International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

Influencing Factors in Determining Research Data Repository Infrastructure for Research Data Management

Iskandar Ishak, NurIlyanaIsmarauTajuddin, Yusmadi YahJusoh, Fatimah Sidi, Rusli Abdullah, HamiruceMarhaban, YusnitaTugiran,YushaidaYusof

Abstract: ***Increasing volumes of data are rapidly being produced by researchers with the advancement of digital technologies. In order to manage these data, a suitable research data repository infrastructure is needed by the higher learning institutions. Apart from storing the data, these data repository need to support the research data life-cycle that include the tasks of data creation, processing, analysis, preservation, access and reuse. The objective of this research is to deeply investigate the influencing factors fordata repository infrastructure in managing research data. A systematic literature review is conducted to perform the investigation where research papers are searched over three electronic journal databases. Selected papers are then analysed and a quality assessment has been conducted to identify the relevant infrastructure for research data repository. As a result, we identified the important components of research data repository infrastructure development.***

Keywords: ***Research Data, Data Repository, Data Infrastructure, Data Management***

I.INTRODUCTION

In recent years, there have been numerous research data repository projects that have been developed either by research centres or institutions of higher learning to manage research or scientific data. The purpose of managing these data is to support research activities based on the research data lifecycle (Wissik&Durco, 2015). The advancement of cloud storage and big data computing even add the exhilaration in creating data repository for research data.However, research data repositories are still in its immature stage where there are no actual standards in determining a specific infrastructure for a good and sustainable research data repository.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
Iskandar Ishak, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
NurIlyanaIsmarauTajuddin, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
Yusmadi YahJusoh, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
Fatimah Sidi, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
Rusli Abdullah, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
HamiruceMarhaban, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
YusnitaTugiran, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
YushaidaYusof, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia

Many literatures discussed about research data repository, but very fewhad put emphasis specifically on the standard or general infrastructure for research or scientific data repository.

Therefore, this paper intends to contribute to look into the infrastructure perspective in which, it tries to determine the factors that influenced institutions or organizations in developing infrastructure for research data repository. In particular, the purpose of the paper is to shed some light on the following questions regarding research date repository among institutes of higher learning and research institutes:

a) What are the common elements in research data repository infrastructure?

b) What are the influencing factors in determining research data repository infrastructure?

The paper is structured as follows. Section 2 describes the definitions of important terms in our research. Section 3 illustrates the review methodology that has been implemented and applied in this research. The results are presented in Section 4. Section 5 summarizes the main findings of this research and concludes the study.

II.OVERVIEW OF RESEARCH DATA REPOSITORY

Research data and Research Data Lifecycle

Research data is an important component in research data lifecycle where it represents information and knowledge about a particular research. It is a key element in research organizations as well as institute of higher learning as it promotes and support research or scientific data sharing among researchers (Austin et al. 2015).University of Edinburgh defined research data as data that is collected, observed, or created for the purpose of analysis to produce original research results (2011).Some literatures such as in Demchenko et al. (2013) and Assante et al. (2016) used the term scientific data to describe research data. Research data lifecycle describes how research data is being used and managed by researchers or data owners. Based on Cortietal. (2014), research data lifecycle include the activities of planning, acquiring data, processing data, analysing data, preserving data and sharing the data.

Research Data Repository and Infrastructure

According to Pampel et al. (2013), research data repository is defined as researchdata repositories need to serve different academic and disciplinarycommunities with

their respective concepts of research data.Through data sharing, it increases the ability to reproduce results, replicate findings, andgenerate new knowledge among researchers thus speed up research practices inter-institution across borders.There is no exact definition in terms of research data infrastructure, but Gray et al. (2018) mentioned that data infrastructure should comprised of relations of databases, software, standards, classification systems, procedures, committees, processes, coordinates, user interface components and many other elements in making and using of research data.

III.REVIEW METHODOLOGY

We conducted our systematic review using guidance from Okoli et al. (2010). The review process contains three stages namely planning, executing and reporting.

The planning stage activities include identifying specific research context, defining reviewing protocols, and constructing research questions. In this review, the issues related to the perception of infrastructure requirement in supporting research data repository is determined to be the main context of the research. A review protocol is also defined in this stage based on the input from the research team that included researchers specializing in research data management.

In the execution stage, the search is conducted on three leading electronic journal databases namely Web of Knowledge, Science Direct and Scopus. A set of criteria has been established in filtering the literatures. In terms of criteria, only literatures from high-quality sources were used in this research. Only English-language peer-reviewed journal and conference papers were chosen forthis research where dissertations, book reviews, case studies and books were excluded. The search was limited to papers published between 2000 and 2018. The search analyse titles and abstracts for a number of related keywords and phrases: *"infrastructure "*, *"infrastructure requirement model"*, *"data sharing"*, *"data repository"*, *"information sharing"*, *"data management"*, and *"big data"*. We used three couples of combinative keywords: (infrastructure AND data repository), (infrastructure requirement model AND information sharing) and (big data AND data management).The article included only if it meet the following criteria:

i. Firstly, its focus was primarily on research data management infrastructure requirement model.

ii. Secondly, the area investigated included organization and university environment.

The articles that fulfilled the above-mentioned criteria are thenanalysed by the authors for additional studies to meet the inclusion criteria for the review. In this stage, the duplicated articles were removed. Further, the articles were screened for relevance, primarily based on the title and abstract.

Table. 1 Quality Assessment Criteria

No.	Item	Answer
Q1	Is there a clear description of the aims and objectives of the investigation?	Yes/No/ Partially
Q2	Is the paper explained the method of analysis pertinent and adequately?	Yes/No/ Partially
Q3	Is the paper supported by primary data?	Yes/No/ Partially
Q4	Is the paper explained the infrastructure in detail?	Yes/No/Partially

In terms of quality assessment, the quality assessment was formulated to evaluate the completeness and advantageous for data extraction. These four questions (Q1-Q4) are presented in Table 1. Each question has only three answer options: Yes=1; Partially =0.5; and No=0.The characteristics of the literatures and the results of the assessment are described in detail in reporting stage presented in the following section.

IV.RESULTS AND DISCUSSION

The literature search identified a total of 261 references. Then, the title, abstract and brief content of these selected papers are evaluated. Based on the evaluation, 41 papers were identified were then filtered by applying the quality assessment criteria. In the very final round, only 23 papers out of 41 papers (55%) were accepted for data synthesis of evidence after executing exclusion criteria (Table 2).

Table 3 describes the type of research for the finalized literatures. A total of 15 or 65.2% literatures are of conceptual type of research. This shows that most of the literatures are basically describing the researchers plan and design about their research data infrastructure. Another type of research that has been categorized for the literatures is empirical type that has 6 literatures (26.1%). The least type of research is case study type with 2 literatures (8.7%). In terms of organizations involved, University scores the highest number of organization in the literatures with 12 literatures or 52.2%. Library and Research Institute both have 5 literatures (52.2%) and Government Agency only has 1 literature (4.4%).

A quality assessment has been performed to the 23 qualified papers based on the criteria set during the planning stage and the summary of the assessment is shown in Table 4. Based on the results, only 5 papers have met all the assessment criteria with the total 4 marks. All papers managed to meet Q1 criteria which means, these papers have clear description in terms of its aims and objectives for research data infrastructure. For Q2, only 18 papers met the requirement while 7 partially met Q2 requirement. This means that majority of the papers explained the method of analysis regarding research data infrastructure. Q3 requirement is about whether the paper showed the use of primary data that support their researches. Based on the results, only 10 papers met Q3 requirements, and 6 papers partially supported by primary data and 6 papers did not supported by primary data.

For Q4, 12 papers met the criteria, where it showed the infrastructure of research data repository in detail. 7 papers only showed it partially and 4 papers did not show any infrastructure of research data repository in detail.

In terms of quality scale, in summary, 16 papers or 69.56% scored a "Very good" quality assessment with scales between 3 to 4. 7 or 30.43% of papers scored a "Good" scale that is between 2 and 3. No papers in this study scored poor (scale 1 to 2) and very poor (scale less than 1). Table 5 and Table 6 describe the results of the overall quality assessment results. In summary, all 23 papers are within the quality assessment intended for this study.

Table. 2 Final set of articles selected for this study

ID	Type of research	Organization	Source
A1	Conceptual	University	Qin (2013)
A2	Conceptual	University	Demchenko et al. (2012)
A3	Empirical	Library	Davidson et al., (2014)
A4	Case study	University	Amorim & Castro (2015)
A5	Empirical	University	Brownlee (2009)
A6	Empirical	University	Schweik et al. (2005)
A7	Case study	University	Hruby et al. (2013)
A8	Conceptual	University	Eifert et al. (2017)
A9	Conceptual	Goverment Agency	Parida & Tripathi (2018)
A10	Conceptual	University	Lee & Stvilia (2017)
A11	Conceptual	Library	Pinfield et al. (2014)
A12	Conceptual	Research Institute	Serrano-Vicente et al. (2018)
A13	Conceptual	Library	Prabhakar & S.V. Manjula Rani (2018).
A14	Empirical	Research Institute	Abrizah, A., Noorhidawati, A., & Kiran, K. (2017)
A15	Conceptual	Research Institute	Ridwan, S.M. (2015)
A16	Conceptual	Research Institute	Uzuegbu (2012)
A17	Conceptual	Research Institute	Oguche, (2018)
A18	Empirical	Library	Baughman et al. (2018)
A19	Conceptual	University	Abdelrahman Omer Hassan (2017
A20	Conceptual	University	Kaka et al. (2018).
A21	Empirical	University	Lovett et al. (2017).
A22	Conceptual	Library	Nemati-Anaraki, L., & Tavassoli-Farahi, M. (2018).
A23	Conceptual	University	Gordon S. A et al (2015)

Table. 3 Type of research

Type of research	Conceptual	Empirical	Case Study	Total
Number of studies	15	6	2	23
Percentage (%)	65.2	26.1	8.7	100

Table. 4 Type of domain

Type of research	University	Library	Research Institute	Government Agency	Total
Number of studies	12	5	5	1	23
Percentage (%)	52.2	21.7	21.7	4.4	100

In order to determine the factors that can influence the design and the development of research data repository infrastructure, we adapted infrastructure factors from Demchenko et al. (2013), Schweik et al. (2005) and Brownlee et al. (2009) as our basis to determine the influencing factors. The factors are technology, data and metadata, policies, operation support and management, security, network infrastructure, shared scientific platform instrument, online storage and stakeholders.

Table. 5 Quality Assessment Result based on sources and questions

ID	Q1	Q2	Q3	Q4	Total
A1	1	1	1	1	4
A2	1	1	1	1	4
A3	1	0.5	0	0.5	2
A4	1	1	1	0	3
A5	1	1	1	1	4
A6	1	1	1	1	4
A7	1	0.5	0	1	2.5
A8	1	0.5	0	1	2.5
A9	1	1	0.5	1	3.5
A10	1	0.5	0.5	1	3
A11	1	1	0.5	1	3.5
A12	1	0.5	1	0.5	3
A13	1	1	0	0	2
A14	1	1	1	0.5	3.5
A15	1	1	0	0	2
A16	1	1	0.5	0	2.5
A17	1	1	1	0.5	3.5
A18	1	1	1	1	4
A19	1	1	0	0.5	2.5
A20	1	1	0.5	1	3.5
A21	1	1	1	0.5	3.5
A22	1	1	0.5	0.5	3
A23	1	1	0	1	3

Table. 6 Quality Assessment Result summary

Quality Scale	Very poor (<1)	Poor(1-<2)	Good(2-<3)	Very good (3-4)	Total
Number of studies	0	0	7	16	23
Percentage (%)	0	0	30.43	69.56	100

The papers are mapped with these influencing factorsto show which factors that each paper is focusing onto. For example, the quality assessments performed showed that the article with ID A1 put more focused on discussing factors including Technology, Data & Metadata and Policies. This shows that the paper with ID A1 have an influence in Technology factor for research data repository.

The result of this quality assessment is shown in Figure 1. Based on the analysis, the most influential factors of research data repository mentioned in the literatures are Technology with 15 papers (65.21%) discussing about it. This is followed by Policies with 13 literatures (56.52%) discussed about policies. Data and Metadata are the third most influencing factors discussed by the literatures with 11 frequencies (47.82%). Other factorsare also mentioned but with less number of frequencies that is Stakeholder (5 or 21.73%), shared scientific platform (3 or 13.04%), operation Support and management (3 or 13.04%), security (3 or 13.04%) and network infrastructure (2 or 8.69%).

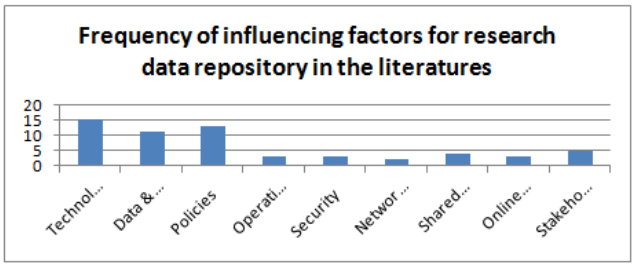

Fig. 1 Frequency of influence factors for research data repository in the literatures

V.CONCLUSION

As a conclusion, this systematic review has successfully analyzed a number of literatures that discussed specifically about infrastructure requirement for research data repository. The review also highlighted the factors of research data management infrastructure. Based on the systematic reviews performed, a total of 23 relevant papers were thoroughly reviewed and analyze. The review process has identified and categorized the influencing factors for research data repository infrastructure.

There are nine factors which have been identified: technology, data & metadata, policies, operation support & management, security, network infrastructure, shared scientific platform instrument, online storage and stakeholder.

Based on the analysis, the most influential factors in developing or designing research data repository are technology, data & metadata and policies. Surprisingly, the analysis also showed that some components that are deemed to be importantsuch as security, online storage and network are not discussed in detail in the literatures that made it to be less influential in developing or designing research data repository infrastructure.For future directions, the impact of these factors towards research data repository should be investigated. The impacts that must be considered are towards the performance of the repository and the impact towardsresearchers' satisfaction as they are the key users of research data repository.

ACKNOWLEDGMENT

This research is supported by Universiti Putra Malaysia under the Putra Grant Scheme (UPM/700-2/1/GP-IPN/2017/9558200).

REFERENCES

1. Abdelrahman O. H. 2017. The status of the University of Khartoum institutional repository. DESIDOC Journal of Library & Information Technology. 7(2), 104-108.
2. Abrizah, A., Noorhidawati, A., &Kiran, K. 2017. Global visibility of Asian universities' Open Access institutional repositories.Malaysian Journal of Library & Information Science, 15(3), 53-73.
3. Amorim R.C., Castro J.A., da Silva J.R., Ribeiro C. 2015. A Comparative Study of Platforms for Research Data Management: Interoperability, Metadata Capabilities and Integration Potential. In: Rocha A., Correia A., Costanzo S., Reis L. (eds) New Contributions in Information Systems and Technologies. Advances in Intelligent Systems and Computing, vol 353.Springer, Cham.
4. Assante, M. Candela, L. Castelli, D. And Tani, A. 2016. Are Scientific Data Repositories Coping with Research Data Publishing? Data Science Journal, 15: 6, pp. 1–24, DOI: http://dx.doi.org/10.5334/dsj-2016-006
5. Austin, C.C., Brown, S., Fong, N., Humphrey, C., Leahey, L., Webster, P. 2015. Research data repositories: review of current features, gap analysis, and recommendations for minimum requirements. Presented at the IASSIST Annual Conference. IASSIST Quarterly Preprint. International Association for Social Science, Information Services, and Technology. Minneapolis.
6. Baughman, S., Roebuck, G. and Arlitsch, K. 2018. Reporting Practices of Institutional Repositories: Analysis of Responses from Two Surveys. Journal of Library Administration. 58(1),65-80.
7. Brownlee, R. 2009. Research data and repository metadata: Policy and technical issues at the university of Sydney library. Cataloging and Classification Quarterly, 47(3–4), 370–379.
8. Davidson, J., Jones, S., & Molloy, L. 2014. Big data: the potential role of research data management and research data registries. Ifla, 1–11.
9. Demchenko, Y., Grosso, P., de Laat, C.andMembrey, P. 2013. Addressing big data issues in Scientific Data Infrastructure, 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, pp. 48-55. doi: 10.1109/CTS.2013.6567203
10. Edinburgh University. 2011. Edinburgh University Data Library Research Data Management Handbook. http://www.docs.is.ed.ac.uk/docs/data-library/EUDL_RDM_Handbook.pdf.
11. Eifert T., Schilling U., Bauer HJ.,Krämer F., Lopez A. 2017.Infrastructure for Research Data Management as a Cross-University Project. In: Yamamoto S. (eds) Human Interface and the Management of Information: Supporting Learning, Decision-Making and Collaboration. HIMI 2017. Lecture Notes in Computer Science, vol 10274. Springer, Cham.
12. Gordon, A.S., Millman, D.S., Steiger, L., Adolph, K.E. and Gilmore, R.O., 2015. Researcher-Library Collaborations: Data Repositories as a Service for Researchers. Journal of Librarianship and Scholarly Communication, 3(2), p.eP1238. DOI: http://doi.org/10.7710/2162-3309.1238
13. Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. Big Data & Society. https://doi.org/10.1177/2053951718786316
14. Hruby, G. W., McKiernan, J., Bakken, S., &Weng, C. 2013. A centralized research data repository enhances retrospective outcomes research capacity: A case report. Journal of the American Medical Informatics Association, 20(3), 563–567.
15. Kakai, M., Musoke, M. G., &Okello-Obura, C. 2018. Open access institutional repositories in universities in East Africa. Information and Learning Science, 119(11), 667-681.
16. Lee, D. J., &Stvilia, B. 2017. Practices of research data curation in institutional repositories: A qualitative view from repository staff. PLoS ONE (Vol. 12)
17. Louise Corti, Veerle Van den Eynden, Libby Bishop, Matthew Woollard,Managing and Sharing Research Data: A Guide to Good Practice, Sage Publications,2014, http://www.sagepub.com/sites/default/files/upm-binaries/61019_Corti__Managing_and_sharing_research_data.pdf
18. Lovett, J. A., Rathemacher, A. J., Boukari, D., & Lang, C. 2017. Institutional Repositories and Academic Social Networks: Competition or Complement? A Study of Open Access Policy Compliance vs. ResearchGateParticipation.Journal of Librarianship and Scholarly Communication.
19. Nemati-Anaraki, L., &Tavassoli-Farahi, M. 2018. Scholarly communication through institutional repositories: proposing a practical model. Collection and Curation, 37(1), 9-17.
20. Oguche, D. (2018). The state of institutional repositories and scholarly communication in Nigeria.Global Knowledge, Memory and Communication, 67(1/2), 19-33.
21. Okoli, C., Schabram, K. 2010. A Guide to Conducting a Systematic Literature Review of Information Systems Research. Sprouts: Working Papers on Information Systems, 10(26).
22. Parida, P. K., &Tripathi, S. 2018. Odisha Spatial Data Infrastructure (OSDI) – Its Data Model , Meta Data and Sharing Policy ., IV(November), 20–23.
23. Pampel H, Vierkant P, Scholze F, Bertelmann R, Kindling M, et al. (2013) Making Research Data Repositories Visible: The re3data.org Registry. PLOS ONE 8(11): e78080. https://doi.org/10.1371/journal.pone.0078080
24. Pinfield S, Cox AM, Smith J (2014). Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. PLoS ONE 9(12): e114734.
25. Prabhakar& S.V. Manjula Rani 2018. Benefits And Perspectives Of Institutional Repositories in Academic Libraries. Scholarly Research Journal for Humanity Science & English Language.5(25).
26. Qin, J. 2013. Infrastructure, Standards, and Policies for Research Data Management. In: Sharing of Scientific and Technical Resources in the Era of Big Data: The Proceedings of COINFO 2013, pp. 214-219. Beijing: Science Press.
27. Ridwan, S.M. 2015. Institutional Repository: A Road Map to Open Access and Resources Sharing in Nigeria (Issues and Challenges). International Journal of Scientific & Engineering Research.6(1),598-605.
28. Schweik, C. M., Stepanov, A., & Grove, J. M. 2005. The open research system : a web-based metadata and data repository for collaborative research, 47, 221–242.
29. Serrano-Vicente, R., Melero, R., &Abadal, E. 2018. Evaluation of Spanish institutional repositories based on criteria related to technology, procedures, content, marketing and personnel. Data Technologies and Applications, 52(3), 384-404.
30. Uzuegbu, C. P. 2012. Academic and research institutions repository: a catalyst for access to development information in Africa. In 78th World Library and Information Congress: International Federation of Library Assciations and Institutions, Helsinki. (August), 1-18.

31. Wissik, T. &Durco, M. 2015.Research Data Workflows: From Research Data Lifecycle Models to
32. Institutional Solutions.CLARIN 2015 Selected Papers • Linköping Electronic Conference Proceedings, No. 123, pp. 94-107.