Research Data Management in Supporting Knowledge Sharing among Univesity Researchers

[1]Yusmadi Yah Jusoh, [1]Rusli Abdullah, [1]Fatimah Sidi, [1]Iskandar Ishak, [2]Suhaimi Napis, [3]Mohammad Hamiruce Marhaban, [4]Yusnita Tugiran, [1]Nur Ilyana Ismarau Tajuddin

[1]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
[2]Faculty of Biotechnology and Biomolecular Science, Universiti Putra Malaysia
[3]Faculty of Engineering, Universiti Putra Malaysia
[4]Research Management Centre, Universiti Putra Malaysia
{[1]yusmadi, [1]rusli, [1]fatimah, [1]iskandar_i, [2]suhaimi, [3]mhm, [4]yusnita}@upm.edu.my,
[1]ilyanaismarau90@gmail.com

Abstract:
Research data management is an important concern for institutions and several platforms to support data deposits have emerged. Thus, many research institutions have established or plan to establish research data curation services as part of their Institutional Repositories (IRs). In this paper we start by viewing the current practices in the data management workflow in one of public university in Malaysia. A survey was conducted using questionnaire for data collection among staff, researcher and postgraduate students from Public University in Malaysia. The finding shows most respondents have awareness in research data management. The finding also provided useful information to a better understanding regarding research data management. Limitation and suggestion for future research are discussed.

Keywords: research data, data repository, data management, data sharing.

1. INTRODUCTION

Research data management has gained increasing recognition for its value an importance among funding agencies and research institutions, as evidenced by the fast growth of data repositories at disciplinary community and institutional levels. The growth of data in general and scientific research data in particular, has been driven by a number of social and technological factors. New technologies have made it possible to gather data more quickly and cheaply than ever before. Not only can we generate data more quickly and cheaply than ever before, but we can also afford to store it, as data storage costs have also dramatically decreased in recent years. Cloud storage solutions like those provided by Google and Amazon make it easy to store unprecedented amounts of data. The purpose of this paper is to explore the practice of data management among researchers in the university.

2. RELATED STUDY

In a work done by (Lind, 2016) which focuses on administrating and managing a database from collaboration, the steps taken were enough to provide a coherent whole of merging distributed dataset. By assembling and build a standardization, the data that has been taken into account will be marked with versions before incorporating new data types. In order to have it accessible to others, the data access will be granted to the internal partners. Gellerman, Svanberg and Barnard(2016) recommended a set of guidelines through Data Sharing Framework (DSF) that facilitate easier data sharing. This is to make light of the possibilities for a global scale data sharing and re-use of data. Collaborative researches increases the value of products, and to increase the transparency, badges to mark the report on usage of data sharing enables comprehensive metadata that are sufficient to enable data interpretation and repurposing (Michener, 2015; Kidwell et al., 2016). Increasing in popularity of

usage, data sharing has evolve to slowly become a common practice.

We focuses on exploiting the good in having cloud storage to manage research data and allowing accessible resources for collaboration due to low cost and easy storage. However, there are some researchers that still refuses to migrate to cloud computing and collaborate due to some issues. (Alkhalil, Sahandi, & John, 2017) listed out the concerns that varies based on the context of existing cloud computing, involving the cost, performance, configuration, legal implications and data ownership, access difficulties, expertise level, privacy and confidentiality. Following the varying cultures and behaviors of willingness to data sharing, (Tenopir et al., 2015) shares that the perceptions differs across age-groups, geographical regions, and subject disciplines which should be understand to be able to promote data sharing, given the needs of different research communities.

Giving access to librarians, IT staff and research administrators that relate to the research would be able to support and provide a structured framework for the data management (Cox & Verbaan, 2016)(L. Federer, 2016) . Professional services staffs are able to conceive their current relation to the research and offer expert advice, infrastructures such as library collection or archive services, introducing researchers to funders and relieving researchers of administrative burdens. In conjunction to the roles played by the staffs, research data management (RDM) programs would greatly help academic library to acquire and enhance their skills to assist researchers proposals (Latham, 2017)(Pinfield, Cox, & Smith, 2014)(L. M. Federer, Lu, Joubert, Welsh, & Brandys, 2015). It helps to clarify the different issues that are involve in the large numbers of factors influencing the layers of activities in the data management of researches. The use of RDM drives the roles of different actors in the development of research, making the whole thing manageable and well-organized.

IT investment is going towards managing and maintaining big data. The challenges that are imposed by big data would be able to manage through Scientific Data Lifecycle Management (SDLM) model that includes all major stages and reflects specifics in data management (Demchenko, Grosso, De Laat, & Membrey, 2013). SDLM can be implemented using cloud based infrastructures services which tailored to support the workflow demands. (Davidson, Jones, & Molloy, 2014) introduced the DCC's DMPonline, a tool to help researchers write Data Management Plan (DMP) in order to simplify and enhance the management and reuse of data for a longer-term retention. This is important to hold the potential value of reusing data by organizing it to be more visible and searchable.

3. MATERIALS AND METHODS

Data for the study were collected through e-mail survey, online survey (Google Form), WhatsApp, and Telegram. The period of collection were done from October 2018 until November 2018. The questionnaire is a web based survey and was sent through emails of researchers in the university. Selection of respondents are based on the involvement of academic staff and postgraduates students in research projects. The method used in collecting the data is called cluster sampling. With cluster sampling, the population is divided into separate groups based on faculties and institutes

4. RESULTS AND DISCUSSION

From the survey, we manage to collect about 164 respondents from one of public university in Malaysia that are made up of 31.7% senior lecturers, 7.7% associate professors, 11.6% postgraduate students and 9.1% research officers. Table 1 below summarize the total respondents' group involvement in the survey.

Table 1.Summary of the total respondents' group involvement in the survey

Position	Percentage
Senior lecturer	31.7%
Associate Professor	17.7%
Postgraduate student	11.6%

Research Officer	9.1%

The result of research data management in supporting data and information sharing among university researchers' surveys collected about 164 data entry. 10 questions are chosen to depict the goal in this survey.

Figure 1 below shows a bar graph that illustrate the number of respondents that collect/create/analyze/ store or reuse their research data and their willingness to share data with external entities.
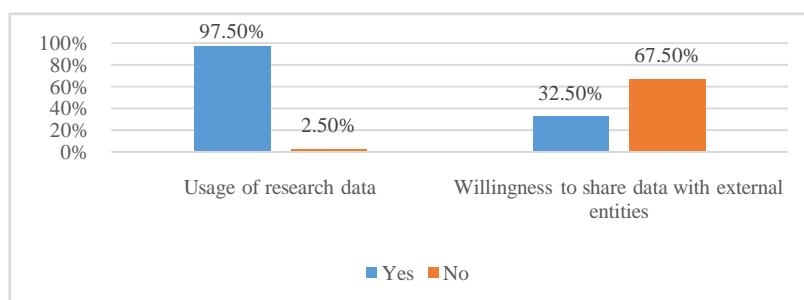

Figure 1. Usage of data research and willingness to share data to external entities

Most of the respondents (97.5%) collect, create, analyze, store and reuse their research data. Meanwhile, only small amount of respondents (2.5%) does not.67.5% of the respondents are willing to share their data with external entities. Meanwhile, the other 32.5% are unwilling to share their research data.
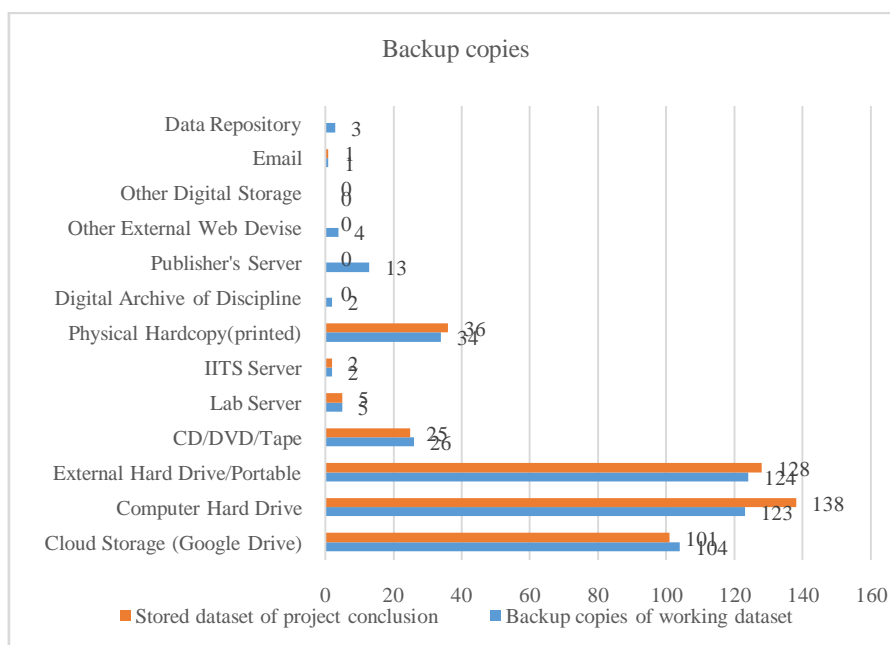

Figure 2.Backup copies of working dataset and the conclusion

Referring to figure 2, a total of 124 respondents (75.6%) respondents stored their backup copies of working dataset during data collection in external hard. Meanwhile, 123 respondents (75%) preferred computer hard drive and 104 respondents (63.4%)chose cloud storage such as Google Drive. The list continued with Physical hardcopy (printed) with

34 respondents (20.7%), CDs, DVDs, Tape with 26 respondents (15.9%) and publisher's server with 13 respondents (7.9%)

A total of 138 respondents (84.1%) respondents stored their dataset of conclusion for their project in computer hard drive. Meanwhile, 128 respondents (78%) preferred external hard drive and 101 respondents

(61.6%)chose cloud storage such as Google Drive. The list continued with Physical hardcopy (printed) with 36 respondents (22%),

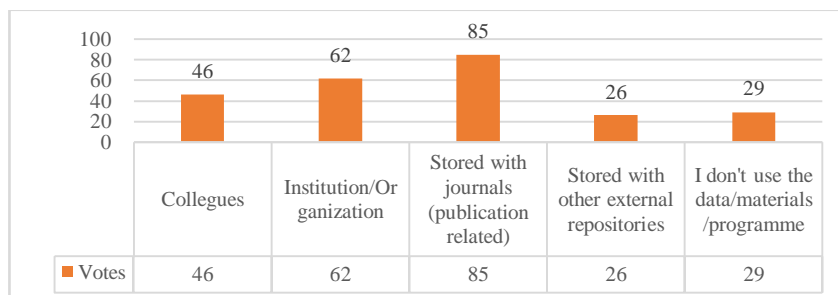CDs, DVDs, Tape with 25 respondents (15.2%) and lab's server with 5 respondents (15.2%)

	Colleges	Institution/Organization	Stored with journals (publication related)	Stored with other external repositories	I don't use the data/materials/programme
Votes	46	62	85	26	29

Figure 3.Location/ access for the research data

There are a total of 85 respondents (54.4%)locate/access the research data/materials/programs of other researchers through storing it with journal (publication related). 62 respondents (39.7%) chose institution / organization, followed by colleagues with 46 respondents (29.5%). 29 respondents (18.6%) do not use the data/materials and programs of other researchers and only 26 respondents (16.7%) chose to store it with other external repositories.
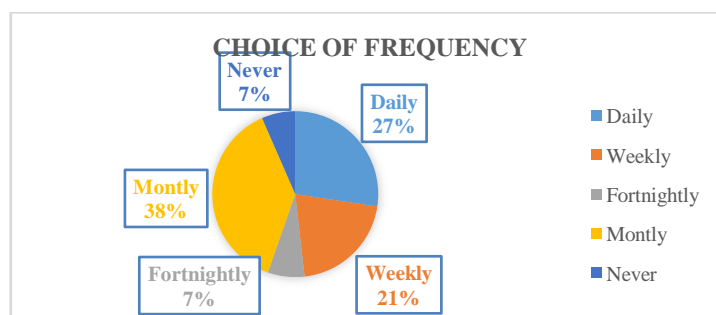

Figure 4. Frequency of back up working data

Referring to figure 4, 64 respondents (39.5%) backup their data monthly. Meanwhile, 46 respondents (28.4%)chose to do it daily, 35 respondents (21.6%) with weekly backup, 12 respondents (7.4%) with every fortnight, and 11 respondents (6.6%) chose to never backup their working data.
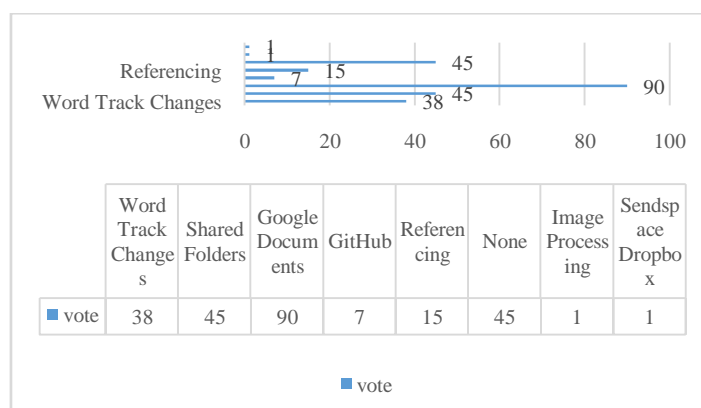

	Word Track Changes	Shared Folders	Google Documents	GitHub	Referencing	None	Image Processing	Sendspace Dropbox
vote	38	45	90	7	15	45	1	1

Figure 5: Software to collaborate with others

Google Document is a popular choice used for collaboration with other people, where based on this bar graph, 90 respondents (54.9%) chose this option. 45 respondents (27.4%) preferred Shared Folder, interestingly, the other 45 have not use any software to collaborate with other people. Word Track Changes is chosen by 38 respondents (23.3%), 15 respondents (9.1%) chose Referencing and the other 7 (4.3%) chose GitHub.
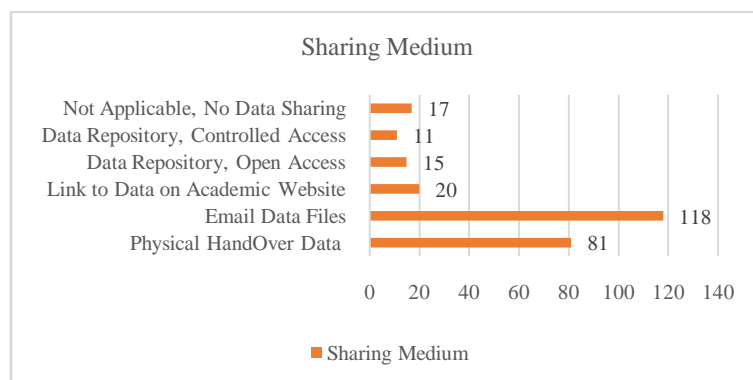

Figure 6. Sharing research data with others

Email data files is a popular choice used for sharing research data with other people, based on this bar graph, 118 respondents (72.4%) chose it. 81 respondents (47.9%) preferred physical hand over through USB, CD, or Portable hard drive. 20 respondents (12.3%) link to data on academic website to share with other people. 17 of them (10.4%) chose to not share. Data repository with open access is chosen by 15 respondents (9.2%), and the other 11 (6.7%) chose data repository with controlled access.

Based on the survey result, we can conclude that respondents have awareness in managing their research data, but not quite updated with the enhancement and innovation in securities and storing their complete backup data. As research data are extremely important and not to be taken lightly of, we should have a change of phase to increase the usage of data repositories that can be customize by universities to cater researchers' needs and demands. As per the data sharing, there is a need in utilizing the services provided by cloud computing to allow secure data storage and easy accessibility for collaboration. The result also reported on the entitlement to the privacy and confidentiality of data and research that emasculate the accessibility of the data to other people

5. CONCLUSION

In conclusion, research data management in supporting data and information sharing is essential as vast quantities of data are being produced in a wide variety of forms at a rapid rate in universities. This study has illustrated how different researchers are adapting to deliver values to the organization in particular and research community in general. While we should evolve the norm in the data sharing issue, effective methods to encourage data sharing must be taken into account seriously.

6. ACKNOWLEDGMENT

7. REFERENCES

1. Alkhalil, A., Sahandi, R., & John, D. 2017. An exploration of the determinants for decision to migrate existing resources to cloud computing using an integrated TOE-DOI model. *Journal of Cloud Computing*, *6*(1). https://doi.org/10.1186/s13677-016-0072-x
2. Cox, A. M., & Verbaan, E. 2016. How academic librarians, IT staff, and research administrators perceive and relate to research. *Library and Information Science*

Research, *38*(4), 319–326. https://doi.org/10.1016/j.lisr.2016.11.004

3. Davidson, J., Jones, S., & Molloy, L. 2014. Big data: the potential role of research data management and research data registries. *Ifla*, 1–11.

4. Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. 2013. Addressing big data issues in Scientific Data Infrastructure. *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013*, (July 2018), 48–55. https://doi.org/10.1109/CTS.2013.6567203

5. Federer, L. 2016. Research data management in the age of big data: Roles and opportunities for librarians. *Information Services and Use*, *36*(1–2), 35–43. https://doi.org/10.3233/ISU-160797

6. Federer, L. M., Lu, Y.-L., Joubert, D. J., Welsh, J., & Brandys, B. 2015. Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff. *Plos One*, *10*(6), e0129506. https://doi.org/10.1371/journal.pone.0129506

7. Gellerman, H., Svanberg, E., & Barnard, Y. 2016. Data Sharing of Transport Research Data. *Transportation Research Procedia*, *14*, 2227–2236. https://doi.org/10.1016/j.trpro.2016.05.238

8. Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Nosek, B. A. 2016. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLoS Biology*, *14*(5), 1–15. https://doi.org/10.1371/journal.pbio.1002456

9. Latham, B. 2017. Research Data Management: Defining Roles, Prioritizing Services, and Enumerating Challenges. *Journal of Academic Librarianship*, *43*(3), 263–265. https://doi.org/10.1016/j.acalib.2017.04.004

10. Lind, E. M. 2016. Unified data management for distributed experiments: A model for collaborative grassroots scientific networks. *Ecological Informatics*, *36*, 231–236. https://doi.org/10.1016/j.ecoinf.2016.08.002

11. Michener, W. K. 2015. Ecological data sharing. *Ecological Informatics*, *29*(P1), 33–44. https://doi.org/10.1016/j.ecoinf.2015.06.010

12. Pinfield, S., Cox, A. M., & Smith, J. 2014. Research data management and libraries: Relationships, activities, drivers and influences. *PLoS ONE*, *9*(12), 1–28. https://doi.org/10.1371/journal.pone.0114734

13. Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., … Dorsett, K. 2015. King James Version: Genesis: Genesis Chapter 9, 1–24. https://doi.org/10.5061/dryad.1ph92