Journal of Scientometric Res. 2019;8(3):135-142
http://www.jscires.org

Research Article

Open Research Data Repositories: A Content Analysis to Comprehend Data Equitable Access

Raj Kumar Bhardwaj
St Stephen's College, University of Delhi, Delhi, INDIA.

ABSTRACT
The present study pertains to content analysis of open Research Data Repositories (RDRs) worldwide to comprehend the growth and development in the area. The study is original because no attempt has been made so far to do a content analysis of these open research data repositories. The study used the content analysis method of open RDRs listed on the registry of research data repositories. The dataset was analyzed using Microsoft Excel. Each unique parameter was given a unique variable for the purpose of analysis. Subsequently, the dataset was analyzed using Microsoft Excel to achieve the objectives of the study. A simple percentage method was followed in analyses and is presented through Tables and Figures. The study found that there are 1997 RDRs are open worldwide, which are indexed in the registry of research data repositories. It was also found that out of these data repositories 1509 (75.6 percent) are 'disciplinary', 'institutional' 398 (19.9 percent) and 'others' 90 (4.5 percent). Majority open RDRs worldwide use the World Data System (WDS) certification followed by Core Trust Seal. The study found total 70 countries in the world have 1997 RDRs and the United States is the leading country with 704 (35.3 percent) open RDRs, followed by Germany 201 (10.1 percent), United Kingdom 183 (9.2 percent), Canada 118 (5.9 percent). Scientific and statistical data formats are available in maximum (1238) open research data repositories, followed by standard office documents (1108), images (1047), plain text (935), raw data (881). It is ascertained that 1049 (52.5) RDRs are not following any metadata standard in open RDRs and 948 (47.5 percent) follow metadata standards. It was ascertained that Dublin Core is being used by (237) data repositories, followed by ISO 19115 (128), Data Documentation Initiative (DDI) (125). It was found that 963 (48.2 percent) data repositories are using unknown and in-house developed software worldwide. Majority of research data repositories contains the material relating to 'Biology' 715 (23.0 percent). Besides this, the study found that open data repositories have been developed in 60 languages worldwide.

Keywords: Research Data, Content Analysis, Data Repository, Researchers, Open Data.

Correspondence
Raj Kumar Bhardwaj
St Stephen's College, University of Delhi, Delhi-110007, INDIA.
E-mail: raajchd@gmail.com

Received: 18-07-2019;
Revised: 27-08-2019;
Accepted: 22-11-2019.
DOI: 10.5530/jscires.8.3.29

INTRODUCTION

In present times, a number of methods are in use to produce various formats of data.[1,2] Scientific research data show a variety across disciplines within research groups and research scholars.[3] Besides this, nowadays, a number of funding agencies have also mandated to share and host their research data in data repositories. However, we do not make any efforts to host the datasets and ask the authors to submit in any suitable open data repositions. Therefore, data should be submitted, giving priority to the discipline-specific and recognized repositories by the community.[4] Funding agencies in the western world made it mandatory to provide a data management plan (DMP) along with a research proposal so that data can be preserved, shared and reused. Research data management addresses data curation and all challenges relating to data collection, description, preservation, sharing, dissemination, reuse and retention.[5] Building a research data repository is a complex process and person responsible for the same have to collaborate with various departments to make it a reality.

The major inhibits in research data management are a multiplicity of disciplines and data types in the discovery of datasets. Furthermore, major concerns are because of inconsistency within disciplines regarding what data are archived, where to archive and at what level to make it discoverable. Besides this, if researchers are unable to locate data conveniently or unable to generate information out of data, then it is located data in the research office. Therefore, we should make dataset discoverable and user-friendly for the users to extract the information out of data conveniently.[6] Various reputed publishers have also taken the lead in data deposition and framed the guidelines and made it obligatory for authors to deposit data associated with the research paper at a disciplinary or subject-specific data repository. Such provision without the broader legal binding could be an effective incentive for researchers

to share their research datasets.[7,8] Nonetheless, it is utmost require for adequate prevention procedure for adhering to the quality of research data repository. Therefore, good data management not only require monitoring and proper action but also need quality data. In order to achieve it, management practices should be followed and metadata from preceding accidents should be better used.[9]

The present study is an attempt to know the open research data repositories growth and development around the world. The study also determines types of contents, author identification system followed, software used, application programming interface, license used, auxiliary features in open RDRs around the world. The present study will help to understand the environment of open RDRs, identify the limitation of open RDRs and guide them and their funders so that global standards can be maintained. 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.

Literature review

Data repositories have grown exponentially over the years and researchers have been educated about the benefits of depositing their data in data repositories. Therefore, the literature review was conducted to understand the data repositories, data sharing, data curation, data reuse to understand overall development in the area. Cheek and Bradigan[10] conducted a study at the United States and Canada and found that just 12.2% of these libraries provided support for "data curation". Steinhart *et al*.[11] revealed only a few university libraries actually involved in research data curation and suggested that libraries should take the lead in research data management. Kuipers and Van der Hoeven[12] revealed accountability for publicly funded research, the inspiration for scientific advancements and re-analysis of previously generated data are the three reasons for research data preservation. Aydinoglu, Dogan and Taskin[13] highlighted that research scholars in Turkey are familiar with the benefits of data management and open to share their datasets. Nevertheless, library professionals lack the competencies and deep knowledge about the subject. Besides this, no support is available to these professionals to initiate the research data management in their institutions.

Consequently, they faced several inhibits in setting up the data repositories and formulating the data management planning draft. The European Commission has been pushing open access since long and encouraged the researchers to publish their research in open access journals and self-archive their publication in an institutional repository.[14] Thus, recognizing the significance of this movement, OECD countries agreed to the Declaration on Access to Research Data for Public Funding in 2004.[15,16] Mayernik[17] highlighted the Australian National Data Services (ANDS) which facilitate users to access research data generated in Australian universities. Furthermore, described that data service also encourages in collaboration in sharing, handling and publishing and reusing datasets. Henderson and Knott[18] examined the research data repository at Virginia Commonwealth University libraries and found that library staff plays a pivotal role in sensitizing and encouraging data sharing. Aydinoglu, Dogan and Taskin[13] found that datasets are being stored at different places and 45.9 percent of students use cloud while graduate assistants (58.9 percent) use the cloud double as much as the professors (30.8 percent). Perrier *et al*.[19] examined the central data repository at Columbia University and emphasized the impact of data repositories.

Faniel and Connaway[20] opined that library professionals have successfully established RDM programs on campus to support the needs of research scholars. However, professionals can further make an impact on the support of the researchers' more efficiently. Furthermore, authors postulated that human resource, technical skills, leadership, researchers' perception, collaboration with human resources and leadership are the key to achieve the goals of research data management. MacMillan[6] recommended that persistent linking of data and publications using identifier have made datasets more accessible. Therefore, the author suggested to provide the persistent identifier to data and link the same to publication so that readers can be encouraged to use the datasets. Hruby *et al*.[21] mentioned about implementing an open-source centralized research data repository (CRDR) and revealed that user acceptance was tested using pretest and post-test and workflow efficiency. Besides this, the quantity and quality of publications were also considered in the process. Yoon and Schultz[22] expressed that financial problems are the major inhibits in developing data repository. Similar views are expressed by Erway and Rinehart[23] viewed that research data management services are costly to process and it requires it require sustainable efforts and funding to achieve the desired goal. Farnel and Shiri[24] stated that research data repository use variety of metadata elements and controlled vocabularies was common across the services. Further, stressed that preservation and unique identifiers are central components of data repository.

Objectives of the study

The study is conducted to comprehend the development of open research data repositories over the years around the world so that clear ideas relating to the same can be illustrated. It was found that there is no document is available online and print to know the status of the same. Thus, the study was initiated to frame clear documentation about open data repositories. The study endeavors to achieve the following objectives:

- To know the country-wise number of open RDRs around the world;

- To identify content types in the open RDRs worldwide;
- To understand the author identification the system followed in managing open RDRs;
- To comprehend API and certificate followed in these open RDRs;
- To identify the data licenses followed in these repositories;
- To ascertain software (s) used and metadata standards being used in open RDRs; and
- To find subject wise number of open RDRs worldwide.

Need and purpose

Research data repositories are a vital collection, curation, preservation, dissemination and reuse of data. However, no content analysis study has been done so far about the open research data repositories so that what their features, a number are of open data repositories worldwide, subject categorization of these repositories. No attempt has been made so far to do a content analysis of these open research data repositories. Therefore, the study is utmost needed to know a number of open research data repositories and availability of features in these data repositories. The study will add valuable literature in the library and information science and help researchers to comprehend the development of data equitable access.

Methodology and Scope of the study

Content analysis is a research technique used to validate inferences using the existing coding text and to interpret the inferences. It evaluates the existing text, systematically converting the quantitative data into qualitative data. Subsequently analyzing the data to achieve the objectives of the study. The method has been used in the library and information science since a long time however no research has been conducted has been.

The content analysis method has been followed in the study to achieve the objectives of the study. The content analysis method each metadata details given on the registry of research data repositories to comprehend the services and standard followed by open data repositions worldwide. The data of the study have been obtained from the registry of research data repositories accessible at https://www.re3data.org/. Wherever data given in the registry is not articulate individual data repository was accessed to capture the details of the elements. Besides this, the study used data repositories listed on the registry of research data repositories and identified the open RDRs for the purpose of content analysis.

The dataset containing the information about open RDRs was exported in Microsoft Excel format. Subsequently, the data were analyzed using Microsoft Excel. Each unique parameter was given a unique variable for the purpose of analysis. Subsequently, the dataset was analyzed using Microsoft Excel to achieve the objectives of the study. A simple percentage method was followed in analyses and is presented through Tables and Figures. Each data repository was also accessed to validate the number of records listed and other details about the repository. Notably, the researcher spent two hours per day during June 1 to 7, 2019 and total of 20 hr to access and validate the data. Interestingly, the data repositories which do not provide transparent information about open data access and relevant parameters of the study were disqualified from the study. The main parameters of the study are as follows:

- Number of open RDRs worldwide;
- Content types in open RDRs;
- Unique identifiers;
- API and certificate followed in these open RDRs;
- Data licenses, data access;
- Data upload restriction;
- Software (s) used and metadata standards; and
- Categorization of open RDRs on the basis of subject

Number of factors as per the objectives of the study were used to collect the data pertaining to open RDRs. The scope of data repositories confined to open RDRs which are listed on the registry of research data repositories. Those data registry which are not listed in the registry were excluded from the study.

Limitations of the study

The study is conducted only on RDRs listed on the registry of research data repositories. The repositories which are not listed could not be included in the study. Besides this, the study does not cover the quality of data repository services and the actual role of library professionals in managing a research data repository. In addition, the impact of research data repositories on research scholars and faculty members is also not explored. Another limitation is that it does not compare the RDRs with closed and paid RDRs.

RESULTS

The results obtained after analyzing the dataset are presented in Tables 1 and 2 and Figures 1-5 below. It is apparent that research data service is popular worldwide. The study found that there are 1997 data access is open and we considered these are open research data repositories worldwide, which are indexed in the registry of research data repositories. It was also found that out of these data repositories 1509 (75.6 percent)

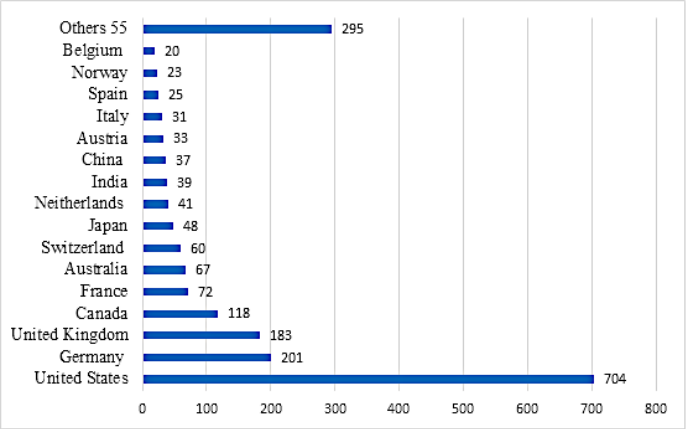

Figure 1: Leading Countries in Establishing open RDRs.

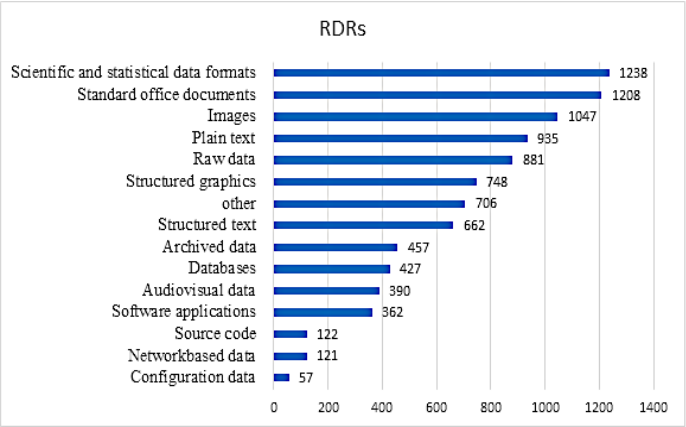

Figure 2: Distribution of Data Repositories on the Basis of Contents.

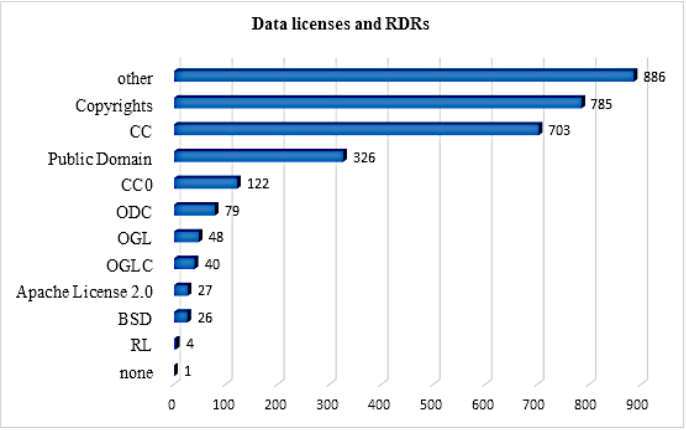

Figure 3: Data Licenses Used.

are 'disciplinary', 'institutional' 398 (19.9 percent) and 'others' 90 (4.5 percent).

Majority open RDRs worldwide use the World Data System (WDS) certification followed by Core Trust Seal, Deutsche initiative für netzwerk information E.V (DINI) certificate, Digital Signature Algorithm (DSA), Rat für Sozial- und Wirtschaftsdaten (RatSWD), Common Language Resources

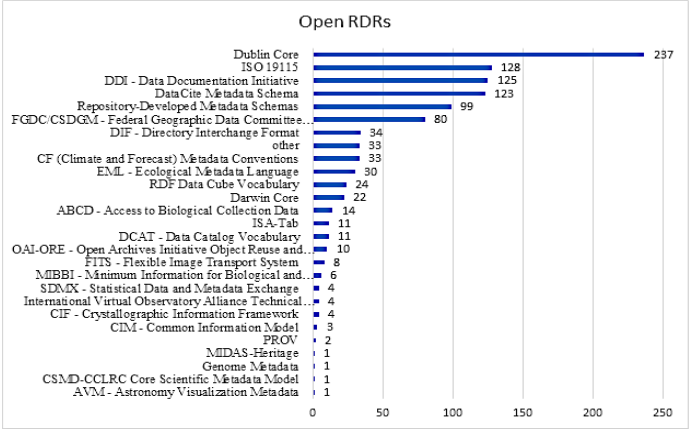

Figure 4: Metadata Standard Used in Open RDRs.

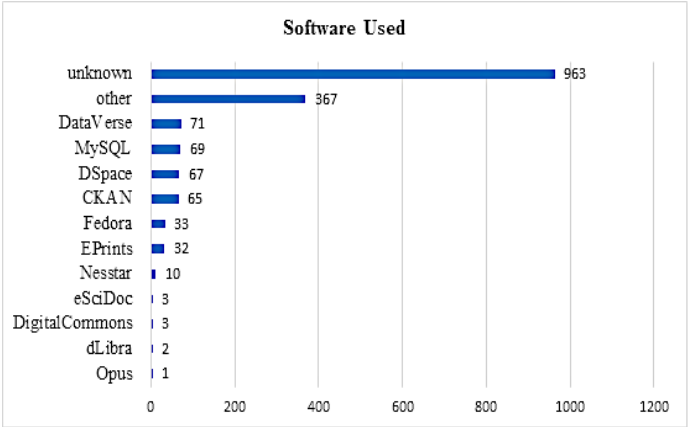

Figure 5: Software used in open RDRs.

and Technology Infrastructure (CLARIN), Deutsches Institut für Normung (DIN 31644). The least used certification in the data repositories is identified as Trustworthy Repositories Audit and Certification (TRAC).

Country-wise distribution of data repositories

Country-wise number of data repositories was ascertained. It was found that a total of 70 countries in the world have 1997 RDRs. Table 1 below shows that the United States is the leading country with 704 (35.3 percent), followed by Germany 201 (10.1 percent), United Kingdom 183 (9.2 percent), Canada 118 (5.9 percent), France 72 (3.6 percent), Australia 67 (3.4 percent), Switzerland 60 (3.0 percent), Japan 48 (2.4 percent), Netherlands 41 (2.1 percent), China 37 (1.9 percent), Austria 33 (1.7 percent), India 39 (2.0 percent) and Italy 31 (1.6 percent), Spain 25 (1.3 percent), Norway 23 (1.2 percent) and Belgium 20 (1.0 percent).

Besides this, the remaining countries collectively have established 295 (14.8 percent) open RDRs. Countries have different data policies, e.g., the SciELO Open Access publishing platform. A repository charging deposit fees for data supporting publications may be a natural extension of charging

Table 1: Country-wise distribution of Open RDRs.

Country	Number	Percentage
United States	704	35.3%
Germany	201	10.1%
United Kingdom	183	9.2%
Canada	118	5.9%
France	72	3.6%
Australia	67	3.4%
Switzerland	60	3.0%
Japan	48	2.4%
Netherlands	41	2.1%
India	39	2.0%
China	37	1.9%
Austria	33	1.7%
Italy	31	1.6%
Spain	25	1.3%
Norway	23	1.2%
Belgium	20	1.0%
Others 55 countries	295	14.8%
	1997	100.0%

Table 2: Subject wise distribution of research data repositories.

Subject	Number	Percentage
Biology	733	25.54%
Geosciences (including Geography)	590	20.56%
Medicine	278	9.69%
Social and Behavioural Sciences	291	10.14%
Physics	243	8.47%
Humanities	207	7.21%
Chemistry	169	5.89%
Agriculture, Forestry, Horticulture and Veterinary Medicine	156	5.44%
Computer Science, Electrical and System Engineering	98	3.41%
Construction Engineering and Architecture	38	1.32%
Materials Science and Engineering	26	0.91%
Mathematics	24	0.84%
Thermal Engineering/Process Engineering	12	0.42%
Mechanical and industrial Engineering	5	0.17%

Note: *Number of data repositories exceeds the total because one repository may cater more than one subject.*

article processing fees.[25,26] However, considering the socio-economic benefits of making the research data open a significant move towards open data access.[26] Information schools should focus on educating and training the academic and research community so that data collection, storage, use and sharing can be achieved.[29]

Distribution of data repositories by content types

Content-type should have opted for long-term access and preservation of data. Subsequently, sharing among wider researchers must be ensured. Thus, 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 the files are migrated to their preservation formats, as well as for any specific software that will be necessary to view or work with the data. Data can be categorized in mainly five categories viz., observational, experimental, simulation, derived or compiled, reference, or canonical. The data repository management must understand that the category chosen for the repository would then have an effect throughout the rest of the data management plan. Thus, choosing the data types will have a crucial impact on research data management. Figure 2 depicts reveals that scientific and statistical data formats are available in maximum (1238) data repositories, followed by standard office documents (1108), images (1047), plain text (935), raw data (881), structured graphics (748), structured text (662), archived data (457), audio-visual data (390), databases (427), software application (362), source code (122) and configuration data (57).

Distribution of Data Repositories by AID Systems and APIs

The study ascertained various AIDs used in research data repositories. The study found that the majority of data repositories (1867) do not use any author identification in research data management. However, ORCID is being used by (121) data repositories, followed by Author Claim (4), Researcher ID (2) and (3) repository uses the ISNI for author identification. Six repositories use other unknown AID in research data management.

Similarly, Application Programming Interfaces (APIs) used in various data repositories help to perform the default task, insert or update any action needed in data repositories. It was found that REST is being used in (345) data repositories, followed by File Transfer Protocol (FTP) used in (303), followed by OAI-PMH (168), Net CDF (70), SOAP (60), Open DAP (45), SWORD (31) and SPARQL (31). However, 248 data repositories have not mentioned the APIs being used by them.

Data Access and Certificates

The study examined data access policies of research data repositories and found that 1997 RDRs are open. The study also explored the standard certificate being followed in these repositories and identified that only 181 repositories are using

the standard certificates, out of 1997 in which maximum use the World Data System (WDS) (51), Core Trust Seal (44), followed by DSA (25), CLARIN certificate B (23), (Rat SWD) (18), DINI certificate (6) and TRAC (1), DIN 31644 (1). Notably, (12) RDRs using other types of license and 1816 RDRs do not use any kind of certificate.

Data licenses and metadata standards followed

The legal position needs to be defined regarding its use and application. Thus, licenses become imperative in releasing research data to the data repositories. It is also found that there are many RDRs which have been using more than one license in the repository. Figure 3 shows that majority of data repositories, 886 (29.1 percent) mentioned 'other' types of license being used by open RDRs it means a license which is not standard and 'Copyright' is being used by 785 (25.8 percent), followed by 'CC0' 703 (23.1 percent), 'Public Domain' 326 (10.7 percent), Open Data Common (ODC) 79 (2.6 percent), Open Government License (OGL) 48 (1.6 percent), Open Government Licence Canada (OGLC) 40 (1.3 percent), Apache License 2.0 27 (0.9 percent), Berkeley Software Distribution (BSD) 26 (0.7 percent) and Restrictive Licence (RL) 4 (0.1 percent).

In addition, the study also explored the metadata standard followed and identified that maximum repositories do not use any metadata standards. Out of the 1997 data repositories, only 1049 (52.5) RDRs are not following any metadata standard in open RDRs and 948 (47.5 percent) follow metadata standards. It was ascertained that Dublin Core is being used by (237) data repositories, followed by ISO 19115 (128), Data Documentation Initiative (DDI) (125), Data Cite Metadata Schema (123), Repository-Developed Metadata Schemas (99), Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (FGDC/CSDGM) (80), Directory Interchange Format (DIF) (34), Climate and Forecast (CF) Metadata Conventions (33), Ecological Metadata Language (EML) (30), RDF Data Cube Vocabulary (24), Darwin Core (22), Access to Biological Collection Data (ABCD) (14), ISA-Tab (11), Flexible Image Transport System (FITS) (8), Data Catalog Vocabulary (DCAT) (7), Open Archives Initiative Object Reuse and Exchange (OAI-ORE) (7), Minimum Information for Biological and Biomedical Investigations (MIBBI) (6), Crystallographic Information Framework (CIF) (4), Statistical Data and Metadata Exchange (SDMX) (4), International Virtual Observatory Alliance Technical Specifications (4), Common Information Model (CIM) (3), PROV (2). Besides this, Astronomy Visualization Metadata (AVM), Core Scientific Metadata Model (CSMD-CCLRC) and Genome Metadata Standard (GMS), MIDAS-Heritage is being followed in one data repository each.

Software Used in Research Data Management

Several software(s) are being used in research data management around the world. Figure 4 illustrates that 963 (48.2 percent) data repositories are using unknown and in-house developed software, followed by Data Verse 71 (3.6 percent), MySQL 69 (3.5 percent), DSpace 67 (3.4 percent), CKAN 65 (3.3 percent), Fedora 33 (1.7 percent), EPrints 32 (1.6 percent), Nesstar 10 (0.5 percent), eSciDoc 3 (0.2 percent); Digital Commons 3 (0.2 percent) dLibra 2 (0.1 percent) and Opus is being used in one data repository. It is found that 311 RDRs are not using any software.

Subject Coverage of research data repositories

Data analyses were done to know subject coverage of research data repositories worldwide. Table 2 shows that majority of research data repositories are related to '*Biology*' 715 (23.0 percent), followed by '*Geosciences' (including Geography)* 627 (20.1 percent), '*Medicine*' 492 (15.8 percent), '*Social and Behavioural Sciences*' 315 (10.1 percent), '*Physics*' 259 (8.3 percent), '*Humanities*' 194 (6.2 percent), '*Chemistry'* 180 (5.8 percent), '*Agriculture, Forestry, Horticulture and Veterinary Medicine*' 138 (4.4 percent), '*Computer Science*, *Electrical and System Engineering*' 93 (3.0 percent), *'Construction Engineering and Architecture'* 31 (1.0 percent), *'Materials Science and Engineering'* 24 (0.8 percent), '*Mathematics*' 24 (0.8 percent), '*Thermal Engineering/ Process Engineering'* 15 (0.5 percent), '*Mechanical and Industrial Engineering'* 7 (0.2 percent).

Data repositories languages

Data repositories have been developed in 60 languages worldwide. Majority of research data repositories are developed in English (1801), followed by Dutch (148), French (145), Spanish (65), Chinese (40), Russian (33), Japanese (29), Portuguese (20), Italian (17), Swedish (13), Arabic (11), Polish (10), Hungarian (9), Català (9), Finnish (8), Norwegian (8), Korean (8), Greek (7), Indonesian (7), Estonia (6), Lithuanian (6), Hindi (6) Slovak (5), Slovenian (5), Bokmål (4), Romanian (4), Thai (4), Ukrainian (4), Latvian (4), Vietnamese (3), Persian (3), Bulgarian (2), Cymraeg (2), Euskara (2), Irish (2), Hebrew (2), Croatian (2), Icelandic (2), Maltese (2) and Serbian (2). Besides those, one research data repository each is developed in the following languages: Yoruba, Tagalog, Tamil, Telugu, Albanian, Romansh, Norwegian Nynorsk, Malayalam, Marathi, Macedonian, Lao, Kalaallisut, Kannada, Galician, Gujarati, Bengali and Afrikaans. Number of data repositories covering more than one language exceeds 1997.

DISCUSSION AND CONCLUSION

Out of 1997 data repositories studied it is found that out of these data repositories 1509 (75.6 percent) are 'disciplinary', 'institutional' 398 (19.9 percent) and 'others' 90 (4.5 percent).

Furthermore, it was identified that World Data System (WDS) certification followed by Core Trust Seal, Deutsche initiative für netzwerk information E.V (DINI) certificate is being most in these open RDRs. It was found that total 70 countries in the world have 1997 RDRs. United States is the leading country with 704 (35.3 percent), followed by Germany 201 (10.1 percent). Study also found that scientific and statistical data formats are available in maximum (1238) data repositories, followed by standard office documents (1108). The study also ascertained that majority of data repositories (1867) do not use any author identification in research data management. Surprisingly, only 181 repositories are using the standard certificates, out of 1997 in which maximum use the World Data System (WDS) (51). And, 1867 Open RDRs do not use any author identification in research data management. Moreover, study revealed that majority of data repositories, 886 (29.1 percent) mentioned 'other' types of license being used by open RDRs. Thus, it is apparent that majority of developers of these open RDRs do not apply standard license. Besides this, 'Copyright' is being used by 785 (25.8 percent), followed by 'CC0' 703 (23.1 percent). Eccentrically, 1049 (52.5) open RDRs are not following any metadata standard. However, Dublin Core and ISO 19115 are the most popular metadata standard being used in these data repositories. It is also found that 311 open RDRs are not using any software. And, 963 (48.2 percent) open RDRs are using unknown and in-house developed software, followed by DataVerse 71 (3.6 percent). As far as subject coverage by these RDRs is concern 'Biology' 715 (23.0 percent), followed by 'Geosciences' (including Geography) 627 (20.1 percent). Libraries have been providing useful information to users about research data management through their websites which is relatively easy and a good starting point.[22] Cordial relationship between the librarian and administrators in other departments in an organization is crucial to develop a successful data repository.[27,28] Researchers use the minimum required approach in metadata entry while uploading data to data repository.[29]

Study is useful to understand the quantitative analysis of research data repositories. Data available on the registry seems unorganized and complicated to analysis and may have given error while presenting in this study. Librarians, researchers, data specialist and administrators ought to work together to transform the data management practices within the research community.[30]

ACKNOWLEDGEMENT

The article was a part of project prepared for partial fulfillment of Refresher Course held on May 31, to June 14, 2019 at CP-DHE, University of Delhi, India.

CONFLICT OF INTEREST

The author declare no conflict of interests.

REFERENCES

1. Borgman CL. The conundrum of sharing research data. Journal of the American Society for Information Science and Technology. 2012;63(6):1059-78.
2. Berman F, Cerf V. Who will pay for public access to research data?. Science. 2013;341(6146):616-7.
3. Gómez ND, Méndez E, Hernández-Pérez T. Social sciences and humanities research data and metadata: A perspective from thematic data repositories. El Profesional de la Información. 2016;25(4):545-55.
4. Nature. Recommended Data Repositorie. Scientific Data. 2019. accessible at: https://www.nature.com/sdata/policies/repositories#general
5. Nature. Data Citation Needed. Scientific Data. 2019;6(1):26-7. doi:10.1038/s41597-019-0026-5.
6. MacMillan D. Data sharing and discovery: What librarians need to know?. The Journal of Academic Librarianship. 2014;40(5):541-9.
7. Smith A, Balazinska M, Baru C, Gomelsky M, McLennan M, Rose L, *et al.* Biology and data-intensive scientific discovery in the beginning of the 21st century. Omics: a Journal of Integrative Biology. 2011;15(4):209-12.
8. Vickers AJ. Making raw data more widely available. BMJ. 2011;(342):d2323. http://dx.doi.org/10.1136/bmj.d2323
9. DeSouza PA. Open data could have helped us learn from another mining dam disaster. Scientific data. 2019;6(1):54.
10. Cheek FM, Bradigan PS. Academic health sciences library research support. Journal of the Medical Library Association: JMLA. 2010;98(2):167.
11. Steinhart G, Saylor J, Albert P, Alpi K, Baxter P, Brown E, *et al.* Digital research data curation: Overview of issues, current activities and opportunities for the Cornell University Library. 2008
12. Kuipers T, DerHoeven JV. Insight into digital preservation of research output in Europe. Survey Report. 2009. Accessible at: http://www. parse-insight, eu/downloads/PARSE-Insight D3-4 SurveyReport_final hq. pdf
13. Aydinoglu AU, Dogan G, Taskin Z. Research data management in Turkey: Perceptions and practices. Library Hi Tech. 2017;35(2):271-89.
14. Horizon 2020. Guidelines on data management in Horizon 2020: version 1.0. 2013. available at: http:// www.gsrt.gr/EOX/files/h2020-hi-oa-data-mgt_en.pdf
15. National Science Foundation. Data management for NSF SBE directorate proposals and awards. 2010. available at: http://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf
16. OECD. OECD principles and guidelines for access to research data from public funding. 2007. available at: ww.oecd.org/science/scitech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm
17. Mayernik MS. Research data and metadata curation as institutional issues. Journal of the Association for Information Science and Technology. 2016;67(4):973-93.
18. Henderson ME, Knott TL. Starting a research data management program based in a university library. Medical Reference Services Quarterly. 2015;34(1):47-59.
19. Perrier L, Blondal E, Ayala AP, Dearborn D, Kenny T, Lightfoot D, *et al.* Research data management in academic institutions: A scoping review. PloS One. 2017;12(5):e0178261.
20. Faniel IM, Connaway LS. Librarians' Perspectives on the Factors Influencing Research Data Management Programs. College and Research Libraries. 2018;79(1):100.
21. Hruby GW, McKiernan J, Bakken S, Weng C. A centralized research data repository enhances retrospective outcomes research capacity: A case report. Journal of the American Medical Informatics Association. 2013;20(3):563-7.
22. Yoon A, Schultz T. Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries' Websites. College and Research Libraries. 2017;78(7):920-33.
23. Erway R, Rinehart A. If you build it, will they fund? making research data management sustainable. OCLC Research. 2016. Available at: at http://www.oclc.org/content/dam/research/publications/2016/oclcresearch-making-research-data-management- sustainable-2016.pdf
24. Farnel S, Shiri A. Metadata for research data: current practices and trends. In International Conference on Dublin Core and Metadata Applications. 2014;74-82. available at: https://www.academia.edu/19654623/Metadata_for_Research_Data
25. OECD. Business Models for Sustainable Research Data Repositories. OECD Science, Technology and Industry Policy Papers. 2017;80:47 doi:10.1787/23074957.
26. E-skills UK. Big Data Analytics: Adoption and Employment Trends, 2012-2017. 2013. Available at: http://www.voced.edu.au/content/ngv:59261 (June 7, 2019)
27. Pinfield S, Cox AM, Smith J. Research data management and libraries: relationships, activities, drivers and influences. PLoS One. 2014;9(12):e114734.
28. Bhardwaj Raj Kumar. Content Analysis of Indian Research Data Repositories: Prospects and Possibilities DESIDOC Journal of Library and Information Tech-

nology. 2019;39(6):280-89.
29. Piwowar HA, Chapman WW. Public sharing of research datasets: a pilot study of associations. Journal of Informetrics. 2010;4(2):148-56.
30. Bhardwaj Raj Kumar. Research Data Management in Higher Educational Institutions. Desidoc Journal of Library and Information Technology. 2019; 39 (6): 269-270

Further Readings

31. Akers KG, Doty J. Disciplinary differences in faculty research data management practices and perspectives. International Journal of Digital Curation. 2013;8(2):5-26.
32. Ball A. How to License Research Data. DCC How-to Guides. Edinburgh: Digital Curation Centre. 2014. Available online: http://www.dcc.ac.uk/resources/how-guides
33. Clements A, McCutcheon V. Research data meets research information management: Two case studies using (a) Pure CERIF-CRIS and (b) EPrints repository platform with CERIF extensions. Procedia Computer Science. 2014;33:199-206.
34. Corti L, DenEynden VV , Bishop L, Woollard M. Managing and sharing research data: A guide to good practice. Sage. 2014.
35. Cox AM, Pinfield S. Research data management and libraries: Current activities and future priorities. Journal of Librarianship and Information Science. 2014 ;46(4):299-316.
36. Davidson J, Jones S, Molloy L, Kejser UB. Emerging good practice in managing research data and research information within UK Universities. Procedia Computer Science. 2014;33:215-22.
37. Gorgolewski K, Margulies DS, Milham MP. Making data sharing count: A publication-based solution. Frontiers in Neuroscience. 2013;7:9.
38. Hanson B, Sugden A, Alberts B. Making data maximally available. Science. 2011;33(6018):649.
39. Holdren JP. Memorandum for the heads of executive departments and agencies: Increasing access to the results of federally funded scientific research. (2013), Executive Office of the President, Office of Science and Technology Policy, Washington, DC. 2013. available at: www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
40. Liu X, Ding N. Research data management in universities of central China: Practices at Wuhan University Library. The Electronic Library. 2016;34(5):808-22.
41. Dillo I, Treloar A, Lusoli W, Kupiainen I, Asch M, Diepenbroek M, *et al*. Business models for sustainable research data repositories. OECD Science, Technology, and Innovation Policy Papers. 2017.
42. Murad MH, *et al*. New evidence pyramid. BMJ Evidence-Based Medicine. 2016;21(4):125-7.
43. OECD "Making Open Science a Reality", Science, Technology and Industry Policy Papers, No. 25, 2015 OECD Publishing, Paris. 2015. http://dx.doi.org/10.1787/5jrs2f963zs1-en.
44. Pinnick J. Exploring digital preservation requirements: A case study from the National Geoscience Data Centre (NGDC). Records Management Journal. 2017;27(2):175-91.
45. Rice R, Southall J. The data librarian's handbook. London, Facet publishing; 2016
46. *Roche DG, et al*. Troubleshooting public data archiving: suggestions to increase participation. PLoS Biology. 2014;12(1):e1001779.
47. Si L, Xing W, Zhuang X, Hua X, Zhou L. Investigation and analysis of research data services in university libraries. The Electronic Library. 2015;33(3):417-49.
48. Smith A, Balazinska M, Baru C, Gomelsky M, McLennan M, Rose L, *et al*. Biology and data-intensive scientific discovery in the beginning of the 21st century. Omics: A Journal of Integrative Biology. 2011;15(4):209-12.
49. Tammaro AM, Casarosa V. Research data management in the curriculum: An `interdisciplinary approach. Procedia Computer Science. 2014;38:138-42.