

## Open Access to Research Data: Sharing Data for Efficient Research

Rishi Tiwari\*, B.B. Chand\*\* Ramesha\*\*

### Abstract

The research is becoming more data intensive today not only because data intensive research has more acceptances today but also data driven research facilitate quality research and show pathways for new fields of research. Research data can be generated for different purposes and through different processes i.e., field survey, laboratory experimentation, may be captured through sophisticated technologies and can be divided into different categories. Each category may require a different type of data management plan. The advantages of sharing research data are many to the individual researchers and to the research community as a whole. The library can play a vital role in facilitating research data service by closely associating with the researchers in planning, managing, preserving and sharing the research data. By creating awareness among the researchers and providing necessary training efficient research data service can be provided to the research community.

**Keywords:** Research data service; Open data; Data curation; Data sharing.

### Introduction

Research cannot flourish if data are not preserved and made accessible (Nature, 2009). One of the next big things for the research libraries are going to be the research data management service. Large scale data is being generated now during the research process and it is expensive and time consuming. Considering this now funding agency are making it mandatory for a data management plan along with the open access to scholarly communication. Research organizations need to provide not only structure and policies for research data preservation, but services to support and educate researchers on concepts of data management and promote the sharing

of data sets that can often be vital for the continuation of research.

Data is one of the basic inputs for any research and at the end of the research it is one of the important content of the output as well. Huge investment is required to gather, process and bring the data to a usable form. The recent development in information infrastructure, investment on data and the open access movement has compelled the researchers, the funders to rack in more return on investment from the research data. It is only when the research data managed properly and shared openly among the research community the true value of the research data will be realised. The sharing of research data and reusing the same data for further research has been supported by both the science and social science research community. The research funding agencies make sure that the data generated from the high value research project be made public and further research should be allowed to carry on with the same datasets.

The research is becoming more data intensive today not only because data intensive research has more acceptance today but also data driven research facilitate quality

---

**Author's Affiliation:** \*Librarian, Birla Institute of Management Technology (BIMTECH), Plot No. 5, Knowledge Park-II, Greater Noida (NCR), U.P. - 201306, \*\*Deputy Librarian, Indian Socio Economic Change (ISEC), Nagarabhavi, Bangalore, Karnataka, \*\*Associate Professor, Dept. LIS, Bangalore University, Bangalore-560056, Karnataka, India.

**Reprint's Request:** Dr. Rishi Tiwari, Librarian, Birla Institute of Management Technology (BIMTECH), Plot No. 5, Knowledge Park-II, Greater Noida (NCR), U.P. - 201306, India.

E-mail: rishi.tiwari@bimtech.ac.in

research and show pathways for new fields of research. The growth of data in the “big sciences” such as astronomy and physics has led to new models of science—collectively known as the “Fourth Paradigm” and to the emergence of new fields of study such as astroinformatics, computational biology, and digital humanities (Borgman, 2009; Hey, T., Tansley & Tolle, 2009).

### *Research Data*

“Data are facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors.” (National Research Council. 1999). A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.” (Consultative Committee for Space Data Systems. 2002). “Any information that can be stored in digital form, including text, numbers, images, video or movies, audio, software, algorithms, equations, animations, models, simulations, etc. Such data may be generated by various means including observation, computation, or experiment.” (National Science Board. 2005)

The term “data” as used in this document

is meant to be broadly inclusive. In addition to digital manifestations of literature (including text, sound, still images, moving images, models, games, or simulations), it refers as well to forms of data and databases that generally require the assistance of computational machinery and software in order to be useful, such as various types of laboratory data including spectrographic, genomic sequencing, and electron microscopy data; observational data, such as remote sensing, geospatial, and socioeconomic data; and other forms of data either generated or compiled, by humans or machines. (Borgman, 2012)

Research data can be generated for different purposes and through different processes i.e field survey, laboratory experimentation, may be captured through sophisticated technologies and can be divided into different categories. Each category may require a different type of data management plan.

### *Types of Research Data*

Research data may include all anything and everything related to the project. From the initial project note to the final report is part of the research data. Data recorded from the field in questionnaire to organized data as SPSS tables are research data.

**Table 1: Types of Research Data**

Data Type	Characteristics	Example
<b>Observational</b>	data captured in real-time, usually irreplaceable	sensor data, survey data, sample data, neurological images
<b>Experimental</b>	data from lab equipment, often reproducible, but can be expensive	gene sequences, chromatograms, toroid magnetic field data
<b>Simulation</b>	data generated from test models where model and metadata are more important than output data.	climate models, economic models
<b>Derived or compiled</b>	data is reproducible but expensive	text and data mining, compiled database, 3D models
<b>Reference or canonical</b>	a (static or organic) conglomeration or collection of smaller (peer-reviewed) datasets, most probably published and curated	gene sequence databanks, chemical structures, or spatial data portals.

**Table 2: Major Research Funding Organization Data Policy Requirements**

	Economic and Social Research Council, UK	Welcome Trust, UK	National Science Foundation, USA	National Institute of Health, USA
Data policy launch year	2000, revised 2010	2007	2011	2003
Are researchers required to offer to deposit data in a central dedicated data centre (where available)?	Y	Y	-	Y
Are researchers required to deposit data locally and make it available to others (where data centres are not applicable / available)?	-	Y	Y	Y
Are researchers required to include a 'data sharing plan' when applying for funding?	Y	Y	Y	Y
Does the policy state that research applications can include requests for funding for data sharing activities?	Y	Y	Y	Y
Does the policy state that proposed data sharing activities will be considered within final assessments / funding decisions for projects?	Y	Y	Y	Y
Does the policy state that secondary users of data are expected to acknowledge their sources?	Y	Y	-	Y

### *Importance of Sharing Research Data*

Sharing research data has long been discussed and recorded for the advancement of the research itself. Many publishers like Nature Publishing Group and American Naturalist has inbuilt the policy guidelines associated with sharing of research data. If we see many articles now has a data component attached to the article. Research publications with associated data sets will bring in more transparency to the research itself. The expanded use of research data not only provides substantial benefits to the research community but also to the society as a whole with increase in the quality and efficiency of research and innovation.

❖ *Scientific Proof:* Publishing research data and along with research papers can allow others to re-use, replicate, validate, or correct your results, thereby improving the science and scientific process.

- ❖ *Research Impact:* When the data are made available for researchers across the globe, those work done by the data collectors and sharers, will be the one first to referred as to know what is the motif behind the research data and what all research has already been on that data set. This will enhance the citation data for the original data sharer.
- ❖ *Promotion of Use of Empirical Data for Policy Decisions:* As because more data is available it compels the decision makers to opt for policy decisions backed by research supported by empirical data. It provides authority to the decisions makers to choose the best policy options available with them.
- ❖ *Improvement of Measurement and Data Collection Methods:* When the data is available for wider access, it is extensively

used for research. If the data requires any improvement with respect to methods of data collection it can be improved in the future rounds.

- ❖ *Data Preservation*: Preserving research data for future use is an important aspect of data curation and sharing. Largely when the data is shared it is well organised and well documented. Centralized preservation will facilitate best information infrastructure to the individual researchers.
- ❖ *New Research Ideas*: Availability of data ignites researchers mind. New research ideas are generated looking the same data from different angle. Data sharing promotes new research ideas among the researchers.
- ❖ *Making Publicly Funded Research Available Publicly*: Major funding agencies throughout the world now wish to see a data management plan along with the research proposals. Access to data has direct impact on the research output in any subject. Considering the real and visible edge the sponsoring agency are focusing more or data sharing as well.

#### *Role of Library in Research Data Service*

The role of library is crucial in organization and preservation of data. The data could be primary or secondary, it may be in the form of microdata or it may be tabulated data. Tools need to be developed to present these data in an well organized manner that it can be accessed and used by the researchers at shortest possible time. Utmost care needs to be taken to check the completeness of these data sets in terms of the data itself and the documents those are associated with these data sets.

Research data services are services that a library offers to researchers in relation to

managing data and can include informational services (e.g., consulting with faculty, staff, or students on data management plans or metadata standards; providing reference support for finding and citing data sets; or providing web guides and finding aids for data or data sets), as well as technical services (e.g., providing technical support for data repositories, preparing data sets for a repository, deaccessioning or deselecting data sets from a repository, or creating metadata for data sets) (Tenopir, Birch, and Allard 2012, 7).

Libraries are looking to assist with all stages of research, by offering guidance and tools for collecting, exploring, visualizing, labelling and sharing data (Monastersky, 2013). The data management services also include additional add on services like data visualization and analytics where the researcher can use the data on the web itself and make their own analysis.

#### *Research Data Management Life-cycle*

The research data management life-cycle can be divided in to 8 stages (Figure -1). The librarian's can play a key role in many stages towards the research data management life-cycle as described below:

**Figure -1 Research Data Management Life-Cycle**



- ❖ *Planning for Research Data:* The planning process for research data starts while preparing the concept note for the research funding itself. The library can facilitate the research data that is already available in the proposed area with the detailed description. Major research output from these available data will facilitate quick and easy literature review and will help the research know the gaps that they can concentrate on.
- ❖ *Managing Research Data:* Managing research data require skills related to data organization and clear understanding of data. While managing data library need to work closely with the research team so as to gather as much details available about the data especially the field manual, coding manual and data filtration process. This will help enrich the data with necessary details for clear understanding of the curated data.
- ❖ *Preserving Research Data:* This is where the library is going to play a major role so far as data management process is concerned. Many libraries are already involved closely with the researcher for open access to scholarly communication. The deep knowledge of metadata librarian can enrich the datasets with value additions which can facilitate easy discovery of desired data. Clear understanding of digital depositories will help establish stable data repositories.
- ❖ *Sharing Research Data:* The research data thus preserved need to be shared with the research community so as to realise the real research potential. Software tools needs to be developed and implemented to facilitate the data discovery. Here the rights need to be managed with respect to specific dataset.
- ❖ Apart from the roles as a part of the data management life-cycle process library needs to play several leading roles in many of the following areas:
  - ❖ *Awareness:* The researchers need to be making aware about the clear benefits of sharing the research data and how the library can extend the necessary support towards achieving the common goal. As a part of the information literacy program the awareness service can be extended to the new users about the services offered by the library.
  - ❖ *Training:* Initial training is very important to increase the comfort level of the researcher so as to organise and deposit the data to the repository. Training will help them keep the research data management in mind while collecting, processing and analyzing the data.
  - ❖ *Data Curation Policy:* Library can take the lead in creating institution wide data curation policy which is important for sustainability and will service as a guide for the research community.

#### *Skills Development for Librarian for Research Data Service*

Newton, Miller, and Bracke, in their investigation of the role of librarian in institutional data set collecting, found strong evidence that although research libraries—through their connections with faculty across campus and their expertise in developing traditional collections—are prime candidates for developing scientific data collections for universities, additional skills are required to populate an institutional repository with relevant data. In particular, libraries need to make use of professional relationships and collaborations with faculty across fields and between institutions to identify materials.



Creamer *et al* found that of twenty needed data competency areas, the greatest need for librarians was technical hands-on training in the digital description and curation of large data sets. Research Library UK (RLUK 2012) point out the following skill which are essential for research data service.

- ❖ Ability to advise on preserving research outputs
- ❖ Knowledge to advise on data management and curation, including ingest, discovery, access, dissemination, preservation, and portability
- ❖ Knowledge to support researchers in complying with the various mandates of funders, including open access requirements
- ❖ Knowledge to advise on potential data manipulation tools used in the discipline/subject

## Conclusion

Research is largely data driven and the research community will always look for the new data in their respective research areas. The funding agencies are also making sure that the data generated from the public funds should be available openly to the researchers. Access to open data will allow the researcher to work on new areas without wasting much time on data collection. It will also improve the return on investment in the research studies. Research data service is one of the important services that are going to be integral part of the future library services. With exposure to metadata creation and digital repositories librarians are technically sound on managing the research data. Re-skilling is required in the area of data manipulation and visualization.

## References

1. Borgman CL, Bowker GC, Finholt TA & Wallis JC. Towards a Virtual Organization for Data Cyberinfrastructure. Joint Conference on Digital Libraries, Austin, TX, ACM. 2009. <http://dl.acm.org/citation.cfm?id=1555459> (accessed on 16/01/2014).
2. Borgman Christine L. The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*. 2012; 63(6): 1059-1078.
3. Committee on Science, Engineering, and Public Policy (U.S.). Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age (2009). Ensuring the integrity, accessibility, and stewardship of research data in the digital age. Washington, DC: National Academy Press, [http://www.nap.edu/catalog.php?record\\_id=12615](http://www.nap.edu/catalog.php?record_id=12615) (accessed on 16/01/2014).
4. Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS). Washington, DC: National Aeronautics and Space Administration; 2002, 1-9. <http://public.ccsds.org/publications/archive/650x0b1.pdf> [accessed on 16/01/2014]
5. Creamer, Andrew *et al*. An Assessment of Needed Competencies to Promote the Data Curation and Management Librarianship of Health Sciences and Science and Technology Librarians in New England. *Journal of Science Librarianship*. 2012; 1(1). doi:10.7191/jeslib.2012.1006.
6. Fienberg, Stephen E II Martin, Margaret E III Straf, Miron L. Sharing Research Data. Washington, DC: National Academy Press; 1985. <http://www.nap.edu/catalog/2033.html> (accessed on 16/01/2014).
7. Hey T, Tansley S & Tolle K (Eds.). The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, WA: Microsoft; 2009. <http://research.microsoft.com/en-us/collaboration/fourthparadigm/> (accessed on 16/01/2014)
8. <http://www.ala.org/acrl/sites/ala.org/acrl/>

- files/content/publications/whitepapers/Tenopir\_Birch\_Allard.pdf [accessed on 15/01/2014]
9. Jones S *et al*. The Data Audit Framework: a first step in the data management challenge. *International Journal of Digital Curation*. 2008; 3(2); 112-120.
  10. Mark P Newton, CC Miller, and Marianne Stowell Bracke. Librarian Roles in Institutional Repository Data Set Collecting: Outcomes of a Research Library Task Force. *Collection Managemet*. 2011; 36(1): 53-67. doi:10.1080/01462679.2011.530546.
  11. National Research Council. A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases. Washington, DC: National Academy Press; 1999, 15. [http://www.nap.edu/catalog.php?record\\_id=9692](http://www.nap.edu/catalog.php?record_id=9692) [accessed on 16/01/2014]
  12. National Science Board. Long-Lived Data Collections: Enabling Research and Education in the 21st Century. Arlington, VA: National Science Foundation; 2005, 13. [www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf](http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf) [accessed on 16/01/2014]
  13. Nature. Data's shameful neglect. *Nature*. 2009; 461(7261): 145. <http://www.nature.com/nature/journal/v461/n7261/full/461145a.html> [accessed on 16/01/2014]
  14. Research Information Network. Researchers' use of academic libraries and their services. *RIN*. 2007. <http://www.rin.ac.uk/files/libraries-report-2007.pdf> [accessed on 16/01/2014 ]
  15. Research Information Network. Stewardship of digital research data: a framework of principles and guidelines. *RIN*. 2008: 28 <http://www.rin.ac.uk/files/Research%20Data%20Principles%20and%20Guidelines%20full%20version%20-%20final.pdf> [accessed on 16/01/2014]
  16. Research Information Network. Data centres: their use, value and impact. *RIN*. 2011. [http://www.rin.ac.uk/system/files/attachments/Data\\_Centres\\_Report.pdf](http://www.rin.ac.uk/system/files/attachments/Data_Centres_Report.pdf) [accessed on 15/01/2014]
  17. Research Libraries UK. Re-skilling for Research. 2012. <http://www.rluk.ac.uk/content/re-skilling-research> [accessed on 16/01/2014]
  18. Richard Monastersky. The library reboot. *Nature*. 2013; 495(7442): 431-432. [http://www.nature.com/polopoly\\_fs/1.12664!/menu/main/topColumns/topLeftColumn/pdf/495430a.pdf](http://www.nature.com/polopoly_fs/1.12664!/menu/main/topColumns/topLeftColumn/pdf/495430a.pdf) [accessed on 16/01/2014]
  19. Shearer Kathleen. Research Data: Unseen Opportunities . Canadian Association of Research Libraries; 2009.
  20. Tenopir C, Birch B, & Allard S. Academic Libraries and Research Data Services. Association of College and Research Libraries; 2012. [www.carl-abrc.ca/about/working\\_groups/pdf/data\\_mgt\\_toolkit.pdf](http://www.carl-abrc.ca/about/working_groups/pdf/data_mgt_toolkit.pdf) [accessed on 16/01/2014]
  21. Tenopir Carol, Sandusky Robert J, Allard, Suzie, Birch Ben. Academic librarians and research data services: preparation and attitudes. *IFLA Journal*. 2013; 39(1): 70-78, 9p. DOI: 10.1177/0340035212473089