015) stated that policymakers need to comprehend and identify the involved parties and their perceptions in order to develop significant policies for data sharing. They also highlighted about the research policies should work towards an effective exchange system by sharing as much data as possible. By this way, the researcher would understand the

benefit of exchanging data and also save time and money when it comes to extract data from other sources. Based on Harvard Research Data Security Policy (2017) stated that when the data security level has been established, researchers are having control in term of making and preserving data documentation, implementing the security controls corresponding to the requirements of the data security level, expanding and following a data security plan and operates over the course of researchers' projects.

3. MATERIALS AND METHODS

Most of the respondent were university's staff which were lecturers, research officers and postgraduate students. This research had been conducted in one of the public university in Malaysia. The purpose of distributing this questionnaire is to develop data management policy in one of the public university in Malaysia. There were 4 questions had been asked, the questions were highlighted about an opinion on the benefit of offering information about data, materials and program sharing policies provided by the researcher, the lack of data sharing requirement and type of researcher data to be included in data repository management. The researcher had applied cluster sampling as a method to select types of respondents to response the questionnaires. Based on Neuman (2003), the cluster sampling method is easy to be used from usefulness viewpoint and this method involves identification a cluster of participants that portray the people are spotted and included in the sample. The researcher had distributed 164 questionnaires to the respondents. Unfortunately, some of the

questions were not fully answered by the respondents. So, the amount of respondents who answer each set of questions were imbalanced. These result of the questionnaires came from the following faculties: Agricultural, Forestry, Veterinary Medicine, Economic Management, Engineering, Educational Studies, Science, Food Science Technology, Human Ecology, Modern Language, Design and Architecture, Medicine and Health Science, Computer Science, Biotechnology, Environmental Studies and Agriculture and Food Science. Aside there are also 7 institutes included in these results. The duration for the respondent to answer the questionnaires was from 25 September 2018 until 10 October 2018. The researcher had distributed the questionnaire by hand to the respondents and Google Form.

4. RESULTS AND DISCUSSION

In this section, there are 4 questions been asked which are consisted benefit of offering information, data, materials and program sharing policies provided by the researcher, the lack of data sharing requirement and types of research data to be included in data repository management. Based on the result of the benefit in having a service offering information about research data policies, there were 161 respondents had given their responses which were 77% agreed that the service of offering information give benefit to the research data policies. Only 18% of the respondents do not know the benefit of it. It shows that some of the respondents are still doubtful.

Table 1: Data, materials and program sharing policies provide by the researcher

Data, materials and program sharing policies provide by the researcher.	Response
Yes	70.4%
No	11.1%
Do Not Know	18.5%

Based on Table1, it is about the possibilities of providing data, materials and program of sharing policies. Out of 164 respondents, there were 70.4% of respondents mentioned 'Yes' about researcher should provide data, materials and program sharing policies. Unfortunately, 18.5% of the respondents mentioned that they do not know about the benefit of providing data and sharing policies. Based on Figure 1, this bar graph represents the reason of the researchers do not want to share their data. The main

reason of researchers to not sharing their data because the data is treated as private ownership and there were 57.4% of respondents give their feedback about it. Then, 48.1% of respondents choose lack of awareness and followed by no established policy (47.5%). Another important reason was fear of misuse of the data also as part of the reason for not to share the data. So, this has affected the researcher not interested to share their data and they really concerned about the authenticity of the data.

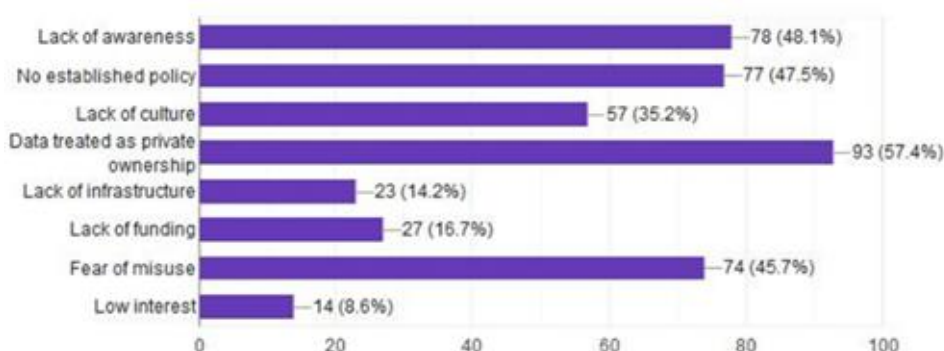


Figure 1: The lack of data sharing requirement

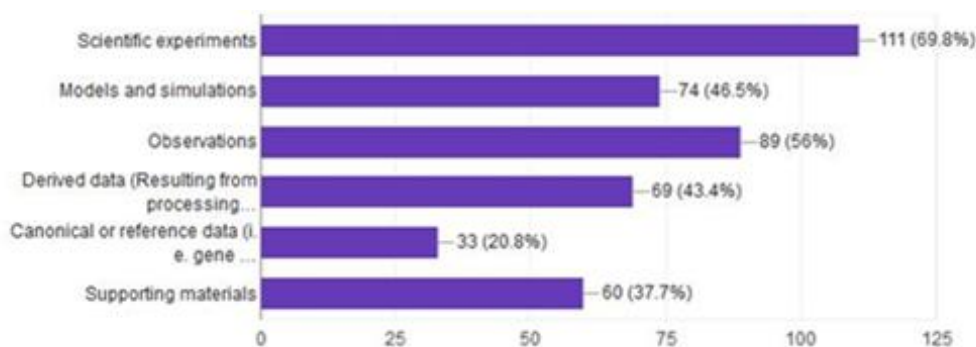


Figure 2: Types of research data to be included in data repository management

Based on Figure 2, only 125 respondents had answered this question's section. So, the scientific experiments were the highest research data to be included in the data repository where the experiments data amount was 69.8%. It shows that the experiments data are filed in huge amount and that mean the experiment data need to save in secure data repository. The scientific data portrays the raw data from the result of the experiment, images of the experiment and so to the progress report. Same goes to the observation, the items are notes, images and

reports. Then, observations data were the second highest research data, which was 56%. Next, the third highest was 46.5% which were from the models and stimulations. The lowest data was canonical or reference data where the result of this kind of research data was 20.8%. Based on the result of data sharing practiced in one of the public university, it shows that the researcher data is serving as confidential and keep among team members only. So, the research data such as data analysis, results and finding would not be reveal until the project is done.

In discussing recommended guidelines in RDM policy, the public university should apply a policy that could give a guideline to the researcher about data sharing among researchers. The result from Figure 3 indicated 70.4% of the respondents gave their feedback that agreed of service on research data policy in the university. In this situation, a policy can make the researcher feel secure to exchange their research data. At the same time, they also allow sharing their data with other scholars if the finding could contribute to society. Based on the result from Figure 3, it mentioned that no established policy as a guideline to the researcher. This situation has suggested that an effective policy should be constructed and clearly educate the researcher to understand data sharing. A policymaker should construct a policy that can attract and understood by the researcher. According to Latham (2017) stated that many higher institutions are practicing RDM for their scholars and libraries, so from this chance, it connecting to present this application once the policy has already been agreed. Similarly, the university should generate data repository system in order to store and share open data with other researchers. On the other hand, this policy should suit all fields such as science, science social, art and humanities. As reflect the lack of data sharing requirement from the Figure 3, the researcher chooses to treat the data as private and lack of awareness toward the data sharing, this situation has given a challenge to the policy maker to construct a policy that can educate and trust data repository system. The security guarantee is a must to highlight in the policy so that the data is in high security. Therefore, it is important to list a clear guideline to the researcher to understand the purpose of data repository policy. For example, Monash University (2017), Harvard University (2017) and University of Birmingham (2019) have listed security guarantee is important to practice in the data repository policy. In RDM, for example, Monash University (2017) is committed to achieve research data management standards for the purpose of

secure data retention and to optimizing the benefits of RDM through collecting, storing and making research data accessible that it can be used in upcoming research by the members. The RDM should be focusing on protecting the research in high ethic which is protecting the rights, dignity, safety, research subject and privacy of the community, the welfare of animals and the integrity of the environment. So, through this policy, it is clear that the awareness toward the ethics in data security could raise the researcher's trust about the RDM policy in the university. In term of acknowledgment aspect, based on University of Birmingham (2018) mentioned that all users of the Research Data should formally cite the data they use, whether the data referred was to generate as an inherent part of research, or whether the primary aim of the research was to create research data sets for the use of others. By this way, a person who likes to refer the data can formally cite it and the owner of the research will not feel uncertain to their research data.

Once the research project is completed, the result of the researcher is ready to be published and share the new knowledge with other researchers. If the researcher chooses to make the data is open access, then it can give an advantage to the other researcher especially when it comes to the reference. Aside, there are also some researchers do not intend to share the research data in the form of open access. The researchers from the medical and science fields are very strict toward their data and these data are known as confidential data. Apart from that, a good policy is offering a good data repository system that can archive research data. The purpose is to preserve important research data. For example, data from the scientific experiment data is crucial to archive and preserve it for the long term. In addition, Zhu, Marciano, Moore, Herr and Schulze (2012) stated that to check the data truthfulness inside repositories, a certain digit is assisted for the purpose of the maintenances and to keep a good copy. This strategy gives a

huge impact to other researcher and they can expand the research. In short, as if there is the preservation and archiving in the data repository, the researcher able to reuse, data sharing and access data. At the same time, as the system data repository is centred in the university itself, so it could save lots of money if other researcher wants to download the research data. The most important in the data repository policy is to make sure that the system is user-friendly to the users.

5. CONCLUSION

In conclusion, the guidelines of proposed data repository policy have highlighted that the data sharing, data security, acknowledgement and data archiving are crucial elements that should be adopted by researchers and stakeholders. Research data policy that has been developed will further facilitate researchers in different fields to carry out interdisciplinary research within a trusted ecosystem whereby collaboration and interaction among each other will be governed by a defined policy. In this way, researchers will have greater freedom and peace of mind which will lead to greater research outputs.

6. ACKNOWLEDGEMENT

The authors would like to express gratitude for the financial support provided under the Putra University Grant Scheme, Grant cost centre: 9558500

References

1. Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing?. *PloS one*, 10(2), e0118053. Retrieved from <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118053>
2. Harvard University. (2017). Harvard Research Data Security Policy. Retrieved from <https://vpr.harvard.edu/pages/harvard-research-data-security-policy>
3. Higgins, E., Taylor, M., Lisboa, P., & Arshad, F. (2014). Developing a data sharing framework: a case study. *Transforming Government: People, Process and Policy*, 8(1), 151-164. Retrieved from <https://www.emeraldinsight.com/doi/full/10.1108/TG-02-2013-0007>
4. Monash University. (2017). Monash University Policy. Retrieved from https://www.monash.edu/__data/assets/pdf_file/0011/797339/Research-Data-Management-Policy.pdf.
5. Indian Council of Social Science Research. (2015). ICSSR Data Service. Retrieved from <http://www.icssrdataservice.in/>
6. Kim, Y., & Adler, M. (2015). Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management*, 35(4), 408-418. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0268401215000432>
7. Kim, Y., & Zhang, P. (2015). Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Library & Information Science Research*, 37(3), 189-200. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0740818815000584>
8. Latham, B. (2017). Research data management: Defining roles, prioritizing services, and enumerating challenges. *The Journal of Academic Librarianship*, 3(43), 263-265. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0099133317301453/pdf?md5=f67c98c4b1e776c3538c718462282a94&pid=1-s2.0-S0099133317301453-main.pdf>
9. MacMillan, D. (2014). Data sharing and discovery: What librarians need to know. *The Journal of Academic Librarianship*, 40(5), 541-549. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0099133314000950>

10. Nanyang Technological University (NTU).(2019). NTU Research Data Policy. Retrieved from <https://research.ntu.edu.sg/rieo/RI/Pages/Research-Data-Policies.aspx>
11. Neuman, W. L. (2003). Qualitative and Quantitative Research Designs.Social Research Methods.Qualitative and Quantitative Approaches.Fifth Edition.pg 137-168. United States of America. Library of Congress Cataloging-in Publication Data.
12. Rice, R., Ekmekcioglu, Ç., Haywood, J., Jones, S., Lewis, S., Macdonald, S., & Weir, T. (2013). Implementing the research data management policy: University of Edinburgh roadmap.*International Journal of Digital Curation*, 8(2), 194-204. Retrieved from <http://www.ijdc.net/index.php/ijdc/article/view/8.2.194> et al. (1987).
13. Sayogo, D. S., & Pardo, T. A. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, 30, S19-S31. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0740624X12001529>
14. The University of Salford.(2019). Research Data Management.Introduction.Retrieved from<http://www.salford.ac.uk/research/research-data-management/introduction>
15. University of Birmingham.(2018). Research Data Management Policy. Retrieved from <https://intranet.birmingham.ac.uk/as/libraryservices/library/research/rdm/Policies/Research-Data-Management-Policy.aspx>
16. Zhu, B., Marciano, R., Moore, R., Herr, L., & Schulze, J. (2012). Digital repository: preservation environment and policy implementation. *International Journal on Digital Libraries*, 12(1), 41-49. Retrieved from <https://link.springer.com/article/10.1007/s00799-012-0082-3>