

DESIGN AND DEVELOPMENT OF A NATIONAL LEVEL RESEARCH DATA REPOSITORY IN INDIA

Jamsheer N.P.

Under the guidance of

Dr. Mohamed Haneefa K.

Assistant Professor
Department of Library and Information Science
University of Calicut.

1. Introduction

Scientific research plays a vital role in the development. The speed and depth of this research depends on nurturing collaborative research and exchange of ideas and resources. Organisation for Economic Cooperation and Development (OECD) and guideline on access to research, OECD (2007) recommends the exchange of ideas, knowledge and data as fundamental for human progress. Human genome project is a good example of collaborative biological research based on data compiled by scientists from various parts of the globe.

During the research, researcher generates data and in many occasions only a portion of it is used for the purpose of analysis and findings. Once the research is over, entire dataset is either discarded or stored in personal or institutional hard drives. Sharing of data reinforces open scientific inquiry; encourages diversity of studies and opinion; promotes new areas of work and enables the exploration of topics not covered by the initial investigators. Besides, open access to research data increases the returns from public investment on research and development.

Government of India has set up platform called open government data platform India, for proactive and open access to the data generated by various Government of India entities. The objective of the open government data is providing access to Government of India owned shareable data in machine readable form through a web portal, permitting sustainable and wider accessibility of valuable government data by public.

2. Need and Significance of the Study

Information and communication technologies (ICTs), now in widespread use throughout all research disciplines, have greatly aided system of free exchange and have opened up new avenues for collaboration and sharing. The area of research data sharing is very complex because of the heterogeneous nature of academic subjects. Thus it is difficult for researchers, funding bodies, publishers and scholarly institutions to select appropriate repositories for storage and search of research data.

More than 1000 research data repositories from different academic disciplines are registered in global registry of research data repositories (www.re3data.org). An effort to set up a national level research data repository is the need of the scientific community which facilitates preservation and provides open access of data sets to researchers as well as to the wider community of users. Requirements of Indian scientific community should be considered and research data repository be initiated in order to foster the speed and depth of research. A national level research data repository would facilitate preservation and storage of research data generated within India.

3. Statement of the Problem

Research data are valuable and reusable. Providing sustainable access to research data is a challenge of the academic community. The long-term preservation together with the principle of open access to research provides unlimited possibilities in the area of research.

Thesis, dissertations, articles and other intellectual outputs are already available through open access repositories and this has reduced to a greater extent the barrier between the academic professionals and wider community. As it becomes clear that it is practicable to achieve the goal of Open Access, more governments are asking that all publicly funded research is made available to all for free, Galsworthy(2013).

As the volume of new research data expands, however, so does the need for the archiving of such data as well as the implementation of efficient procedures that enable researchers to locate, search, retrieve and reanalyse these data.

Public access to research data helps reap the fruits of publicly funded research in optimum level. In countries like India, where less than 1 percentage of GDP is being spent for research and development, it is high time to share data of publicly funded research for further development. Useful research data are stored in the personal hard drives of scholars, which could have been used for reanalysis of the existing problems.

Development of repository of research data at national level would unleash the possibility of preservation and reuse which could bring more value to the money spent from public exchequer.

The title of the study is "Design and Development of a National Level Research Data Repository in India."

4. Definition of the Key Terms

4.1 Design

Definition given by Oxford Universal Dictionary is used in the study.

"the general form or arrangement of something". As far as this study is concerned, design is the process which involves the creation of the architecture and functional workflow of the proposed national level research data repository.

4.2 Research Data

The term research data is used differently in various contexts, Blue Ribbon Task Force on Sustainable Digital Preservation and Access has defined research data as "the primary inputs into research, as well as the first-order results of that research" BRTC (2010). Office management and Budget department of US government OMB (1999) defines research data as "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings"

However, for the purpose of this study definition adopted by OECD (2007) is used "research data are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated."

4.3 Research Data Repository

According to Crow (2002) institutional repositories are digital collections capturing and preserving the intellectual output of a single or multi-university community. For the purpose of the present study research data repository can be defined as the digital collection of research data.

5. Literature Review

Many datasets, collected in the beginning could not be used for the lack of proper documentation and data curation and hence no reuse was possible. This fact is

reflected in the survey findings of Kuula and Borg (2008). The responsibility of development and maintenance of national level repository of research data is a well debated issue. Pullinger and Wagner (2010) have the opinion that it would be useful to arrive at a decision on the respective roles of national libraries, national archives and research data centres. Because of complexity in formats and metadata standards used and the growing nature of research, research data centres be assigned with the prime responsibility of managing repository.

A researcher would be more responsible for the data he/she shares, which would increase the quality of research argues Galsworthy (2013).

OECD has published detailed recommendations on the need and significance of open access to research data, OECD (2007), problems challenges and solutions for its member countries. According to OECD "Effective access to research data, in a responsible and efficient manner, is required to take full advantage of the new opportunities and benefits offered by ICTs."

6. Objectives of the Study

The main aim of the study is to design and develop a national level research data repository in India. In order to fulfill this aim, the following specific objectives are identified:

1. To investigate the need and significance of open access to research data.
2. To study the present practice of storage, preservation and dissemination of research data.
3. To understand the attitudes, concerns, and interests of the research scholars/scientists in contributing data to proposed research data repository.
4. To identify tools, techniques and technical problems related with the design of research data repository.
5. To design and develop a sustainable national open access research data repository.

7. Methodology

The study depends heavily on primary research literature, experience and expertise, tools and techniques. Interactions with system administrators of repositories and digital libraries to understand challenges involved in establishing national level systems, problems encountered, trouble shootings and maintenance of the system. Interview with software experts, information professionals and research supervisors

would be used to identify problems and challenges. Possible area of problems would also be identified through responses of research data generators using questionnaire method. Pilot study would be conducted by collecting sample research data from various disciplines at university level. Problems of data collection, data curation, and metadata standards will be studied by installing software and setting up a prototype. Technical feasibility of incorporation of research data sets with the present digital library/institutional repository software would also be explored.

8. Scope and Limitations of the Study

The study proposes to identify the need and significance of preserving and sharing data generated as a part of research at universities and other research institutes in India. Role of various stakeholders in the development of research data repositories and skills required for maintenance would be studied. Study takes sample from prominent research institutes and universities in India.

The ethical and legal issues related with preservation and sharing of research data are not covered under this study. The study will use a quota sampling of research scholars and scientists. Data will be collected from the prominent research institutions and universities in India.

9. Conclusion

This study will investigate the present practice of storage, preservation and dissemination of research data. It will assess the attitudes and concerns of research scholars and scientists towards open access of research data. This study will design and develop a sustainable open access research data repository in India. This is very essential for the sustainable research and development in India.

References

- B. R. Task Force. Sustainable Digital Preservation and Access.
- Crow, R. (2002). The case for institutional repositories: a SPARC position paper. ARL Bimonthly Report 223.
- Galsworthy, M. (2013, June 25). Funding bodies will have to force scientists to share data. *The Hindu*. Retrieved from <http://www.thehindu.com/sci-tech/funding-bodies-will-have-to-force-scientists-to-share-data/article4849229.ece>
- Kuula, A., & Borg, S. (2008). Open access to and reuse of research data- The state of the art in Finland. *Finnish Social Science Data Archive*, 31.

Office of Management and Budget, Circular A-110 (revised 11/19/93, as further amended 9/30/99), available at www.whitehouse.gov/omb/circulars/a110/a110.html (last visited December 2, 2014).

Pilat, D., & Fukasaku, Y. (2007). OECD principles and guidelines for access to research data from public funding. *Data Science Journal*, 6, OD4- OD11.

Principles, O. E. C. D. (2007). Guidelines for Access to Research Data from Public Funding. URL: <http://www.oecd.org/science/scienceandtechnologypolicy/oecdprinciplesandguidelinesforaccessto-researchdatafrompublicfunding.Htm>.

Pullinger, J., & Wagner, G. G. (2010). *On the respective roles of national libraries, national archives and research data centers in the preservation of and access to research data* (No. 153). German Council for Social and Economic Data (RatSWD).

