ISSN (Print): 0972-2467
ISSN (Online): 0976-2477
SRELS Journal of Information Management, Vol 56(2), April 2019, p. 112-118 DOI: 10.17821/srels/2019/v56i2/131578

Emergence of Research Data Literacy with Special Reference to India

Prashant Shrivastava* and Dinesh K. Gupta

Department of Library and Information Science, Vardhman Mahaveer Open University, Kota – 324021, Rajasthan, India; pshrivastavalib@vmou.ac.in, dineshkumargupta@vmou.ac.in

Abstract

Preservation of research data is an important policy requirement for universities and research organizations so as to manage research data at organizational level and make it available lifelong. However, the importance research data management especially in the context of e-research and data intensive research is not widely recognized both by researchers and research organizations due to lack of research data literacy. In the Indian context, it is essential to formulate a national data sharing and accessibility policy so that Indian universities and research organizations can cope research data.

Keywords: Research Data, Research Data Literacy, Research Data Management, Research Data Literacy-India

1. Introduction

Universities and research organisations have always been focussed on research and, in recent years, the need to make research data freely accessible has been emphasized; many institutional repositories and subject/national repositories have been established. In the course of any research a large amount of data gets generated. Not all of this data will find a place in the final theses / dissertation or research report which necessitates managing this data. Though research data literacy concept is about a decade old it has not been explored so far in this country and this study is an attempt in this direction.

1.1 Objectives

The core objectives of this study are to:

a. Examine the relevance for research data globally.
b. Examine the status of research data management in prominent universities and to discuss relevance and barriers of research data literacy.
c. Analyse research data literacy preparedness in India

1.2 Scope

Research data management is being considered significant for better research outcomes at the global level. The scope of this study is to examine the mandate for research data management in prominent universities and organizations.

1.3 Methodology

This paper is divided into two parts; part one covers global developments about research data management and research data literacy. This part is based on secondary sources of information. Various research papers, reports, repositories, policy documents have been used. For part two, which deals with India's preparedness, secondary sources along with primary information received about the CSIR (re3data.org) is used along with the observations about the developments taking place in the country.

2. Towards Research Data Literacy: Developments at the Global Level

21st century is the age of data which is generated, utilized, converted, analysed, transformed and applied in day to day life. during the course of research, data gets generated, which when applied generates more data. Problem specific data have been utilized in various decision and policy framing activities. Data is available everywhere and is

*Author for correspondence

openly accessible worldwide, while its interpretation and orientation decide its value (Gurria, 2007). It is essential to develop research data literacy to achieve global scientific standards. Research data and data literacy are both basic pillars of research data literacy. In the words of John Wilbanks, "Modern research may be conducted on the shoulders of research data dealing parameters those define capacity of archiving of data, analysis and management of data (Shearer, *et al.*, 2010). Present data intensive research demands extensive support of research data management (Hey, *et al.*, 2009).

OECD (Organization for Economic Co-operation and Development) - a special group of thirty countries to address global challenges - has realized that global development will happen through innovative scientific growth and has recommended guidelines for research data. It will only be possible through global exchange of research data based on agreed principles. The aim of OECD is to assemble a research data efficient global science infrastructure (Gurria, 2007). The outline report of UK Data service in 2014, recommended various measures including enabling researcher to understand their responsibility about research data. Secondly it recommends transparency of generated research data as required by governments and research funding agencies. Finally, efficient management of research data is also a requirement for economic development (Mandinach, & Gummer, 2013).

In 2003 the Committee of Support of Research at U.K prepared a report on policies and futuristic demand for curation of primary research data. The recommendations related to digital research data preservation and suggested high end management of research data as having the potential to contribute to save funds to avoid infrastructure expenditure (Woollard, & Corti, 2014).

The data working group of Cornell University submitted a digital research data curation report in 2008 suggesting the need to develop an environment that supports efficient curation, transformation, analysis and sharing of research data (Albert, *et al.*, 2008).

Understanding of research data demands data literacy curriculum starting at higher secondary education level. Data literacy abilities of researchers should make them capable of dealing with the process of transformation from raw data to linked information. The set of skills for data literacy comprises proper understanding of data with required archiving techniques and organizational knowledge (Vahey, *et al.*, 2006).

3. Conceptualising Research Data Literacy

According to Engineering and Physical Sciences Research Council "Research data is recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created" (EPSRC expectations on Research Data Management, n.d.).

Elsevier support services indicate that "the precise notion of what constitutes research data will differ from field to field". It may be any type of data which is directly linked to the output and results. Research data may be observational, statistical, data without analysis, the records and output of a scientific device, data from surveys, digital scans / readings (Research data policy at Elsevier support services, n.d.).

There have been many other institutions and organizations that have defined research data (Research data policy, n.d.; Defining Research data, n.d.; Data guide, n.d.; University of Edinburgh data policy, n.d.; Steen, 1999.

4. Research Data Management Policies

Canadian Association of Research Libraries Data Management Sub-committee prepared a toolkit for data management in 2008 The report admitted the role of efficient management of research data to achieve swift scientific progress. It was observed in the report that shared research data can increase the visibility of research. The report recommended management of research data in projects funded by funding agencies (CARL Data Management Sub-Committee, 2009).

Monash University research data management policy explained that research data for various disciplines may be different in nature. Social science researchers may generate research data in the form of statistics, results and interviews while humanities research work may be

in texts. Sometimes research involves using existing data, collecting or creating new data itself" (Managing research data, n.d.).

5. Barriers to Research Data Literacy

Data literacy is a combination of information, digital, visual and media literacy. The present working generation in research and education is not fully digitally literate but due to versatile experience they have authority to command and frame the future development. It has been identified many times on failure of policies (Ridsdale, *et al.*, 2015).

It has been observed in various fields that data skill of employees always enhances their performance and output, but daOpen data initiatives are getting attention around the world. Research data literacy may be developed by connecting open data repositories with capabilities of translation, format interchange and analytical tools, also back end support of an efficient federated search engine.

6. Echo of Research Data Management

Research data management comprises organization of research data during complete life cycle of data. Many universities and research organizations. The League of European Research Universities (LERU) Roadmap for Research Data has acknowledged the new era of data driven science in universities (Clements, 2013). Research data policies/practices at different universities around the world are summarized.

The University of Iowa identified pitfalls in research data management by interviewing individual researchers and drafted a data management report (Educause, 2006). Monash University, Australia published a *Research Data Management Strategy and Strategic Plan 2012–2015*, (Johnsson, & Ahlfeldt, 2015). Loughborough University started a new platform to manage research data. It had been launched to maintain long-term archiving and provides unique search facilities (Johnston, 2014). University of St Andrews initiated research information management infrastructure to support research data management (Lin, 2012).

Michigan State University, University of Minnesota and University of California, University of Pittsburg in the USA have also initiated research data management (Max, 2016; Nordenberg, 2009; Rice, 2013; Courses: Research Data Management and Sharing, n.d.). Lund University, Sweden and La Trobe University, Australia have also initiated research data management (CSIR Central : Mandate, n.d.). It was observed that in most cases the research data gathered by researchers is stored in user devices and there was a lack of awareness of the importance of long-term preservation and sharing of research data. Users surveys have suggested a loss of productivity and resources. Research data management needs to be made a component of research policy of universities and research organizations so as to educate researchers (Trewhella, 2014). Te Pokai Tera University, New Zealand introduced a Research Data Management framework in 2016. A research data working group was established to diagnose status of research data in all units. After long deliberations of research data working group several key benefits are noted in the report. Recommendations have also been made to start programmes about knowledge and use of research data (MANTRA: Research Data Management Training, n.d.).

University of Edinburg started Research Data Management planning to support nationwide research data service. It has been initiated with a mission to change researcher perception regarding research data. The roadmap for management of research data has been framed with a governance model that includes combined efforts of academicians and technologists. The University developed DMPonline tool for research data management life cycle (Wright, 2012).

7. Learning Opportunities for Research Data Literacy

Research data literacy and data literacy are two related concepts while research data is a subset of data. The need of research data literacy has risen due to importance of data in research. Various Suitable training for research data management have been proposed: For example, the University of Edinburg offers a free 9-module online course, MANTRA (Shrivastava, & Gupta, 2018). University of North Carolina and the University of Edinburg offer a MOOC on Research Data Management and Sharing.

8. Research Data Literacy for India

Global review strongly indicates the importance of research data management. Preparedness for research data literacy in India cannot be assessed easily as all Indian research output is still not available in digital form. Research data understanding in Indian information professionals and research communities is developing slowly, after launching of programmes as Shodhganga – a national thesis and dissertation repository in 2009, National Data Sharing and Accessibility Policy – 2012 and Digital India programme in 2015 and The Data (Privacy and Protection) Bill in July 2017.

Indian government efforts at open data movement have been started by framing of National Public Information Infrastructure. Also, National Spatial Data Infrastructure as a specific division of Government of India has been established to evaluate open data projects worldwide. Such developments for open data in Indian government policies and practices are indicators of data intensive governance, which may be achieved by sharing research data created by government or public funded research.

CSIR, being one of the leading institutions for research in India is considered here to sense readiness. For this purpose, Scimago institutions ranking has been used taken. CSIR (Council of Scientific and Industrial Research) has got first rank in Scimago ranking.

CSIR is a leading research organization of India collaborating with industrial partners. Presently, it has 44 specialized laboratories which conduct research in various fields. Under CSIR open access mandate each laboratory is required to develop an (Institutional Repository (IR) to provide open access to full texts of research publications with in respective laboratory. There is also a CSIR-Central repository.

DataCite is a not for profit organization, which has developed Re3data.org, a registry to index research data repositories. Identification of IR as research data repository is done on various parameters (presented in Schema 3 of re3data.org).

Awareness of research data literacy in Indian research community can be gauged by development of IRs by CSIR laboratories. So, all working IRs of CSIR were submitted to registry of re3data.org.

Only 32 of the 44 CSIR laboratories have registered IRs (Annexure 1). Only 13 IRs were found working. Central Glass and Ceramic Research Institute has highest number of items registered.

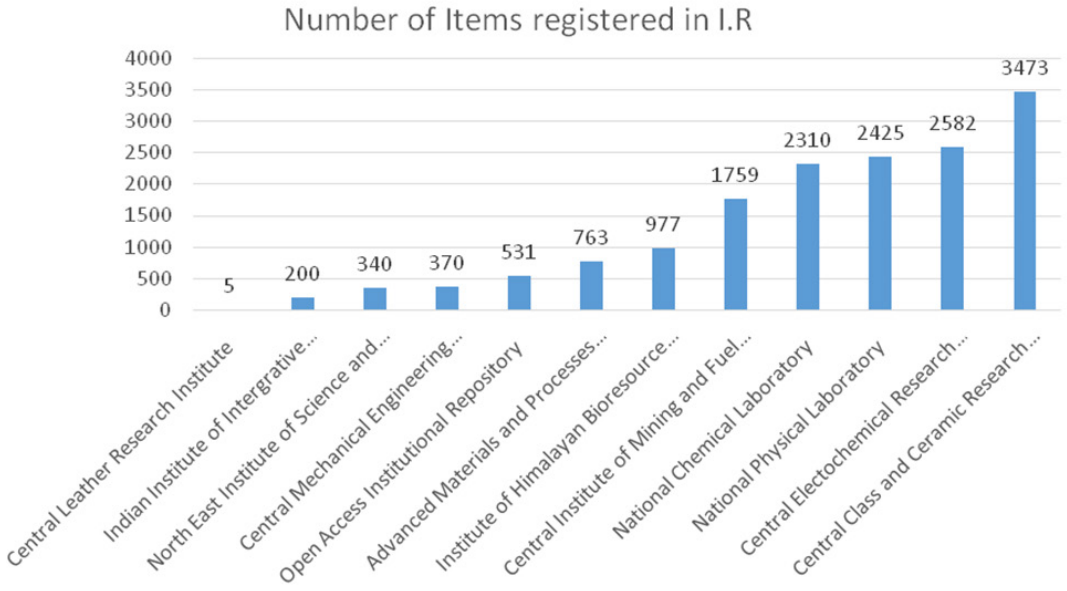

Figure 1. CSIR working I.R with number of registered items

Re3data.org registry has an open contribution service for research data repositories. IRs of all CSIR laboratories were submitted by the authors to the registry for evaluation. After analysis of the contents and its structure none of the IRs was found suitable to be considered as research data repository by re3data.org.

It was advised by the registry to develop focused research data preservation to qualify for registration.

A previous study by the authors had suggested that Re3data.org registry has around 43 Indian research data repositories. These are in preliminary phase of repository registration and the IRS are not populated with large data. This emphasises the need for research data literacy in India. In a recent development, Indira Gandhi National Centre for Arts has started a new part-time one-year Post Graduate Course on Digital Library and Data Management.

9. Conclusion

Research data management needs proper planning covering organizational and technological aspects. The planning phase of research data management requires mandatory research data literacy for researchers and stake holders. Management of research data is essential for present and future. It is widely recognized that research data literacy for successful research data management. Research data literacy should be an integral part of higher education curriculum. Long term and need based research data literacy need to be introduced in India for bringing out a change in the mindset about research data management.

10. References

Achard, P., Ayris, P., Fdida, S., Graddmann, S. (2013). *LERU roadmap for research data.*

Albert, P., Alpi, K., Baxter, P., Brown, E., Corson-rikert, J., Hirtle, P. and Westbrooks, E. L. (2008). *Digital research data curation: Overview of issues, current activities, and opportunities for the Cornell University Library* . Cornell University Library. Retrieved from http://hdl.handle.net/1813/10903.

Averkamp, S., Xiaomei, G., and Rogers, B. (2014). *Data management at the University of Iowa : A university libraries report on campus research data needs*. https://doi.org/10.17077/t5j5-ry3i

Beitz, A., Dharmawardena, K. and Searle, S. (2012). *Research data management strategy and strategic plan 2012-2015*. Monash Research Committee.

Brewerton, G., Steering Committee and IT Services . (2015 Oct). Research data management : A case study. *Ariadne, 74*, 1–11.

CARL Data Management Sub-Committee. (2009). *Research Data: Unseen opportunities*. Retrieved from http://www.carl-abrc.ca/uploads/pdfs/ data_mgt_toolkit.pdf.

Clements, A. (2013). Research information meets research data management in the library? *Insights: The UKSG Journal, 26*(3), 298–304.

Courses: Research Data Management and Sharing. Retrieved from https://www.coursera.org/learn/data-management.

CSIR Central : Mandate. Retrieved from www.csircentral.net/mandate.pdf.

Data guide. Retrieved from https://www.bu.edu/dev/data/guides.

Defining Research. Retrieved from https://www.lib.ncsu.edu/data-management/define.

Educause - Research Data Management Guidance. (2009). *Michigan State University*. Retrieved from http://net.educause.edu/ir/library/pdf/ers0908/rs/ers09087.pdf.

EPSRC expectations on Research Data Management. Retrieved from http://www.ucl.ac.uk/library/research-support/research-data/policies/epsrc/guidance.

Gurria, A. (2007). *OECD principles and guidelines for access to research data from public funding*. Organisation for Economic Co-Operation and Development, USA.

Gurria, A. (2007). *OECD principles and guidelines for access to research data from public funding*. Organisation for Economic Co-Operation and Development, USA.

Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday, 16*(2). https://doi.org/10.5210/fm.v16i2.3316

Hattwig, D., Bussert, K., Medaille, A. & Burgess, J. (2013). Visual literacy standards in higher education: New opportunities for libraries and student learning. *Libraries and Academy. 13*(1), 61–89. https://doi.org/10.1353/pla.2013.0008

Hey, A., Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research.

Johnsson, M., and Ahlfeldt, J. (2015). *Research libraries and research data management within the humanities and social sciences*. Lund University, Sweden.

Johnston, L. (2014). *A workflow model for curating research data in the University of Minnesota Libraries: Report from the 2013 Data Curation Pilot*, University of Minnesota, Digital Conservancy.

Lin, E. (2012). *The responsible stewardship of research data : A roadmap for the University of California*, Merced. University of California, Merced.

Managing research. Retrieved from https://www.monash.edu/library/researchdata.

Mandinach, E. B., and Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42*(1), 30–7. https://doi.org/10.3102/0013189X12459803

MANTRA: Research data management training. Retrieved from https://mantra.edina.ac.uk/.

Max, W., Howard, A., Lise, M., Brian, F., Shari, H., Lynch, H. et al. (2016). *Research data management framework report*, Te Pokai Tera University , New Zealand.

Nordenberg, M. (2009). *A guidelines on research data management*. University of Pittsburgh.

Research data policy. Retrieved from http://www.mopp.qut.edu.au/D/D_02_08.jsp.

Research data policy at Elsevier support services. Retrieved from https://www.elsevier.com/about/our-business/policies/research-data.

Rice, R., Çuna, E., Jeff, H., Sarah, J., Stuart, L., Stuart, M., and Weir, T. (2013). Implementing the research data management policy: University of Edinburgh Roadmap. *International Journal of Digital Curation, 8*(2), 194–204. https://doi.org/10.2218/ijdc.v8i2.283

Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., et al. (2015). *Strategies and best practices for data literacy education: Knowledge synthesis report*.

Shearer, K., Argaez, D., and Swanson, M. (2010). *Addressing the research data gap: A review of novel services for libraries*. Retrieved from http://forskningsdata.deff.wikispaces.net.

Shrivastava, P., and Gupta, D. K. (2018). Research data preservation in India: An analysis based on research data registry. *World Digital Library: An International Journal, 11*(1).

Steen, L. (1999). Numeracy: The new literacy for a data-drenched society. *Educational Leadership, 57*(2), 8–13.

Trewhella, J. (2014). *Research data management policy 2014*, Trobe University.

University of Edinburgh data policy. Retrieved from https://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy.

Vahey, P., Yarnall, L., Patton, C., Zalles, D., Swan, K. (2006). *Mathematizing middle school: Results from a cross-disciplinary study of data literacy*. Annual Meeting of the American Educational Research Association, San Francisco, CA.

Woollard, M., and Corti, L. (2014). *Case study 4: A national solution - the UK Data Service*. Delivering Research Data Management Services: Fundamentals of good practice, Facet Publishing.

Wright, G., Abraham, S., and Shah, N. (2012). *Open government data study: India*. (2012).

Annexure 1

Sl.no	Institutional Repository	CSIR Institute	Items	Full text	Open access	Working Status
1	ir@ampri	Advanced Materials and Processes Research Institute	763	13%	2%	Working
2	ir@csir-ceeri	Cental Electronics Engineering Research Institute	NA	NA	NA	Working
3	ir@cbri	Central Building Research Institute	N/A	N/A	N/A	N/W
4	ir@cgcri	Central Class and Ceramic Research Institute	3473`	83	6	Working
5	ir@cdri	Central Drug Research Institute	N/A	N/A	N/A	N/W
6	ir@cecri	Central Electochemical Research Institute	2582	98	46	Working
7	ir@cftri	Central Food Technological Research Institute	N/A	N/A	N/A	Working
8	ir@cimfr	Central Institute of Mining and Fuel Research	1759	43	6	Working
9	ir@clri	Central Leather Research Institute	5	100	100	Working

10	ir@cmeri	Central Mechanical Engineering Research Institute	370	0	0	Working
11	ir@crri	Central Road Research Institute	N/A	N/A	N/A	N/W
12	ir@csmcri	Central Salt & marine Chemicals Research Institute	N/A	N/A	N/A	N/W
13	ir@c-mmacs	Centre for Mathematical Modelling and Computer Simulation	N/A	N/A	N/A	N/W
14	ir@iict	Indian Institute of Chemical Technology	N/A	N/A	N/A	N/W
15	ir@iiim	Indian Institute of Intergrative Medicine	200	100	1	Working
16	ir@iip	Indian Institute of Petroleum	N/A	N/A	N/A	N/W
17	ir@iicb	Institute of Chemical Biology	N/A	N/A	N/A	N/W
18	ir@igib	Institute of Genomics and Integrative Biology	N/A	N/A	N/A	N/W
19	ir@ihbt	Institute of Himalayan Bioresource Technology	977	83	11	Working
20	ir@imtech	Institute of Microbial Technology	N/A	N/A	N/A	N/W
21	ir@immt	Institute of Minerals and Materials Technology	N/A	N/A	N/A	N/W
22	IR@NAL	National Aerospace Laboratories	N/A	N/A	N/A	N/W
23	ir@ncl	National Chemical Laboratory	2310	77	49	Working
24	IR@NEERI	National Environmental Engineering Research Institute	N/A	N/A	N/A	N/W
25	IR@NIIST	National Institute for Interdisciplinary Science and Technology	N/A	N/A	N/A	N/W
26	IR@NIO	National Institute of Oceanography	N/A	N/A	N/A	N/W
27	IR@NISCAIR	National Institute of Science Communication and Information Resources	N/A	N/A	N/A	N/W
28	IR@NML	National Metallurgical Laboratory	N/A	N/A	N/A	N/W
29	ir@npl	National Physical Laboratory	2425	100	29	Working
30	ir@neist	North East Institute of Science and Technology	340	100	100	Working
31	ir@neeri	Open Access Institutional Repository	531	99	16	Working
32	IR@SERC	Structural Engineering Research Centre	N/A	N/A	N/A	N/W