DESIDOC Journal of Library & Information Technology, Vol. 39, No. 6, Nov 2019, pp. 280-289, DOI : 10.14429/djlit.39.6.15137

Content Analysis of Indian Research Data Repositories: Prospects and Possibilities

Raj Kumar Bhardwaj

St. Stephen's College, University of Delhi, India
E-mail: raajchd@gmail.com

ABSTRACT

The study aims to trace the development of Indian research data repositories (RDRs) and explore their content with the view of identifying prospects and possibilities. Further, it analyses the distribution of data repositories on the basis of content coverage, types of content, author identification system followed, software and the application programming interface used, subject wise number of repositories etc. The study is based on data repositories listed on the registry of data repositories accessible at http://www.re3data.org.The dataset was exported in Microsoft Excel format for analysis. A simple percentage method was followed in data analyses and results are presented through Tables and Figures. The study found a total of 2829 data repositories in existence worldwide. Further, it was seen that 1526 (53.9 %) are open and 924 (32.4 %) are restricted data repositories. Also, there are embargoed data repositories numbering 225 (8.0 %) and closed ones numbering 154 (5.4 %). There are 2829 RDRs covering 72 countries in the world. The study found that out of total 45 Indian RDRs, only 30 (67 %) are open, followed by restricted 12 (27 %) and 3 (6 %) that are closed. Majority of Indian RDRs (20) were developed in the year 2014. The study found that the majority of Indian RDRs (17) are'disciplinary'. Further, the study also revealed that statistical data formats are available in a maximum of 31 (68.9 %) Indian RDRs. It was also seen that the majority of Indian RDRs (28) has datasets relating to 'Life Sciences'. It was identified that only 20% of data repositories have been using metadata standards in metadata; the remaining 80% do not use any standards in metadata entry. This study covered only the research data repositories in India registered on the registry of data repositories. RDRs not listed in the registry of data repositories are left out.

Keywords: India; Research data; Data repositories; License; Dataverse; Author identification.

1. INTRODUCTION

Research datasets are being generated in ever-increasing volume in different formats[1-2]. Computational methods have spurred research across nations, and experts of diverse fields have produced huge amounts of research data[3]. Scientific research data illustrate a heterogeneity that varies across disciplines, and between research groups and research scholars[4]. Data generated during research work are normally collected as part of the academic research process[5]. Such research data needs to be managed through research data repositories before the datasets deteriorate[6-8]. Nonetheless, the research data repository must frame clear data depository agreements written in plain language so that the depositor can follow certain data submission rules. It should be noted that data producers are accountable for quality of research data while the repository is answerable for the quality of storage and accessibility of data[9]. Therefore, data repository should be developed in every research institution to increase visibility and discoverability of research data[10]. Moreover, good data management practices need to be promoted by institutions. Also, publishers and funding bodies should come forward to advocate linking of research data with publication. Research data repositories need to build trust and a sense of prestige among the academic community so that researchers' concerns and interests can be protected.

Indian Government has mandated data sharing and framed the National Data Sharing and Accessibility Policy (NDSAP). The policy authorises all Government and Government funded institutions to publicly share data which are produced through the Government funded research[11]. The objective of the policy is to facilitate access of data to the public in machine-readable formats over the internet. The Government of India has prepared implementation guidelines which instruct that different types of datasets generated by different ministries and/or departments ought to be classified as shareable data and non-shareable data. The data sharing policy of the Government of India consists of the following principles: machine-readable, openness, flexibility, transparency, quality and security. Furthermore, the Government of India has developed an open list and a negative list of research data. Negative list contains datasets which are confidential and can compromise the country's security, and personal information of citizens. The positive list covers datasets which do not fall under the negative list. The Government of India has defined three types of access of data i.e., open access, registered access and restricted access. Open access facilitates a timely and user-friendly way without any process of registration or validation, while in case of registered access data, an individual has to go through a prescribed

Received : 24 Sept 2019, Revised : 2 November 2019
Accepted : 5 November 2019, Online published : 11 December 2019

registration process. Data in restricted access can be accessed only after authorisation[12].

The current study traces the growth of research data repositories (RDRs) in India. Besides this, the study ascertains types of contents, author identification system followed, software used, application programming interface, license used, auxiliary features in Indian RDRs. Outcome of the study will help to understand the environment of RDRs, identify shortcomings of Indian RDRs, and guide them and their funders so that global standards can be maintained. Tracing the growth of data repositories in India is a valuable piece of essential documentation. Also, a comparaison of global and Indian repositories can give valuable hints for developing and improving repositories in India. Besides this, RDRs in India are established by various institutions and it is cumbersome for researchers, publishers and academic institutions to identify the appropriate RDR and their features. In India, increasing globalisation and digitisation have brought numerous challenges in the management of research data. In recent years in India concerns over data management have been at centre stage so that better policy decisions can be taken based on data[13]. The study strives to achieve the following objectives:

- To trace year wise development of research data repositories (RDRs) in India
- To identity content types in research data repositories in India
- To understand the author identification system followed in managing data in research data repositories
- To comprehend application programming interfaces (API) and certificate followed in Indian RDRs
- To identify the data licenses, data access and data upload restrictions followed in Indian research data repositories
- To ascertain software (s) and metadata standards being used in Indian RDRs
- To know the subject coverage of Indian RDRs.

2. LITERATURE REVIEW

Various research articles have been reviewed pertaining to issues associated with research data management and data repositories. Yoon and Schultz[14] conducted content analyses study in the United States to examine research data management services. Authors found that libraries need to advance and engage more actively to provide data services to library users. The study also found wide variation among library data management services. European Commission[15] confirms in a report that data repositories in Europe are heterogeneous. Therefore, a strategy should be developed "to overcome the fragmentation and enable research communities to better manage, use, share and preserve data". Gómez, Mendez and Hernández-Pérez[4] analysed the contents of data repositories in the field of social sciences and humanities, and observed that data and metadata schemas are less homogeneous in humanities than in social sciences. Furthermore, they found that the trends of Data Documentation Initiative (DDI) metadata schema usage are apparent in social sciences. However, authors postulated that it may be because of maturity of standards and number of implementations. Borgman[1] highlighted that metadata is crucial to make the data valuable in describing, dimensioning and contextualising. Consequently, data can be found irrespective of varied disciplines, and enable reuse, crossing subject boundaries. Force and Auld[16] highlighted that disciplinary data repositories differ in the metadata elements and the appropriate level of granularity for data citation. Further, they postulated that representatives of data repositories and publishers ought to coordinate with each other to make data citable and discoverable for the various stakeholders in the data lifecycle. Uzwyshyn[17] found that 74 per cent research institutions provide data archiving services and only 13 per cent have data-specific repositories. Interestingly, 13 per cent researchers use more general digital repositories and 74 per cent use temporary text-centric repositories, in place of data repositories to comply grant guidelines. Pinnick[18] examined geosciences data retention requirements and concluded that the academic community demanded data that are reliable, reusable, in preferred formats. The author emphasised that the creation of high-quality metadata is vital and depositors of data should be trained to submit data set conveniently to the repositories.

Greenberg et al.[19] studied the metadata best practices followed in Dryad data repository which includes two points; first, it has been established to address the Semantic Web technology using metadata application profile and second, to make content available in DSpace using an extensible markup language (XML) schema. Therefore, the present study has also considered metadata standard in contents analysis of Indian data repositories. Metadata Working Group[20] devised the schema and instructions for managing large amount of datasets. Nevertheless, this schema and policies of data repositories ought to be reviewed periodically. The report highlighted that several groups have been working to incorporate various other elements in Dublin Core schemas. Therefore, the format of data and data upload provisions have been covered in the contents analysis of Indian data repositories in the present study. Si et al.[21] examined the current status of research data services in 87 university libraries and found that 50 libraries had data repositories and offered data services of six types viz., conceptual understanding of research data and its formats, data management planning guide, data storage, curation, training, data management reference and resource recommendation. Furthermore, the authors found that research data introduction is the most frequently provided service in libraries. Another study by Austin et al.[22] described that Research Data Canada Standards and Interoperability Committee (RDCSINC) surveyed 32 online data repositories and found heterogeneity of features and services in surveyed online data platforms. Further, it also found non-standardised use of terms, uneven compliance and a less certification in online data repositories. Grunzke, R. et al[23] postulated that research data should be stored in a structured way so that it can be discovered conveniently. Consequently, data in the data repositories can be accessed by content and context. Further, it was suggested that usage of metadata shall be automatic and seamless to foster usability. Hence, the above studies describe that research data in RDRs should be organised in a structured way. However, no content analysis study has been conducted on Indian RDRs. Therefore, the present study not only enriches literature in the field of library and information science, but also helps developers and administrators, cataloguers, policymakers of RDRs to take appropriate decisions in the data repository life cycle.

3. METHODOLOGY AND SCOPE OF THE STUDY

Content analysis method was used in this study. This method assists to make replicable and valid inferences by interpreting the contents of text data. Content analysis was found most suitable because it permits examination of data services and contents listed in data repositories. The study used data repositories registered on the registry of research data repositories accessible at: http://www.re3data.org to understand the development of Indian research data repositories and conduct content analysis. The registry re3data.org offers researchers orientation in the heterogeneous landscape of RDR. It provides information to users on their diverse roles as producers of data and users of data. The registry facilitates publishers and academic institutions to identify research data repositories so that researchers can deposit and share research data at the most appropriate site. Besides this, re3data.org also foresees the establishment of a more coherent and integrated ‘‘ecosystem of data repositories’’.

Each data repository was examined through content analysis in the study because this was the easiest vehicle to comprehend data equity and access. The selected registry of research data repositories is a global initiative covering research data repositories from diverse academic disciplines. The registry presents the list of repositories for permanent storage and access of data sets to researchers, publishers, as well as academic and research institutions. Indian research data repositories listed in the registry were identified for content analysis. In all, 45 Indian RDRs are listed in the registry and all were selected for the study. There is no record revealing the total number of RDRs in India. Therefore, the RDRs not indexed in re3data.org were excluded in the study.

Once the subset of Indian RDRs was identified, a quantitative analysis of the metadata record was done. Each data repository was accessed to verify the number of records listed and the technology used in managing the research data. Author spent two hours per day from May 1 to 30, 2018; total 60 hours were invested in access and validation. The requisite data was obtained manually from each data repository listed in the registry. The dataset was exported in Microsoft Excel format for tabulation and for generating statistical figures. Subsequently, the Microsoft Excel dataset was imported into the Statistical Package for Social Sciences (SPSS) version 20 for statistical analysis and interpretation for achieving the objectives of the study. The analysed data is presented in Tables and Figures. The parameters of content analysis of these Indian RDRs were as follows: year of establishment, types of RDRs viz., disciplinary institutional, others and types of content. Moreover, data relating to the availability of AID Systems and APIs in each RDR have also been collected. In addition, data from each repository were collected to know whether the repository is data provider or service provider. The data were collected pertaining to provision of data access licenses, data upload, software used, subject coverage and other auxiliary features in Indian RDRs. All the above parameters were chosen as variables for content analysis of Indian RDRs.

4. RESULTS

Results obtained after analysing the dataset are presented in Tables I-II and Figures I-VII. The study found that a total of 2829 data repositories are listed in the registry of research data repository. Interestingly, out of these, 1782 are open research data repositories. The following types of research data repositories were found viz., disciplinary 1289 (72.3 %), institutional 390 (21.9 %), and others (commercial, portals etc.) 103 (5.8 %). It was identified that majority (596) of data repositories worldwide use the World Data System (WDS) certification, followed by Digital Signature Algorithm (DSA), Rat fürSozial- und Wirtschaftsdaten (RatSWD), Common Language Resources and Technology Infrastructure (CLARIN), CoreTrustSeal, Deutsche initiative fürnetzwerkinformation E.V (DINI) certificate, DeutschesInstitutfürNormung (DIN 31644). The least used certification in the data repositories was identified as Trustworthy Repositories Audit & Certification (TRAC).

4.1 An Overview of Research Data Repositories Worldwide

Country wise number of data repositories was ascertained. It was found that 2829 data repositories are established in 72 countries. Figure 1 shows that majority of research data repositories have been established in Europe 1295 (45.8 %), followed by North America 1138 (40.2 %), Asia 176 (6.2 %), Australia 93 (3.3 %), South America 19 (0.7 %) and Africa 13 (0.5 %). Besides this, 95 research data repositories have been established with the collaboration of various countries in different continents.

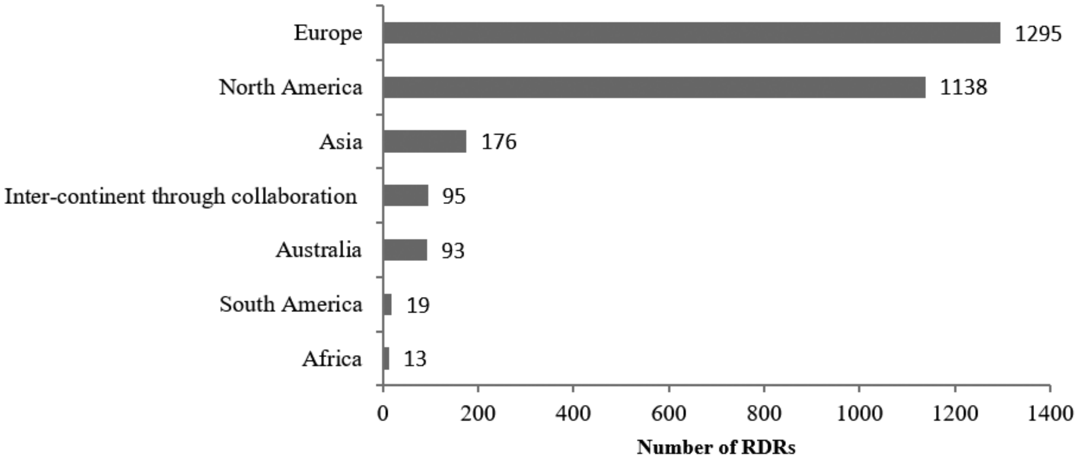

Figure 1. Continent-wise number of research data repositories.

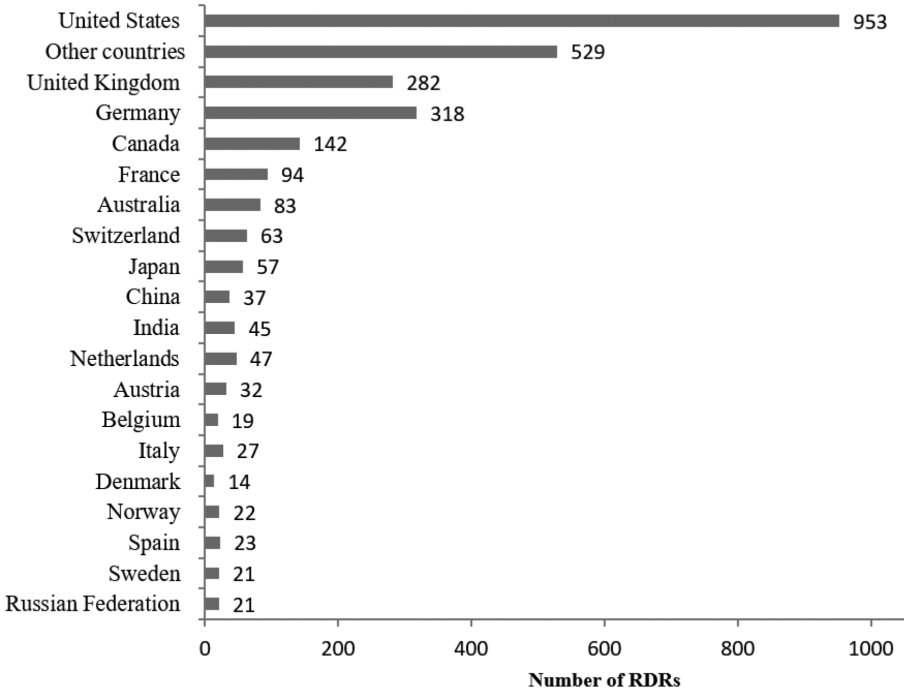

Figure 2. Leading countries in establishing research data repositories.

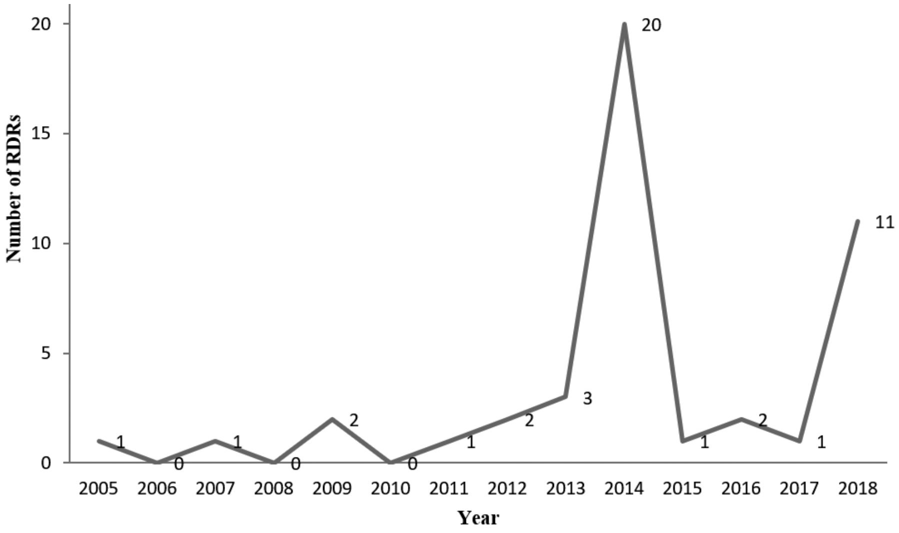

Figure 3. Year wise growth of research data repositories in India.

Figure 2 shows that United States is the leading country with 953 data repositories (33.7 %), followed by Germany 318 (11.2 %), United Kingdom 282 (10.0 %), Canada 142 (5.0 %), France 94 (3.3 %), Australia 83 (2.9 %), Switzerland 63 (2.2 %), Japan 57 (2.0 %), Netherlands 47 (1.7 %), China 37 (1.3 %), Austria 32 (1.1 %), India 45 (1.6 %) and Italy 27 (1.0 %). Countries, where the number of data repositories is found less than 1.0 % of the total, are: Spain 23 (0.8 %), Norway 22 (0.8 %), Sweden 21 (0.7 %), Russia 21 (0.7 %) and Denmark 20 (0.7 %). Besides this, the remaining countries collectively have established 372 (13.1 %) data repositories.

4.2 Growth of Research Data Repositories in India

The present study found that 45 research data repositories have been developed in India. Figure 3 illustrates the yearly growth which reveals that the first research data repository was established in 2005 and the second in 2007. Two research data repositories were developed in 2009 and three in 2013. A huge increase in the number of research data repositories was registered in 2014; 20 research data repositions were developed that year. Subsequently, growth remained stagnant until 2017. Eight research data repositories have been established in 2018. It is expected that the number of research data repositories would increase in the years to come.

4.3 Research Data Repository Types

It is seen that mainly three type of data repositories are popular viz., disciplinary, institutional and others. However, if the data type does not fall under any of these categories of data repositories, other types of repositories are used for data deposit. Generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data, supplementing primary data in a data-type-specific repository. Yu[24] highlighted that two noticeable trends are prevalent worldwide i.e., more engagement and enlarged scope, in data-driven services. Moreover, the study found that researchers prefer submitting research publications to institutional repositories and selected a data repository suitable for their data.The present study ascertained types of data repositories in India. Figure 4 shows that the majority of (17) data repositories are ‘disciplinary’, followed by ‘’others’ type of data repositories (9) and ‘institutional’ data repositories (1). In addition, the study identified that eight data repositories are

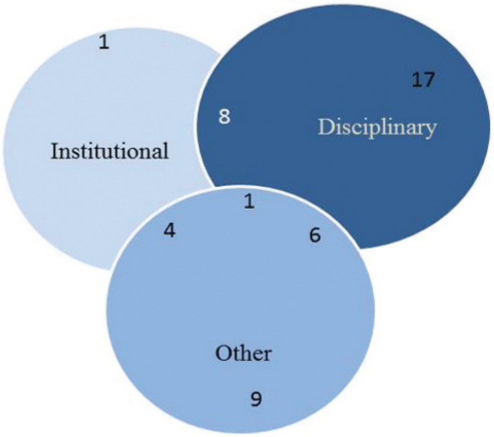

Figure 4. Types of research data repositories in India.

Table 1. Growth of research data repositories vs. open access research data repositories

Country	Number of RDR and percentage n= 2829	Number of OA RDR & percentage n=1782	Rank
United States	953	750	1
United Kingdom	282	188	2
Germany	318	178	3
Canada	142	75	4
France	94	65	5
Australia	83	54	6
Switzerland	63	41	7
Japan	57	38	8
China	37	35	9
India	45	35	10
Netherlands	47	29	11
Austria	32	29	12
Belgium	19	17	13
Italy	27	26	14
Denmark	14	12	15
Norway	22	15	16
Spain	23	15	17
Sweden	21	14	18
Russian Federation	21	13	19
Greece	11	8	20
Mexico	11	9	21
Israel	10	7	22
New Zealand	9	8	23
Taiwan, Province of China	9	9	24
Ireland	7	8	25
Czech Republic	8	7	27
Brazil	6	7	28
Others	458	91	

'disciplinary-cum-institutional', six data repositories fulfill the criteria of 'disciplinary-cum-others', four data repositories fall under the criteria of 'institutional-cum-others', and only one data repository fulfills all the criteria i.e., 'institutional', 'disciplinary' and 'others'.

4.4 Content Types

This study ascertained types of content in the data repositories. A file format encodes information within a computer file, and recognises the application and access. Research data is generated in various formats e.g., text, multimedia, numeric, software, structured graphics, images etc. File name extension is generally indicated by a full stop (dot) followed by three letters. File format enables the computer to recognise whether the file should be processed as text or video. Proprietary formats confine to software patents or built-in encryption, to prevent open usage[25]. Content type should opt for long-term access and preservation of data. Subsequently, sharing among a wider circle of researchers must be ensured. Hence, it is recommended to choose open standards and formats that are easy to reuse. The format being used in data repositories must be included in the documentation. It helps when files are migrated to their preservation formats, as well as for any specific software that would be necessary to view or work with the data. Data can be categorised into five main categories viz., observational, experimental, simulation, derived or compiled, reference or canonical. Data repository management must understand that the category chosen for repository would have impact throughout the rest of the data management plan. Thus, the choice of data types has a crucial place in research data management. Table 1 reveals that scientific and statistical data formats are available in maximum 31 (68.9 %) data repositories, followed by standard office documents 24 (53.3 %), structured graphics 21 (46.7 %), plain text 18 (40.0 %), raw data 17 (37.8 %), images 15 (33.3 %), others 8 (17.8 %), archived data 8 (17.8 %), databases 7 (15.6 %), structured text 6 (13.3 %), audio-visual data 3 (6.7 %), networked data 3 (6.7 %). Besides

Table 2. Distribution of data repositories on the basis of content types

Content type	Number (n=166)	Number (n=45)
Scientific and statistical data format	31	0.69
Standard office documents	24	0.53
Structures graphics	21	0.47
Plain text	18	0.40
Raw data	17	0.38
Images	15	0.33
Other	8	0.18
Archived data	8	0.18
Databases	7	0.16
Structured text	6	0.13
Audio visual data	3	0.07
Networked data	3	0.07
Source code	3	0.07
Software applications	2	0.04
Configuration data	0	0.00

this, the study found source code in three (6.7 %) research data repositories, followed by a software application in two (4.4 %) research data repositories. Surprisingly, no research data repository in India preserves configuration data.

4.5 Data Provider vs. Service Provider, and the Policy Framework

The study analysed datasets to know whether research data repositories are data providers or service providers. Figure 5 illustrates that the majority of research data repositories in India are data providers, 28 (62.2 %), and only four (8.9 %) are service providers. Besides this, 13 (28.9 %) of the research data repositories play dual roles as data provider and service provider. Data provider who offers research data and its metadata exposes metadata to users. It creates the dataset and distributes among users. However, a service provider harvests the metadata of the research data from data providers and adds value to provide the service[26].

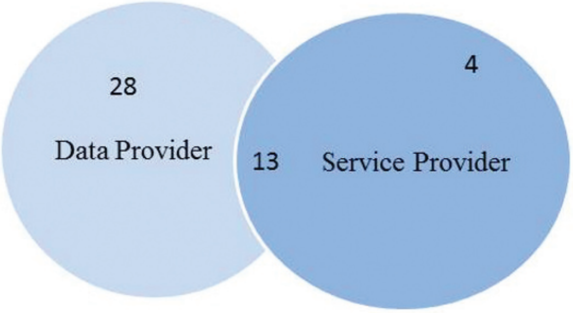

Figure 5. Data provider vs. service provider research data repositories.

Furthermore, the study found that 13 (28.9 %) research data repositories have provision of versioning and 32 (71.1 %) do not have provision of versioning. Also, 14 (31.1 %) research data repositories have given citation guidelines while 31 (68.9 %) do not have provision for citation guidelines.

Scholarly contents on the Internet is growing at a rapid pace. Hence, integration of all the scientific information by linkage helps to keep the publication process smooth and efficient. Therefore, publications should also link the associated data, resulting in 'Enhanced Publications'[27]. Enhanced Publications should capture data in digital format and facilitate database deposit alongside manuscript publication. An enhanced publication should make articles fully machine-readable, providing intelligent markup and structured digital abstracts[28]. The study also explored the provision of enhanced publications in research data repositories and found that 16 (35.6 %) have provision of enhanced publication; 19 (42.2 %) research data repositories have listed 'unknown' for enhanced publication and eight (17.8 %) mentioned they do not have the provision of enhanced publications.

4.6 Distribution of Research Data Repositories by AID Systems and APIs

Author Identification (AID) helps the manager of the data repository to identify works of an individual author. Unique author identification is vital in data repository services for meeting search requirements of users. Author identification that is universally unique and persistent helps users to identify contents of the appropriate author. Therefore, the study ascertained numerous AIDs used in research data repositories. It was found that the majority of data repositories, 29 (64.4 %), do not use any author identification and API in research data management. AID is being used in four data repositories; 41 research data repositories do not use any AID. Further, it was found that only 12 (26.7 %) of research data repositories have been using APIs and 33 (73.3 %) do not use APIs. Major benefits of using APIs in data repositories are to perform default tasks, insert or update any action needed etc. It was found that the majority of research data repositories using APIs, six(50.0 %), used the REST API. Besides this, file transfer protocol (FTP), NetCDF, OAI-PMH, SOAP, SPARQL and SWORD are being used by one research data repository each.

4.7 Provision of Data Access Licenses And Provision of Data Upload

The study tried to understand data access level provisions in Indian research data repositories. The level of data access could be 'restricted', 'closed' or 'unrestricted'. Research data are being produced by researchers in varied disciplines around the world. The planned release of research data has become common practice. Therefore, the need for data licensing arises after data release[29]. Releasing data after defining certain terms and conditions may be counter productive. The legal position needs to be defined regarding use and application. Thus, licenses become imperative in releasing research data to data repositories. The study ascertained data access and licenses used in data repositories in India. Figure 6 illustrates that the highest number of data repositories in India, 30 (67 %), are open, followed by restricted 12 (27 %) and three (6 %) closed. Data repositories that are restricted require registration to access the contents.

Furthermore, the study ascertained licenses used in Indian data repositories. It was found that 25 (55.6 %) data repositories

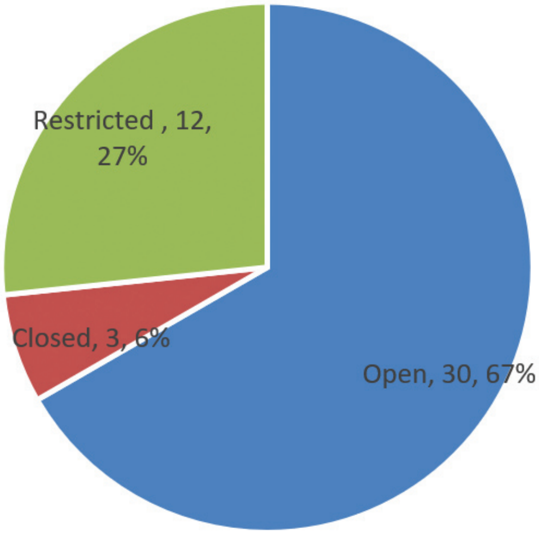

Figure 6. Database access of data repositories in India.

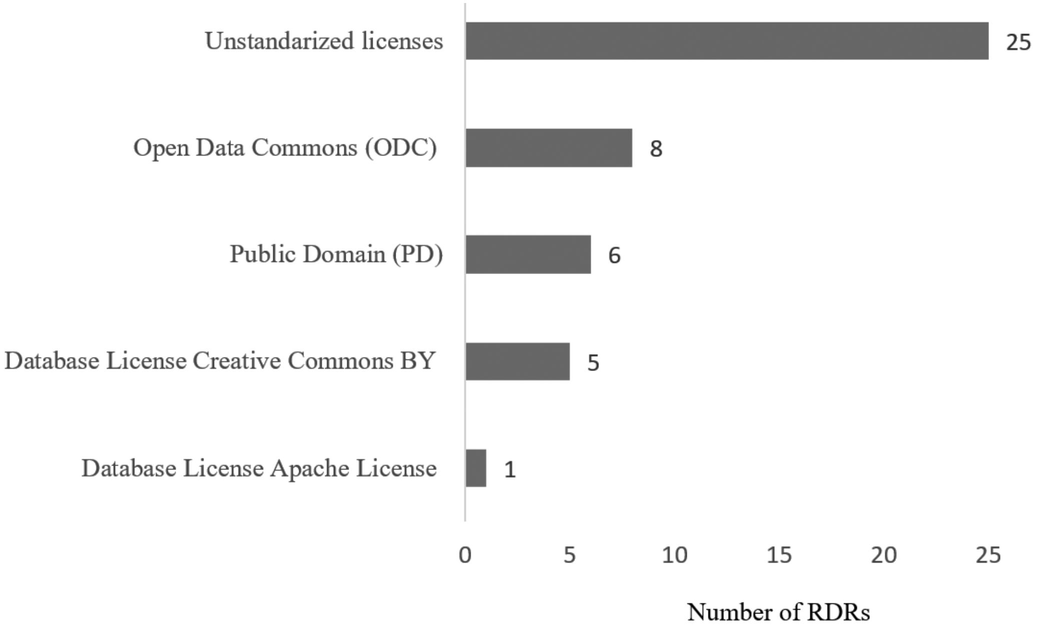

Figure 7. Database licenses used in indian data repositories.

have non-standardised licenses and 20 (44.4 %) use standard licenses. Further, the dataset was analysed to know the license type being used. Figure 7 illustrates that a maximum number of data repositories, 25, use non-standardised licenses that are not popular worldwide. (These licenses are not standard licenses and are being prepared in-house by repository developers.) They are followed by Open Data Commons (ODC) (8), public domain (6), creative commons (5) and Apache License (1).

This study reveals uploading provisions in 45 Indian data repositories. Majority of data repositories, 20 (44.4 %), data uploading is 'closed', followed by 'restricted' 7 (15.6 %), 'restricted-registration' 7 (15.6 %), 'restricted-institutional member' 7 (15.6 %), 'restricted-institutional member' 4 (8.9 %), 'restricted-other' 4 (8.9 %). It is discouraging to see that only 3 (6.7 %) data repositories have 'open' provision of data uploading. Hence, volunteer contributors cannot upload data in the majority of Indian data repositories.

4.8 Software(S), Persistent Identifier and Metadata Standards Used

The study analysed software(s) used in Indian data repositories and the metadata standards used. The majority of data repositories 23 (51.1 %) use 'unknown software'; 'other' is mentioned by 10 (22.2 %). Besides this, it was found that 6 (13.3 %) data repositories mention that they are not using any software, and 2 (4.4 %) have not mentioned name of the software. Interestingly, only one data repository namely, ICSSR data repository, is developed using 'Nada' and only one namely, CSISA Data Repository, is using 'Dataverse' software.

Furthermore, it was found that majority of Indian RDRs, 34 (75.6 %), do not use persistent identifier, and only 11 (24.4 %) have the provision of the persistent identifier in accessing the datasets. The study found that digital object identifier (DOI) is the most popular persistent identifier in Indian data repositories. Besides this, meta-standards used in data repositories were investigated. It was identified that only 20% data repositories use metadata standards in metadata entry and the remaining 80% do not use any standards in metadata entry. IMEx data repository is using '*Minimum Information for Biological and Biomedical Investigations (MIBBI)*' standard; '*ISO 19115, DCC*' standard is being followed by IODP data repository, '*DDI - Data Documentation Initiative, DCC*' is being followed by ICRISAT Research Data repository and '*Dublin Core*' is being used by CSISA Data Repository. 'DCC' is followed in metadata creation in the National Genomic Resources Repository while '*Ecological Metadata Language*' is used in India Biodiversity Portal (IBP). 'DDI' metadata standard is being followed in Indian Council of Social Science Research (ICSSR), and Census Repository has been using the Developed Metadata Schemas, DCC.

4.9 Subject Coverage of Research Data Repositories

The study analysed data on the basis of subjects to comprehend subject wise coverage of Indian research data repositories. It was found that the majority of research data repositories (28) in India have content relating to 'Life Sciences'. In these 28 research data repositories, maximum contents relate to 'biology', 'medicine'. Second highest data pertains to the subject 'Natural Sciences', which is available in (21) research data repositories. In these research data repositories, the most popular subjects relating to Natural Sciences are 'Geophysics and Geodesy'. Physics, Geography, water research are popular subjects. Third highest number of research data repositories relate to 'Humanities and Social Sciences' (17), wherein maximum data relates to the subjects Social and Behavioural Sciences, Economics and Humanities. Fourth highest number of data sets pertain to 'Agriculture, Forestry, Horticulture and Veterinary Sciences, (6) and fifth highest is 'Engineering Sciences' (3). The total number of research data repositories subject-wise exceeds 45 because there are repositories which deal with more than one subject area. However, data in the social sciences and humanities are not collected digitally all the time. In archeology, the results of observational data can be more closely linked to the object, photographs, or videos[30].

4.10 Provision of Auxiliary Features

The study also explored auxiliary features available in Indian data repositories. It was found that all the 45 data repositories provide additional information to their service and 38 (84.4 %) data repositories have framed policies so that users can understand repository policies in details. Also, 38 data repositories have mentioned uniform resource locator (URL) of the detailed policy paper so that users can download the policy document to read and study. Besides this, 15 (33.3 %) data repositories have given the application programming interface (API) and only 3 (6.7 %) data repositories have the provision of data dissemination policy. Interestingly, all the data repositories provide additional information to its service in the data repository. Further, it was identified that only 13 (28.9 %) data repositories in India have mentioned 'data citation guidelines' on the website and 17 (37.8 %) data repositories have indicated a 'mission statement'. 'Quality management' is another significant parameter in a data repository and it was found that only 19 (42.2 %) data repositories have 'quality management' system. In addition, only 9 data repositories provide 'alerting services'; 26 (57.8 %) do not have such provision. Furthermore, it was found that out of nine data repositories, five provide really simple syndication (RSS); ATOM and REST are being provided by two data repositories each.

5. DISCUSSION

Researchers seem hesitant in contributing data to repositories. This is a major cause of concern since researchers already have several issues in sharing data. Therefore, library professionals should be proactive in adopting a standard license so that more researchers can share their data without any hesitation. Several libraries tend to deliver some basic services through their websites e.g., overview of data management, while significantly fewer libraries provide a detailed level of information about data documentation, administration, and reuse. Libraries have been providing useful information to users about research data management through their websites, which is relatively easy and a good starting point[14]. The cordial relationship between the librarian and administrators of other departments within an organisation is crucial to develop a successful data repository[31]. Research data management services are being provided by a variety of academic and research libraries across the world. As a result, data management competencies render the librarian's job much more demanding[32]. Nevertheless, the majority of research data repositories are developed because of funders' prerequisites and the number of repositories does not seem enough for managing research data across the globe. Out of 2829 data repositories worldwide, only 1526 (53.9 %) are open. Countries have different data policies e.g., the SciELO Open Access publishing platform, which originated in Brazil, has now been taken up in a number of other countries. A repository charging deposit fees for data supporting publications may be a natural extension of charging article processing fees for publishing in some countries and disciplines, while it may not be applicable in other countries[33]. Moreover, considering the socio-economic benefits of making research data open, a significant move towards open data access calls for trained manpower who can collect, store, manage, reuse and make research data openly accessible in academic and research institutions. It is anticipated that data specialists in large companies in the United Kingdom would increase at a growth rate of 243 % in five years[34]. Several Information Schools around the world have started short term courses in data management to train students. Nevertheless, information schools should not merely focus on training of students, but must aim to educate and train the academic and research community so that data collection, storage, use and sharing are optimised. Majority of Indian RDRs, 20, were developed in the year 2014. The study ascertained that the majority of Indian RDRs (17) are 'disciplinary'. It was found that statistical data formats are available in maximum 31 (68.9 %) Indian RDRs. It was also found that the majority of Indian RDRs (28) have datasets relating to 'Life Sciences'. Unknown software is being used in maximum Indian RDRs, 23 (51.1 %). This shows that either developer of these repositories are not aware of open source software(s) available to developed RDRs, or are not comfortable in using the existing software(s). It was identified that only 20% data repositories have been using metadata standards in metadata entry and the remaining 80% do not use any standards in metadata entry. Researchers use the minimum required approach in metadata entry while uploading data to a data repository. Therefore, on the data repository, it is not sufficient to mention the data upload mechanism[35]. Interestingly, metadata can be accurately assigned to publications by librarians. However, assigning metadata to datasets also requires the contribution of researchers. Therefore, knowledge of the domain and recording the dataset production context are utmost required[1].

6. CONCLUSIONS

Academic libraries are sources of research support in research data management and understand the needs of researchers in developing data services[36]. However, there is not much discussion on research data management (RDM) and the relationship between data sharing, development of policies and practices in academic institutions. No emphasis seems to have been given to data sharing and researchers' concerns to protect their rights on the data[1],[37]. Research data generated after collaborative research projects in universities must be shared in a controlled, organised and structured manner[38]. To fulfill this, administrators must engage in dialogues and debates on open research data. It is observed that some institutions embrace the RDM challenge proactively whereas others rely on funders' requirements to manage research data[39].

Nowadays researchers collecting large datasets have enhanced knowledge and skills in managing data. These competencies are vital to ensure data quality, integrity, shareability and reuse of data. Therefore, the funding agencies have consistently formulated regulations for submission of research data. Besides this, some funding agencies mandate the submission of a research data management plan along with the research proposal[40]. Librarians, researchers, data specialists and administrators ought to work together to transform data management practices within the research community. Making research data openly accessible is not a new idea, nonetheless its adoption among the research community has been slow. Therefore, funding agencies, research organisations and researchers should come forward in this data-centric world so that new hypotheses can be proposed and tested to achieve better results. Moreover, considering the socio-economic benefits of making research data open, a significant move towards open data access would call for a body of trained manpower that can collect, store, manage, reuse and make research data openly accessible in academic and research institutions. It is anticipated that the number of data specialists in large companies would increase at a growth rate of 243 % in a five-year-period[33]. Several Information Schools around the world have started short term courses in research data management to train students. Nevertheless, information schools should not merely focus on training students, but must educate and train the academic and research community so that data collection, storage, use, and sharing can be achieved. Besides this, empirical and conceptual work should be done by students of LIS schools on the research data management to enhance understanding of the reality of research data. Further, we need to explore how research data can be used for the needs and objectives of research evaluation[41].

The study analysed Indian RDRs on the basis of subject coverage, software used, data access and restriction, data licenses and metadata standards followed, content type etc. Hence, the findings of the study can be used by researchers, librarians, data scientists, to identify the appropriate RDR suitable to the need for their research. However, it did not explore quality of data repository services and the actual role of library professionals in managing research data repositories. In addition, the impact of research data repositories on research scholars and faculty members was also not explored. The content analysis does not provide information about usefulness from the users' perspectives or how librarians work with the researchers. Thus, a further expanded study may be useful to comprehend the role and interaction of librarians with researchers in developing a research data repository. Besides this, further study may also be conducted on RDRs of other countries using the same methodology.

REFERENCES

1. Borgman, Christine L. Data, disciplines, and scholarly publishing. *Learned Publishing*, 2008, **21**(1), 29-38.
   doi: 10.1087/095315108X254476.
2. Berman, F. & Cerf, V. Who will pay for public access?, *Science*, 2013, **341**(6146), 616-617.
   doi: 10.1126/science.1241625.
3. Tripathi, M.; Chand, M.; Sonkar, S.K. & Jeevan, V.K.J. A brief assessment of researchers' perceptions towards research data in India. *IFLA J.*, 2017, **43**(1), 22-39.
   doi: 10.1177%2F0340035216686984.
4. Gómez, N.D.; Méndez, E. & Hernández-Pérez, T. Social sciences and humanities research data and metadata: A perspective from thematic data repositories. *El Prof. De La Inf.*, 2016, **25**(4), 545-555.
   doi: 10.3145/epi.2016.jul.04.
5. Pryor, G. Why manage research data?, In Managing Research Data, edited by G Pryor, Facet Publishing, London, 2012, 1-16.
   doi: 10.29085/9781856048910.002.
6. Cox, A.M. & Pinfield, S. Research data management and libraries: Current activities and future priorities. *J. Libr. Inf. Sci.*, 2013, **46**(4), 1-18.
   doi: 10.1177/0961000613492542.
7. Higgins, S. The lifecycle of data management, In Managing Research Data, edited by G. Pryor, Facet Publishing, London, 2012, 17-46.
   doi: 10.29085/9781856048910.003.
8. Lavoie, B.F. Sustainable research data, In Managing Research Data edited by G. Pryor, Facet Publishing, London, 2012, 67-82.
   doi: 10.29085/9781856048910.005.
9. Green, A.; Macdonald, S. & Rice, R. Policy-making for research data in repositories: A Guide, EDINA, 2009, available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.218.467&rep=rep1&type=pdf. (accessed on 1 February 2019).
10. Wolski, M. & Richardson, J. A framework for university research data management. CCA-EDUCAUSE Australasia Conference, Sydney, Australia, 2011. Available at: http://www98.griffith.edu.au/dspace/bitstream/handle/10072/39672/69936_1.pdf (accessed on 1 February 2019).
11. Government of India. National data sharing and accessibility policy-2012, NDSAP-2012, available at: http://ogpl.gov.in/NDSAP/NDSAP-30Jan2012.pdf (accessed 10 May 2018).
12. Open Government Data (2015). Implementation guidelines for national data sharing and accessibility policy (NDSAP), available at: https://data.gov.in/sites/default/files/NDSAP_Implementation_Guidelines_2.2.pdf (accessed 7 May 2018).
13. PWC (2018). An overview of the changing data privacy landscape in India. available at: https://www.pwc.in/assets/pdfs/publications/2018/an-overview-of-the-changing-data-privacy-landscape-in-india.pdf (accessed on 1 February 2019).
14. Yoon, A. & Schultz, T. Research data management services in academic libraries in the US: A content analysis of libraries' websites. *College Res. Libr.*, 2018, **78**(7), 920-933.
    doi: 10.5860/crl.78.7.920.
15. European Commission (2009). ICT Infrastructures for e-Science, Available at https://ec.europa.eu/eurostat/cros/system/files/COM%20%282009%29%20ICT%20INFRASTRUCTURES%20FOR%20e-SCIENCE.pdf (accessed on 20 December 2018).
16. Force, M.M. & Auld, D.M. Data citation index: Promoting attribution, use and discovery of research data. *Inf. Serv. Use*, 2014, **34**(1-2), 97-98.
    doi: 10.1007/s10822-014-9768-5.
17. Uzwyshyn, R. Research data repositories: The what, when, why and how, 2016. Accessible at: https://digital.library.txstate.edu/handle/10877/7597. (accessed on 1 February 2019).
18. Pinnick, J. Exploring digital preservation requirements: A case study from the National Geoscience Data Centre (NGDC). *Records Manage. J.*, 2017, **27**(2), 175-191.
    doi: 10.1108/RMJ-04-2017-0009.
19. Greenberg, J.; White, H.C.; Carrier, S. & Scherle, R. A metadata best practice for a scientific data repository. *J.*

Libr. Metadata, 2009, **9**(3-4), 194-212.
doi: 10.1080/19386380903405090.

20. Metadata Wokring Group (2015). Data cite international data citation metadata working group. Data Cite metadata schema for the publication and citation of research data. Version 3, 2015, Available at. http://schema.datacite.org/meta/kernel-3.1/doc/DataCite-MetadataKernel_v3.1.pdf. (accessed on 21 September 2019).
21. Si, L.; Xing, W.; Zhuang, X.; Hua, X. & Zhou, L. Investigation and analysis of research data services in university libraries, *Electron. Libr.*, 2015, **33**(3), 417-449.
doi: 10.1108/EL-07-2013-0130
22. Austin C.C.; Brown, S.; Fong, N.; Humphrey, C.; Leahey, A. & Webster, P. Research data repositories: Review of current features, gap analysis, and recommendations for minimum requirements. *IASSIST Q.,* 2015, **39**(4), 24-38.
doi: 10.29173/iq904.
23. Grunzke, R.; Hartmann, V.; Jejkal, T.; Kollai, H.; Prabhune, A.; Herold, H.; Deicke, A.; Dressler, C.; Dolhoff, J.; Stanek, J. & Hoffmann, A. The MASi repository service — Comprehensive, metadata-driven and multi-community research data management. *Future Gener. Comput. Syst.*, 2019, **94**(879-94).
doi: 10.1016/j.future.2017.12.023.
24. Yu, H.H. The role of academic libraries in research data service (RDS) provision: Opportunities and challenges, *Electron. Libr.*, 2017, **35***(*4), 783-797.
doi: 10.1108/EL-10-2016-0233.
25. The University of British Columbia Library (2018), available at: http://guides.library.ubc.ca/ld.php?content_id=12389607. (accessed on 21 September 2019)
26. Vierkant, P.; Spier, S.; Rücknagel, J.; Pampel, H. & Gundlach, J. (2013). Schema for the description of research data repositories, *re3data. Org, Version*, *2*., accessible at: http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:758898:6/component/escidoc:775891/re3data_schema_v2-2_public_final-2014-12-03.pdf (accessed 12 August 2018).
27. Woutersen-Windhouwer, S.; Brandsma, R.; Hogenaar, A.; Hoogerwerf, M.; Doorenbosch, P.; Dürr, E.; Ludwig, J.; Schmidt, B. & Sierman, B. (2009), "Enhanced publications: linking publications and research data in digital repositories", accessible at: https://pure.uva.nl/ws/files/1101083/72745_311760.pdf (accessed on 9 February 2019).
28. Seringhaus, M.R. & Gerstein, M.B. Publishing perishing? Towards tomorrow's information Architecture. *BMC Bioinf.*, 2007, **8**(1), 17.
doi: 10.1186/1471-2105-8-17.
29. Ball, A. (2014). How to license research data. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides. (accessed 9 July 2018).
30. Frank, Rebecca D.; Yakel, Elizabeth & Faniel, Ixchel M. Destruction/reconstruction: preservation of archaeological and zoological research data". *Archival Sci.*, 2015, **15**(2), 141-167.
doi: 10.1007/s10502-014-9238-9.
31. Pinfield, S.; Cox, A.M. & Smith, J. Research data management and libraries: Relationships, activities, drivers and influences, *PloS One*, 2014, **9**(12), p.e114734.
doi: 10.1371/journal.pone.0114734.
32. Knight, G. Building a research data management service for the London School of Hygiene & Tropical Medicine, *Program: Electron. Libr. Inf. Syst.*, 2015, **49**(4), 424-439.
doi: 10.1108/PROG-01-2015-0011.
33. E-skills Report (2013). Big data analytics: Adoption and employment trends, 2012-2017, available at: http://www.voced.edu.au/content/ngv:59261. (accessed 1 July 2018).
34. OECD (2017), "Business models for sustainable research data repositories", *OECD Sci., Technol. Ind.Policy Pap.*, No. 47, OECD Publishing, Paris.
doi: 10.1787/302b12bb-en.
35. Piwowar, H.A. & Chapman, W.W. Public sharing of research datasets: A pilot study of associations, *J. Informetrics*, 2010, **4**(2), 148-156.
doi: 10.1016%2Fj.joi.2009.11.010.
36. Akers, K.G. & Doty, J. Disciplinary differences in faculty research data management practices and perspectives, *Int. J. Digital Curation*, 2013, **8**(2), 5–26.
doi: 10.2218/ijdc.v8i2.263.
37. Wallis, J.C.; Rolando, E. & Borgman, C.L. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology", in NunesAmaral, L.A. (Ed.), *PloS ONE*, 2013, **8**(7), e67332.
doi: 10.1371/journal.pone.006733.
38. Corti, L.; Van den Eynden, V.; Bissell, A. & Woollard, M. *Manage. Sharing Res. Data: A Guide to Good Pract.*, SAGE Publications Ltd, Los Angeles, CA, 2014.
39. Rice, R. & Southall, J. *Data Libr. Handbook*. Facet Publishing. London, 2016.
40. Holdren, J.P. (2013), "Memorandum for the heads of executive departments and agencies: Expanding public access to the results of federally funded research", Executive Office of the President, Office of Science and Technology Policy, Washington, DC, February 22, available at: www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf (accessed 12 May 2018).
41. Schöpfel, J.; Prost, H. & Rebouillat V. Research data in current research information systems, *Procedia Comput. Sci.*, 2017, **106**(1), 305-20.
doi: 10.1016/j.procs.2017.03.030.

CONTRIBUTOR

Dr Raj Kumar Bhardwaj is a Librarian at St. Stephen's College, University of Delhi (India). He holds MCA, MLIS and M.Phil, PhD from University of Delhi. He has also qualified UGC-NET. He has published 4 book, 40 research paper in various reputed journals, and has delivered several invited talks at various international conferences and seminars. He is the reviewer for five international LIS journals of repute and elected member of IFLA Standing Committee of Indigenous Matters and corresponding member of Law Libraries Section.