ISSN (Print): 0972-2467
ISSN (Online): 0976-2477
SRELS Journal of Information Management, Vol 55(3), June 2018, p. 147-152 DOI: 10.17821/srels/ 2018/v55i3/116820

Indian Data Repositories in re3data: A Study

Tanmay Mondal*

Scientific Information Resource Division, Indira Gandhi Centre for Atomic Research, Kalpakkam - 603102, India; tanmaymondal555@gmail.com

Abstract

A registry is a source of information. Registry of Research Data Repositories (re3data.org) is a global data registry for more than 1500 data repositories available in different academic disciplines worldwide. Different data repositories from India are listed in re3data. Majority of these data repositories are in scientific discipline. Most of the data repositories don't follow any metadata standards for datasets. Only few provide permanent data links.

Keywords: Data Repository, Indian Data Repository, re3data

1. Introduction

Amount of data is growing day by day. Data once generated can be used for various purposes. Data include statistics, results of experiments, measurements, observations, survey results, interview recordings and many more. There are data repositories in different disciplines. But knowing all these data archives is not easy. There exist a few data registries which provide information about various data repositories. It helps end users to search many data repositories in one platform. Registry of Research Data Repositories (re3data.org) is one of the largest and most comprehensive registries for data portals. It has already identified more than 1,500 research data repositories with detailed information. Information (How to find an appropriate research data repository, 2013) like general information (e.g., short description about data repository, content types, keywords), responsibilities (e.g., institutions responsible for funding, content or technical issues), policies (e.g., policies of data repository, restriction), legal aspects (e.g., licenses of the database and datasets), technical standards (e.g. APIs, versioning of datasets, software of the repository), quality standards (e.g., certificates, audit processes) is easily available on re3data.org. Different access policies about various data repositories is available at Registry of Research Data Repositories (LSE Impact Blog, 2013). This paper discusses various facets about Indian data repository listed in re3data.

2. Registry of Research Data Repositories (About-re3data.org)

Registry of Research Data Repositories acts as a discovery tool. The idea behind Registry of Research Data Repositories (re3data) is to make an exhaustive web-based registry for data repositories available in almost all disciplines. It promotes a culture of sharing, increased access and better visibility of research data. Information regarding any data repositories' policies and other contextual information are in one platform. It helps researchers to identify different data repositories which can be used for accessing various types of data. It can also be used by funding organizations to store and provide access to data generated from their funding. Publishers and academic institutions also can recommend authors, scientists for data archiving purposes.

Re3data.org is funded by the German Research Foundation (DFG). Project partners of re3data are the Berlin School of Library and Information Science (http://www.ibi.huberlin.de) at the Humboldt Universitätzu Berlin, the Library and Information Services (LIS) department (http://bib.gfzpotsdam.de) of the GFZ German Research Centre for Geosciences, the KIT Library (http://www.bibliothek.kit.edu) at the Karlsruhe Institute of Technology (KIT) and the Libraries (https://www.lib.purdue.edu) of the Purdue University. In December 2012, re3data launched an alpha version. The web address of re3data is www.re3data.org.

*Author for correspondence

3. Criteria for Data Repository to be Included in re3data.org

Rücknagel, *et al.* (2015) state that any repository registered in re3data.org must fulfill the following:

- Have focus on research data;
- Be operated by a legal entity with an organizational framework that provides sustainability (e.g. library, university);
- Clarify access conditions to the repository and research data; and
- Provide terms of use.

Entries in the registry are gathered in a sequential workflow backed by the experience of an editorial team. Any user or institute can suggest new repositories for inclusion in the list. The editorial team will check the repositories on the minimum requirements set by the re3data.org. A repository is indexed when the minimum requirements of the re3data.org's registration policy is met. Before a new record is published in re3data.org, all gathered information is reviewed by a second editor.

4. Method

The data is collected only from re3data.org as it is the largest and most comprehensive registry of data repositories available on the web (Re3data.org).

4.1 Data Collection

Various search facilities are supported by re3data.org. Keyword search is the default option. Other filters like subject, content, countries, etc. are also important to search and explore all possible information about a repository. Browsing facility by subject, content type and country is available. Browse by 'country' shows that there are 30 data repositories available in re3data.org from India (re3data, 2017). A list of data repositories is listed in Annexure 1.

4.2 Limitation of the Study

The study is based on data available only in the re3data.org database. Other important Indian data repositories available but not listed in re3data are not considered for this study.

4.3 Objectives

The objectives of the study are:

- To analyze different aspects of Indian data repositories,
- To examine various access policies of data repositories, and
- To know various formats of data available in data repositories.

5. Results and Discussion

5.1 Content Type

Most of the repositories are in scientific discipline and many of them contain scientific and statistical content. Twenty repositories contain scientific and statistical data and 18 contain structured graphics. Text, graphics, office documents, images are also preserved in these repositories. Figure 1 shows nature of data preserved by Indian data repositories.

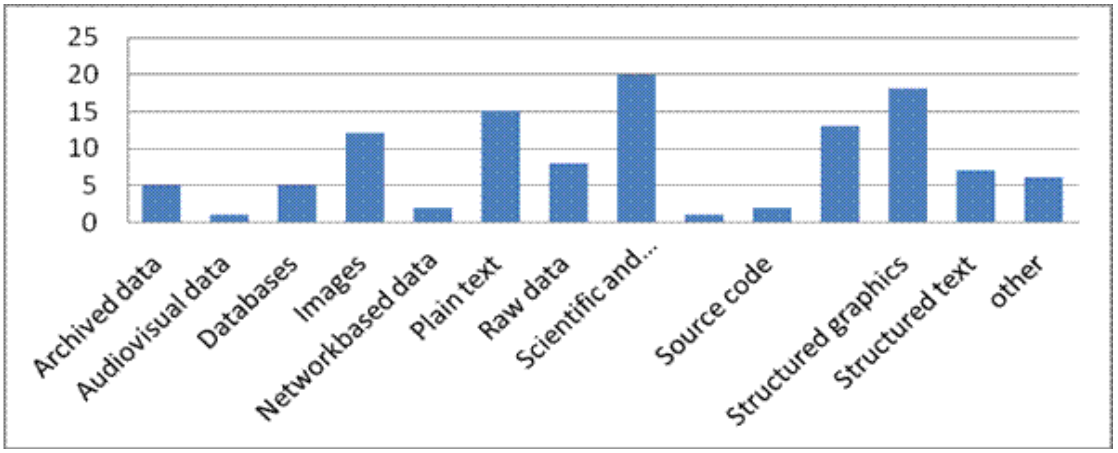

Figure 1. Content type in data repositories.

5.2 Subject Area

Most of the data are in biological disciplines. Subjects are categorized accordingly DFG Classification (Classification of Subject Area, Review Board, Research Area and Scientific Discipline, 2016 - 2019). The broad subject classification in Indian data repositories is as follows:

- Humanities and Social Sciences
  - Social and Behavioural Sciences
- Life Sciences
  - Biology
  - Medicine
  - Agriculture, Forestry, Horticulture and Veterinary Medicine
- Natural Sciences
  - Chemistry
  - Physics
  - Geosciences (including Geography)
- Engineering Sciences

Figure 2 shows various subject categories available in Indian data repositories. 56% data are related to life science whereas 43% data are in natural science domain.

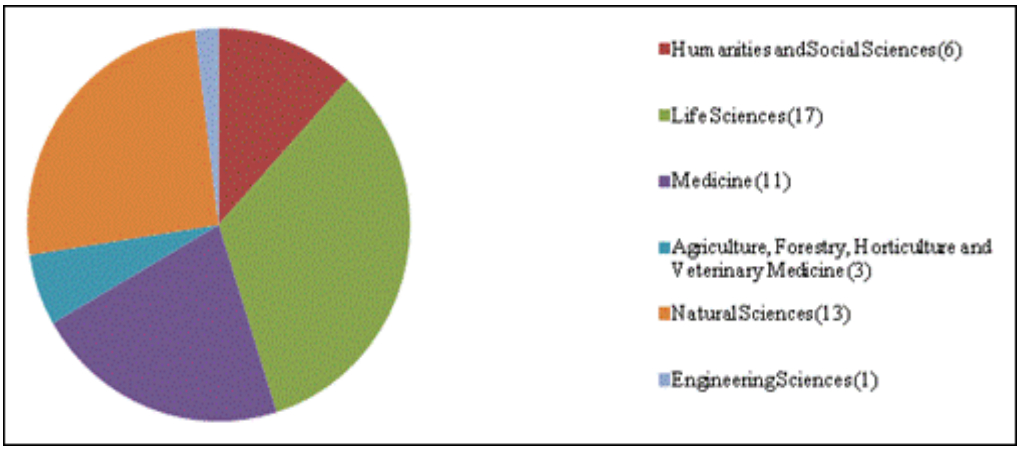

Figure 2. Subject categories of data repositories.

5.3 Collaboration of Countries

Same data repositories are handled by many countries in a collaborative environment. Figure 3 shows that other than Indian contribution (96%), UK, USA, Germany and few other countries are partners for some of these data repositories.

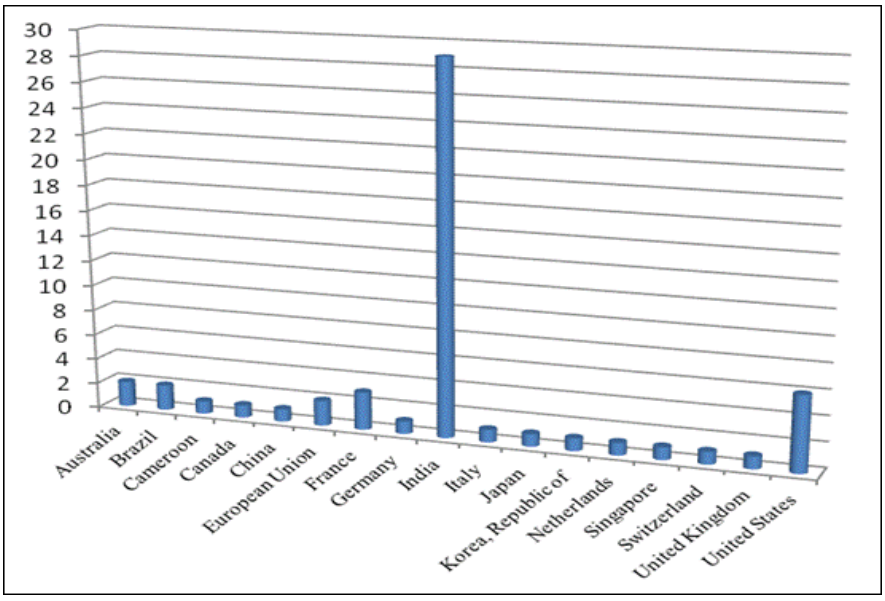

Figure 3. International collaboration of data repositories.

5.4 Data Access

Out of 30 repositories, 80% (24) are providing open access to data. Some of them use all the access policies depending upon the nature of data. Natures of data access are:

- Open (24),
- Restricted (9),
- Closed (2), and
- Embargoed (1).

5.5 Data Licenses

Majority of data archives follow their own data policies. Besides self-copyright license, Creative Commons (CC) is used as it is easy for many repositories for implementation. Public domain data sets are also available in some of these repositories. Following are the available data licenses:

- Copyrights (8),
- Other (11),
- Public Domain (7), and
- CC (5).

5.6 Data Uploaded Method

Most of the data uploading methods are restricted. Data managers collect data from researchers through various medium like email, CD, direct upload using user login etc. Out of 30 repositories, 50% (15) are using closed based for data uploading. Other methods used for data uploading are:

- Closed (15),
- Restricted (14), and
- Open (1).

5.7 Permanent Link

Most of the data repositories (90%) are running without any permanent link for data sets. Only one provides Digital Object Identifier (DOI) for data sets. Other used links are listed below:

- DOI (1),
- HDL (Handle system) (1),
- None (27), and
- Other (1).

5.8 Metadata Standard

Most of the data described in these data repositories are not following any such specific metadata standard. Only a few are following standards. Following are the standards used by Indian data repositories:

- DDI - Data Documentation Initiative (1),
- EML - Ecological Metadata Language (1),
- ISO 19115 (1), and
- MIBBI - Minimum Information for Biological and Biomedical Investigations (1).

5.9 Repository Language

Most of the repositories are providing services only in English language. Hindi is also used for data services besides 'eng'. One is providing service in French language also.

- 'eng' (29),
- 'fra' (1), and
- 'hin' (3).

6. Findings of the Study

Major findings of this study are:

- Majority of Indian data repositories listed in re3data.org are in scientific domain,
- Most of the data repositories follow open access data policy, and
- Majority of these data repositories are not following any metadata standard.

7. Conclusion

Reliable data provide quality information. Implementation of data repository in any organization requires good data management guidelines. Polices related to data access, copyright issues, various standards should be taken care of at first before going for data archive. Description of data requires metadata standards as majority of data archives are not maintaining any such standards. Some of the data archives are not having any such specific data policies. Persistent identifiers are also not available for datasets in many data archives. Data managers must list their data repositories in various registries for its wide publicity.

8. Reference

How to find an appropriate research data repository (2013). Available at: http://blogs.lse.ac.uk/impactofsocialsciences/2013/11/29/how-to-find-an-appropriate-research-data-repository/

About-re3data.org. Available at: http://www.re3data.org/about.

Rücknagel, J, *et al.* (2015). Metadata Schema for the Description of Research Data Repositories: version 3.0. DOI: http://doi.org/10.2312/re3.008.

Re3data.org. Available at: www.re3data.org.

Browse by country (2017). Accessed on: 6th June, 2017. Available at: http://www.re3data.org/search?query=&countries[]=IND ().

Classification of Subject Area, Review Board, Research Area and Scientific Discipline (2016 - 2019). Available at: http://www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/fk-wahl2015/2015_fachsystematik_2016_2019_en.pdf.

Annexure 1. List of Indian data repositories in re3data.org (Accessed on 6th June, 2017)

No.	Name	Institution(s)	Data access	Persistent identifier system(s)	Repository software
1.	ACEpepDB: Peptide Database	CSIR Central Food Technological Research Institute	Open	None	Unknown
2.	Chickpea Transcriptome Database	National Institute of Plant Genome Research	Open	None	Unknown
3.	Clinical Trials Registry - India	CMR National Institute of Medical Statistics (International)	Open	Other	Unknown
4.	Database on Indian Economy	Reserve Bank of India	Open, restricted, registration	None	Unknown
5.	District Information System for Education	National University of Educational Planning and Administration, Department of Educational Management Information System	Open	None	Unknown
6.	Experimental Tropical Watersheds	Indian Institute of Science (International)	Open	DOI	Unknown
7.	Export Import Data Bank	Government of India, Ministry of Commerce and Industry, Directorate General of Commercial Intelligence and Statistics	Open	None	Unknown
8..	HIstome	Indian Institute of Science Education and Research, Tata Memorial Centre, Advanced Centre for Treatment, Research and Education in Cancer	Open	None	Unknown
9.	Human Protein Reference Database	Institute of Bioinformatics (International)	Restricted, other	None	Unknown
10.	Human Proteinpedia	Institute of Bioinformatics (International)	Open, closed, restricted, other	None	Unknown
11.	ICRISAT Dataverse Network	Dataverse Network Project (International)	Open, registered, closed, restricted	HDL	DataVerse

12.	IMEx	Molecular connections (Collaborative)	Open	None	Unknown
13.	India Biodiversity Portal	French Institute of Pondicherry (Collaborative among Indian institutes)	Open	None	Unknown
14.	India Energy Portal	The Energy and Resources Institute, National Knowledge Commission	Open	None	Unknown
15.	India Environment Portal	National Knowledge Commission, Centre for Science and Environment	Restricted, other	None	Unknown
16.	India Water Portal	National Knowledge Commission, The Arghyam Foundation	Open	None	Unknown
17.	Indian Genetic Disease Database	Council of Scientific and Industrial Research, Indian Institute of Chemical Biology	Open	None	Unknown
18.	Indian Space Science Data Center	Indian Space Research Organization (ISRO) Telemetry, Tracking and Command Network, Indian Deep Space Network	Restricted, registration	None	Unknown
19.	Integrated Ocean Discovery Program	Integrated Ocean Discovery Program,	Open, restricted, registration	None	Unknown
20.	Marine Microbial Database of India	CSIR National Institute of Oceanography	Open	None	Unknown
21.	MolTable	National Chemical Laboratory, Digital Information Resource Center	Open	None	Unknown
22.	North East Resources Databank	North Eastern Development Finance Corporation Ltd.	Restricted, registration	None	Unknown
23.	Ocean Data and Information System	Indian National Centre for Ocean Information Services, Ministry of Earth Sciences	Open	None	Unknown
24.	Open Government Data Platform India	National Informatics Centre (International)	Open	None	Drupal
25.	Oral Cancer Gene Database	Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre	Open	None	Unknown
26.	TBNet India	Institute of Bioinformatics, National JALMA Institute of Leprosy, Ministry of Science and Technology, Department of Biotechnology	Open	None	Unknown
27.	The Human Protein Atlas	Lab Surgpath (International)	Open, embargoed	None	Unknown
28.	TropFlux	National Institute of Oceanography, Indian National Centre for Ocean Information Services (International)	Open	None	Unknown
29.	World Data Centre for Geomagnetism	Indian Institute of Geomagnetism, Department of Science and Technology	Restricted, registration	None	Unknown
30.	WorldClim - Global Climate Data	Ashoka Trust for Research in Ecology and the Environment (International)	Open	None	Unknown