The Social Sciences 12 (8): 1310-1316, 2017
ISSN: 1818-5800

Developing a Coordinated and Collaborative Research Data Management Framework using e-Infrastructural Approach

[1]Zanariah Jano and [2]Nurul Aqilah Ahmadi
[1]Centre for Languages and Human Development, Universiti Teknikal Malaysia Melaka,
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
[2]Institute of Technology Management and Entrepreneurship,
Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

Abstract: Hitherto, various database systems exist in archiving research data at institutions of higher learning in Malaysia. The library, Centre of Research Management, Centres of Excellence, Publisher and e-Repository have separate databases to archive and manage the data. This poses a challenge as no one department is in a position to 'own' the resulting set of services. Data are complex and heterogeneous and little agreement is reached on standards to describe data across disciplines. Hence, a centralized research data management is vital to act as a single source of 'signposting' information which offers a valuable resource for researchers from multi-disciplinary fields for the different stages of research cycle. This study investigated the ways in which universities in the advanced countries implemented data management infrastructure through a qualitative content analysis of websites. The data are useful toward devising a conceptual framework of research data management for the Malaysian Higher Institutions. Furthermore, the proposed framework with will be useful for Research Management Centres in universities to provide a centralized platform for researchers in guiding them through the research data life cycle and offer fresh insights.

Key words: Research data management, e-Infrastructural, framework, Malaysian higher institutions, qualitative

INTRODUCTION

In the case of Malaysia, research publications have been produced due to the efforts of government in triggering research efforts. The Ministry of Higher Education Malaysia (MOHE) allocates grants totaling RM741 million under the Tenth Malaysia Plan (RMK-10) comprising of four main grant schemes to intensify the research efforts of the higher learning institutions. The four schemes are Fundamental Research Grant Scheme (FRGS), Exploratory Research Grant Scheme (ERGS), Long-Term Research Grant Scheme (LRGS) and Prototype Research Grant Scheme (PRGS). For 2012, the four main grant schemes worth a total of RM170.47 million. Other efforts include the establishment of potential Centers of Excellence (CoEs) in Higher Education Institutions (HEIs) in order to further leap them towards research, innovation and internationalization.

This move is to support the third thrust of National Higher Education Strategic Plan (NHESP), towards gearing up the level of the country's research and development (R&D) and innovation. The vision is to enable the HICoEs to be educated-wide knowledge enterprises that will initiate and lead intra and inter institutional research in collaboration with international and industrial partners in line with the global best practices, connecting innovation systems by fostering discovery, creativity and new talent. Furthermore, the mission is to provide a conducive environment that nurtures scientific exploratory works and creativity of knowledge discovery among scientists, policy makers and technopreneur for wealth creation. Thus, HICoEs will be supported and facilitated by MOHE so that they will become the "focused vehicles" that will drive R&D and innovation agenda, particularly in fundamental research as well as contributing to human capital development. These agendas, hence, trigger the needs to manage the extensive knowledge by means of academic publications. However, currently, Malaysia and other countries in Asia are lacking devices to access research data and reusable research for future reference (Johare *et al*., 2009).

Literature review

Research data management: Data management represents different things for different people. Data repositories perceive it in terms of preservation and curation whereas

Corresponding Author: Zanariah Jano, Centre for Languages and Human Development, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100, Durian Tunggal, Melaka, Malaysia

researchers relate it with structuring of data in a database, or the organizing of files and folder. As a whole research data management includes the processes undertaken from the initial phases to the research output. Research inputs include journal articles, research notes whether hand written or electronic through archaic manuscripts and to advanced images. This study is utilizing this definition of research data management in the study as central coordination is vital in developing institutional strategies.

Infrastructural approach: Infrastructural approach recognizes that the qualities that make a sound data management are the sets of generally assumed technologies and knowledge that enable individuals to manage data efficiently. Every aspect is vital in which the infrastructures underpin the distinctly "coordinated" and "collaborative" existence. The particular aspects of infrastructure include university policy, systems for the preservation and documentation of research data, training and support, software tools for the visualisation of large images and creating and sharing databases utilizing the web (database as a service) (Wilson *et al*., 2011). According to Australian National Data Service (ANDS) there are four elements to create research data management such as "university policies", "information technology infrastructure", "support services" and "managing data". The study has resulted in the production of a website under the purview of research service unit at the university, detailing all the pertinent features required by the researchers.

Beagrie *et al*. (2010) conduct a study on mechanism on keeping research data safe and ascertain that the constant turnover of post-doctoral researchers always results in lost data. Further, the current mechanisms lack the routine practices in collecting and organizing the data that post-doctoral researchers generate as there is insufficient information on how the data was created. Hence, it is a high time to generate and discover how to sustain well-curated data.

Previous works: Few projects have been launched at universities with regards to research data management. Ball (2009) has written a comprehensive report detailing the research data management in the United Kingdom, focusing on projects like DataShare, SPECTRa, Dryad Project, DCMI science and metadata community and Go-Geo Go-Geo.

DataShare Project is a 2 years, Joint Information System Committee (JISC) funded project analyzing steps in integrating data of research into university repositories involving the University of Edinburgh, Oxford and Southampton. The main aim of the project is to develop paradigms for handling of research data through e-Prints, DSpace and Fedora. University of Edinburgh has designed a profile of Dublin Core Metadata Terms and DSpace metadata with the specific needs of multi disciplinary research data.

Another project is entitled Data Audit Framework (DAF) conducted by HATII at the University of Glasgow, in partnership with the University of Edinburgh, UKOLN at the University of Bath, Kings College London, Imperial College London and University College London. It is a methodology and complementary tools to identify and assess the data, policies and procedures for managing it. It is the collection of management metadata covering the data assets identified by the audit (Jones *et al*., 2008). The metadata set focuses on data management.

SPECTRa refers to Submission, Preservation and Exposure of Chemistry Teaching and Research Data Project which was conducted between 2005 and 2007, a joint venture between the University of Cambridge and Imperial College London. They have developed open Source tools to cater for high volumes of chemical data and are to be placed into a DSpace repository and subsequently reused. The project prioritizes that these tools should integrate with established workflows of the chemists generating the data. An extended version of the eBank profile was utilized. The principal differences of the SPECTRA profile are as:

- Distinguish the chemist owning the data ranging from the spectroscoper or crystallographer, subject to who the owner
- An adaptation to suit organic and computational chemistry
- Include information with regards to open access embargoes

Another project is known as the Dryad Project which is collaboration between the Metadata Research Center of the University of North Carolina and the (United States) National Evolutionary Synthes is Center (NESCent), funded by the National Science Foundation (NSF). It aims to design published evolutionary biology datasets repository and to link those datasets with major journals and databases in the relevant fields. DCMI Science and Metadata Community is another project as a follow up discussions from the 2008 International Conference on Dublin Core and Metadata Applications. Science and Metadata Community forum was set up by the Dublin Core Metadata Initiative for individuals and organizations to exchange information and knowledge about metadata describing scientific data. Go-Geo Go-Geo is another project to provide a geospatial data collections portal. It

is collaboration between EDINA National Data Centre, University of Edinburgh and the United Kingdom Data Archive, University of Essex. The Go-Geo! Portal was initially run as a trial service to the United Kingdom Higher and Further Education communities by EDINA, and by November 2008 became a full Joint Information System Committee (JISC) service. The portal operates through the Z39.50 protocol, querying a number of different data repositories (ANSI-NISO, 1995) and collating the results. This is subject to the supplies of consistent metadata by each of the Z39.50 targets and is utilizing the Academic Geospatial Metadata Application Profile ISO 19115 profile and UK GEMINI superset, specially produced by the Go-Geo project as standards for the United Kingdom Higher and Further Education Communities.

The report outlines several recommendations for stakeholders to ensure collaborative efforts in sustaining research data management (Ball, 2009). Any scientific data application profile must be fully supported by key stakeholders. The key stakeholders include:

- Repository software developers; to adjust and modify their software to suit the profile
- Institutional repository managers; to manage the Profile for the repository in terms of changing the workflows of the repository to suit and interacting with data creators to ensure in-depth comprehension and knowledge of the metadata requirements
- Cross-search tools (e.g., repository search)
- Data centre's; to adjust their systems to support the output of metadata in a form compliant with the profile
- Researchers producing scientific data; to supply the metadata required by the profile
- Researchers seeking scientific data; to discover and re-use existing data

All in all, all stakeholders play essential roles to sustain research data management in institutions. And they are responsible to make a research data management is more efficient and systematic.

MATERIALS AND METHODS

This project proceeded through three main phases. The first phase was compilation of research data management elements, the second phase was an identification of standard of research data management and the third phase was a development of framework research data management. A qualitative content analysis of website was performed on the selected top five research universities listed in Times Higher Education World University Rankings 2014-2015 in which the universities implement research data management. Content analysis was used, focusing on the content and internal features text (Naidu, 2011). According to Hsieh and Shannon (2005) qualitative content analysis is the process of coding to identify some of themes or patents in the context of text data.

Samples and data collection: The samples consisted of the Research Data Management materials collected through qualitative content analysis of top five universities listed in the Time High Education World University Rankings 2014-2015. This project proceeded through the following three main phases: compilation of research data management elements, identification of standard research data management and development of framework.

Phase 1: Compilation of research data management elements the relevant websites of top 5 universities in the Time World University Ranking 2014-2015 were gathered. A flexible coding scheme was utilized on the basis of prompt discovery'.

Phase 2: Identification of standard research data management the coding schemes were used to analyze qualitatively the patterns of standardization which existed among the samples.

Phase 3: Development of framework-Based on the standard features employed among universities, a framework was devised.

RESULTS AND DISCUSSION

Elements of research data management were based on qualitative content analysis of websites. The top five of university websites were accessed and coded. The content of standard research data management elements was divided into six; data management plan, data organization, storage and backup, data sharing, training and consulting and research data management policy. However, according to Parsons (2013) nine elements are needed to develop research data management website which are "what is research data?"; "research data lifecycle"; "data management planning"; "creating data"; "organizing and storing data"; "data sharing and data archiving"; "research data showcase"; "training, advice and support" and "contact us". Figure 1 shows the standard elements of research data management which have been used in top 5 university websites listed in the

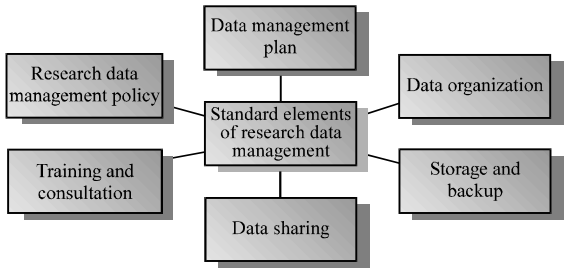

Fig. 1: Standard elements of research data management

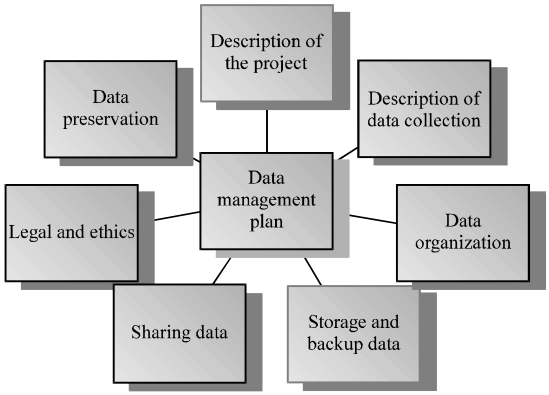

Fig. 2: Data management plan

time high education world university rankings 2014-2015; data management plan, data organization, storage and backup, data sharing, training and consultation and research data management policy. However, the universities use various terms in referring to the same items.

Data management plan: Data management plan is used by all five universities. The standard sub-elements used for data management plan include description of the project, description of data collection (methodology), data organization, storage and backup data, sharing data, legal and ethics and data preservation. In United States, the National Science Foundation (NSF) required a research data management plan as a part of the funding proposal process (NSF, 2010). The requirement of data management plan is not only during proposal but everyday operation in which researchers need to update data management plan (Beagrie *et al*., 2010). Figure 2 shows standard elements for data management plan.

Several research funders in the United Kingdom require data management plan as a crucial component of research grant application (DCC, 2015a, b). In fact University of California Libraries have developed a system, service and tools to facilitate user friendly

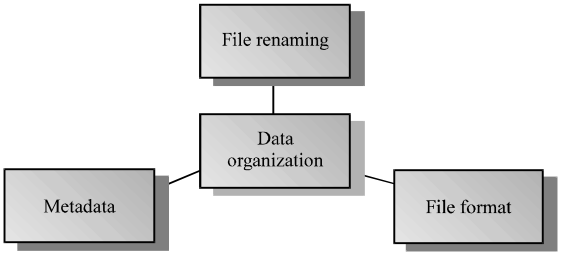

Fig. 3: Data organization

research data management such as "DMP Tool", "EZID", "Web Archiving Service" and "data management programs" to meet requirement of the National Science Foundation and other funding agencies (Starr *et al*., 2012).

Each university has provided data management plan template to researchers. Currently, Digital Curation Center has also offered data management checklist which includes "administrative data", "data collection", "documentation and metadata", "ethics and legal compliance", "storage and backup", "selection and preservation", "data sharing" and "responsibilities and resources" and also including "guidance and questions are need to be consider" when preparing a data management plan (DCC, 2013). Hence, the standard elements are in rhyme with the ones offered by the Digital Curation Centre. Universities can use this guideline to develop data management plan template.

Data organization: The second element is data organization which includes sub-elements like file renaming, file format and metadata. Figure 3 shows standard sub-elements of data organization.

Data organization is very important because researchers are always losing a research data and procedure or data collection are not systematic. Researchers need to organize and manage a research data throughout the project from the beginning of research project until the end of research project. Data organization must be always organize and systematic for future reference. There have a lot benefit to researchers if research data are organized and systematic.

Data organization is a part of research data management. Data organization includes file renaming, file format and metadata. Several universities have provided guidelines on how to organize a data management. The top five university websites guide researchers on how they can use file naming, file format and how to create metadata or data documentation. For example, for ease in managing file format, they are proposing researchers to use "PDF", "MPEG", "XML" and others. University can

provided a software tool for researcher to file renaming ranging such as "Bulk Rename Utility", "Renamer" and "PSRenamer" (CDL, 2015). Other guides for file naming are also included.

Metadata or also known as data documentation is summaries of information about research data. Metadata is information of research data for researchers in order to provide structure of the research data and to help a research data is more reusable in the future and help the other researchers to cite a research. Metadata must be start in the beginning of research project until the end of research project. Standard sub-elements of metadata are title of the research project, researcher name, date, description of research project, identifier, language, publisher, file format, file naming, sources, methodology, rights and others. In short, most universities have provided metadata tools and guidance to researchers on creating metadata. CDL (2015) also explains how to create metadata in their "data management general guidance" as a guideline to university or institutions.

Storage and backup: The third element is storage and backup which offers a storage solution for researcher and university or department and how universities backup their research data or material during a research project. Universities must also consider the cost of storage and backup of research data. Pertinent questions like how research data can be stored and backup during the research project and who are responsible must be considered (DCC, 2013). However, a few universities have given a guideline to researchers on how they can storages and backup their research data.

Data sharing: The fourth elements are data sharing which includes sharing with other researchers, digital repositories, intellectual property and citing data. Figure 4 shows standard sub-elements of data sharing. Research data must be useful and provide knowledge and can be access for sharing with other research communities from the first step of research project until the end. Researchers can share research data in many ways such as using email, sharing conference papers and others but few considerations when sharing data are essential. Data sharing is important to researchers because of the efforts of researchers to generate the data are useful to others and can be reused in the future (Wallis *et al*., 2013). There are many ways to share a research data.

CDL (2015) two types of repositories are utilized which are discipline specific and institutional. Before researchers can share and publish their research data,

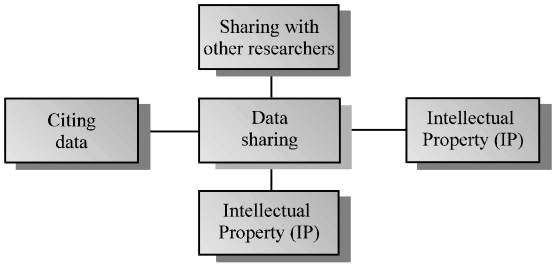

Fig. 4: Data sharing

intellectual property or any ethical issues must be considered before data can be preserved and sharing (DCC, 2013). CDL (2015) defines five core elements when make citation such as "creator", "title", "publication year", "publisher" and "identifier".

Training and consultation: The fifth element is training and consultation. Universities can provide any training or workshop for researchers to ensure that researcher can abide to all rules in data management plan and meet a research funder requirement. Several universities have provided trainers and service centers to help researchers. Many bodies in the United Kingdom and United States have offered their services of training, support or tool to help researchers and also universities with research data management (Qin and Solinger, 2011). The top five universities have provided consultation and workshop sessions on data management plan for their researchers.

Research data management policy: The sixth element is research data management policy. The policy is integral to guide universities. However, most websites did not reveal their policies virtually. Some of university presented research data management policy in their websites and this elements includes purpose or objective of research data management, guidance (training and support), collaboration of research project, meet funding requirement (data management plan), data sharing, definition of terms, retention of research data, research data ownership, removal of research data, responsibilities such as universities responsibilities, researchers responsibility, updating policy and others. Some universities especially in United Kingdom have developed research data management policies (DCC, 2015a, b). Hence, in comparison, elements in research data management policies consist of "research data management", "research data ownership and storage" "access to database and archives", "retention of materials and research data following a research project", "collaborative research project with others research

organizations", "ethics, codes, funding requirement and relevant legislative framework", "data sharing and re-use", "secure and safe disposal of research data", "intellectual property; copyright and patents" and "record management" (Australian National Data Service).

Implementation of research data management policies is very challenging in order to deal with any issues in science data management (Jian and Solinger). Australian National Data Service also states that university should constantly update policy in line with current issues in research data management. The creation of research data management policies is the first step in the process of developing a research data management (Jones *et al*., 2013). In a nutshell, the integral elements that comprise research data management framework in the present study are data management plan, data organization, storage and backup, data sharing, training and consultation and research data management policy. However, developing research data in universities is dependent on the needs and requirements of the university where it involves the collaboration of all departments such as researcher, research centers, faculty, library and others.

CONCLUSION

The indication presented in the qualitative content analysis of the top five research universities listed in Times Higher Education World University Rankings 2014-2015 shows interesting findings. In conclusion, the study develops a research data management framework based on standard best practices of research universities which consists of data management plan, data organization, storage and backup, data sharing, training and consultation and research data management policy. The study is limited in terms of samples, scope of study and genre used which prevent the generalization of the findings. More universities from other countries should be included to obtain more understanding and gain information on how they develop research data management. This study does not only fill the gap in the literature but also provide insights on best practices of research data management systems utilized by universities in advanced countries. The study offers a proposed model which will provide understanding of infrastructural approach that induces coordination and collaboration among faculties and departments in Higher Learning Institutions. Instead of working in silo, the study penetrates the pool of talents in each department to contribute to the system so that equilibrium in the eco-system is sustained. Most importantly, the framework will be the base for the Research Management Centres in Malaysia and other countries to utilize.

REFERENCES

ANSI-NISO., 1995. Information retrieval: Application service definition and protocol specification. American National Standards Institute, Washington, USA.

Ball, A., 2009. Scientific data application profile scoping study report. MSc Thesis, University of Bath, Bath, England.

Beagrie, N., B. Lavoie and M. Woollard, 2010. Keeping research data safe 2: Final report. Master Thesis, Faculty of Social Sciences, Universities of Cambridge, Oxford, UK.

CDL., 2015. Data management general guidance. California Digital Library, Oakland, California. https://dmptool.org/dm_guidance

DCC., 2013. Checklist for a data management plan, v4.0. Digital Curation Center, Edinburgh, Scotland. http://www.dcc.ac.uk/resources/data-management-plans

DCC., 2015a. Funders data plan requirements. Digital Curation Center, UK. http: //www .dcc .ac. uk/ resources/data-management-plans/funders-requirements

DCC., 2015b. United Kingdom instituional data policies. Digital Curation Center, UK. http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies

Hsieh, H.F. and S.E. Shannon, 2005. Three approaches to qualitative content analysis. Qual. Health Res., 15: 1277-1288.

Johare, R., A.M. Yunus, I.K.A. Kadir and H. Mohamed, 2009. Managing primary research data and records for research in research institution and related organizations: Examples from the TEAM Malaysia case studies. Proceedings of the Interpares 3rd International Symposium, April 6, 2009, Universiti Teknologi MARA, Shah Alam, Malaysia, pp: 1-9.

Jones, S., G. Pryor and A. Whyte, 2013. How to develop research data management services-a guide for HEIs how to develop research data management. Digital Curation Centre, UK.

Jones, S., S. Ross and R. Ruusalepp, 2008. Data Audit Framework Methodology. University of Glasgow, Glasgow, Scotland,.

NSF., 2010. NSF data management plan requirement. National Science Foundation, Virginia, USA. https://www.nsf.gov/eng/general/dmp.jsp

Naidu, S., 2011. Qualitative Research: Data Collection and Data Analysis Techniques. Universiti Utara Malaysia Press, Sintok, Malaysia,.

Parsons, T., 2013. Creating a research data management service. Int. J. Digital Curation, 8: 146-156.

Qin, J. and C. Solinger, 2011. Institutional policies on science research data: A pilot analysis. Proceedings of the 2011 Conference on iConference, February 8-11, 2011, ACM, Seattle, Washington, USA., ISBN:978-1-4503-0121-3, pp: 761-762.

Starr, J., P. Willett, L. Federer, C. Horning and M.L. Bergstrom, 2012. A collaborative framework for data management services: The experience of the University of California. J. E Sci. Librarianship, 1: 109-114.

Wallis, J.C., E. Rolando and C.L. Borgman, 2013. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PloS One, Vol. 8, 10.1371/journal.pone. 0067332.

Wilson, J.A., M.L. Uribe, M.A. Fraser and P. Jeffreys, 2011. An institutional approach to developing research data management infrastructure. Int. J. Digital Curation, 6: 274-287.