

DST-CENTRE FOR POLICY RESEARCH



INDIAN INSTITUTE OF SCIENCE

**Policy Brief**

Published June 2022

# Open Research Data for India: Exploring Infrastructure and Governance



**Moumita Koley**

STI Post-Doctoral Policy Fellow, DST Centre for Policy Research, Indian Institute of Science, Bangalore, 560012.



**Suchiradipta Bhattacharjee**

STI Senior Policy Fellow, DST Centre for Policy Research, Indian Institute of Technology, Delhi, 110016.



**Perminder Jit Kaur**

STI Senior Policy Fellow, DST Centre for Policy Research, Indian Institute of Science, Bangalore, 560012.



**Famida Khan**

STI Young Policy Fellow, DST Centre for Policy Research, Indian Institute of Science, Bangalore, 560012.

## Summary

There is a growing trend of Open Research Data (ORD) sharing worldwide. International agencies like the OECD, UNESCO, etc., actively promote ORD practices. ORD is essential for new knowledge creation, reproducibility, transparency, or reuse and can generate economic benefits. The adoption of open research data practices offers a significant promise of progress for the Indian STI ecosystem. The considerable challenges are the absence of focused ORD policies, cognitions of data management plans by institutions, inadequate infrastructure, lack of governance, and legal frameworks for research data sharing. Transforming scientific culture, incentivizing, and investing in capacity building are essential and primary requirements to encourage research data sharing.

## Introduction

Moving towards a knowledge-based economy is crucial for low to middle-income countries (LMICs) to avoid being trapped in the mid-income loop. Access to knowledge and data can play an enabling role in uplifting the economic status of these countries. As per the news published in *The Economist*, data is even more valuable than the oil reserves of any country. The database of any country can be fed to machine learning software to generate solutions in numerous fields ranging from defense, medicines, gene therapy, electronics, and the environment. Easy availability of data helps in analyzing past trends to predict the future and can become a pillar for the socio-economic growth of any nation<sup>1</sup>.

Among various types of data, research data (input and output data from research

activities) is an important category. Research data is the primary evidence of many scientific discoveries and is ubiquitous in the present scientific ecosystem. Presently, scholarly articles are the most revered means to represent a comprehensive set of results. However, access to such articles and associated data files are locked behind the publisher's paywall. Especially in Indian contexts, this aspect can not be ignored where only a few institutions and universities can pay for access to scholarly literature. The adoption of data sharing culture can play a crucial role in the ease of doing research. It can increase transparency, trustworthiness, reproducibility, and optimum usage of resources. While there are initiatives and conversations actively addressing how to increase access to both, the issues around data sharing are more nuanced.

Some countries have adopted open data frameworks and mechanisms to upload data periodically in a public repository that is easily accessible to all citizens. Guidelines for implementing open research data sharing are available for the European Union and can be extrapolated and replicated in India and other Low Middle-Income Countries (LMICs). Adopting an open research data (ORD) sharing policy, the establishment of infrastructure and the governance mechanism can guide the research data sharing and stimulate data usage from public-funded research. Thus, the objective of the present policy brief is to illustrate the current status of open research data (ORD) policies at the international level, their need, and recommended policy actions in the Indian context.

## Current Open Research Data Practices

The Universal Declaration of Human Rights (1948) asserts access to cultural, art, and scientific progress as the fundamental human rights<sup>2</sup>. Ensuring access to data

1. Kiran Bhageshpur, Forbes Technology Council, 2019; <https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/?sh=2b8b28637304>, on 20.12.2021

2. Report of United Nations on Universal Declaration of Human Rights, 1948; <https://www.un.org/en/universaldeclaration-human-rights/>

Ensuring access to data generated from research not just upholds our fundamental rights but shared access to research data can help broaden the researchers' view and explore "additional research questions"<sup>3</sup>. Cross-disciplinary collaborations and access to data across medical and socio-economic disciplines promptly have proven crucial to the quest for solutions of the Covid-19 pandemic. However, for such partnerships to continue beyond the emergencies requires global adoption of open science practices, of which open sharing of research data is an essential element. Data and codes can provide a comprehensive view of any research by validating the replicability of outcomes under different geographical conditions and providing further directions in research. The utilitarian points of ORD are already an established phenomenon, and there are global initiatives to make access to research data a common practice across several disciplines.

Some notable examples of subject-specific ORD sharing are Genbank, Worldwide Protein Data Bank (wwPDB), UniProt in the life sciences, Space Physics Data Facility and Set of Identifications, Measurements, and Bibliography for Astronomical Data (SIMBAD) in the space sciences, etc<sup>4</sup>. Apart from the subject-specific data repositories, discipline-neutral research data repositories such as ZENODO from CERN, Figshare, Dryad<sup>5,6,7</sup>, etc., are also used for open data sharing.

## Current status of research data sharing principles and policies

The past couple of years have seen notable efforts by the international community to promote the culture of ORD sharing. Declaration on Access to Research Data from Public Funding (2004) has defined guidelines for open access for thirty OECD countries<sup>6</sup>. The UNESCO Open Science Policy Recommendations draft (2018) also advocates for ORD, especially in public-funded research<sup>7</sup>. The European Open Science Cloud (EOSC) is an environment for hosting and processing research data to support EU science<sup>8</sup>. One of the successful examples of the open data policy is Open Government Data Policies and Practices in the Republic of Korea. The Korean Strategy to Promote Sharing and Use of Research Data for Innovative Growth is an excellent example of an approach toward maximizing the benefits from research data<sup>9</sup>.

Different universities have also come up with additional research data-sharing practices. Universities like Harvard<sup>10</sup>, and the University College of London<sup>11</sup>, provide free access to data. Funding agencies like NSF, NIH, Research UK, Research England, etc., advocate for more openness in publicly funded research so that research work can impact public health and wellbeing<sup>12</sup>. Similarly, as per directives of the European Union, all research funded by European

3. National Institute of Health, Genomic Data Sharing Policy, available at <https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing-faqs/>

4. European Open Science Cloud (EOSC); <https://eosc-portal.eu/about/eosc>

5. <https://zenodo.org>

6. <https://figshare.com>

7. <https://github.com/CDL-Dryad/dryad>

8. Report of OECD Legal Instruments on Declaration on Access to Research Data from Public Funding, 2022; <https://legalinstruments.oecd.org/public/doc/157/157.en.pdf>

9. Report of UNESCO on Open Science, 2019; <https://en.unesco.org/science-sustainable-future/open-science/recommendation>

10. European Open Science Cloud; <https://eosc-portal.eu/>

11. Report of US Government on Public access to Federally funded research in the United States, 2019;

[https://community.oecd.org/servlet/JiveServlet/downloadBody/149047-102-1-](https://community.oecd.org/servlet/JiveServlet/downloadBody/149047-102-1-263336/USA%20ANNEX%20to%20ENHANCED%20ACCESS%20TO%20PUBLICLY%20FUNDED%20DATA%20FOR%20SCIENCE_Final4sec)

[263336/USA%20ANNEX%20to%20ENHANCED%20ACCESS%20TO%20PUBLICLY%20FUNDED%20DATA%20FOR%20SCIENCE\\_Final4sec](https://community.oecd.org/servlet/JiveServlet/downloadBody/149047-102-1-263336/USA%20ANNEX%20to%20ENHANCED%20ACCESS%20TO%20PUBLICLY%20FUNDED%20DATA%20FOR%20SCIENCE_Final4sec)

12. 1. Open access at UCL, <https://www.ucl.ac.uk/research/integrity/policies-and-guidelines/data-management-protection/open-access>

research bodies must be published via open access and be available to the whole European community<sup>13</sup>.

## Need for a comprehensive open data policy in India

In India, novel scientific knowledge generated from research is currently transformed chiefly into journal articles, reports, and grey literature. Findability, interoperability, and reuse of the data from the research outputs is not an easy task, and hence the data is hardly ever available, discussed, or reused. The data formats in the public domain are generally not machine-readable to be utilized in AI, machine learning-driven discoveries. They are also not compatible with the FAIR (Findable, Accessible, Interoperable, and Reusable) principle<sup>14</sup>, globally the most accepted data sharing guidelines. Moreover, the negative results aren't published as the journals focus on publishing only positive outcomes. Such practices amount to a massive data loss.

So far, there is no constructive approach to research data sharing generated using public funds in India. The National Data Sharing and Accessibility Policy (NDSAP, 2012) is directed mainly towards governance data, not a targeted policy for open research data<sup>15</sup>. The draft Biological Data Storage, Access, and Sharing Policy of India formulated by the Department of Biotechnology (DBT) is a good example of promoting research data sharing<sup>16</sup>. Similar to biological research data, other data domains should be accessible to the public. Therefore it becomes crucial to have an Open Research

Data (ORD) policy to drive and guide research data sharing in India. Like many countries, an open research data portal for the scientific community for storing, sharing, and reusing scientific data and results, supported by high-capacity cloud solutions with computing capacity, should also be considered in India at the national level. A well-defined research data portal federated by the government can help in the country's data-driven science and overall innovation landscape. Researchers, innovators, companies, and the general public can publish, find and reuse research data and tools for research, innovation, and educational purposes under trusted conditions. A data-centric system can further boost innovative governance initiatives by promoting decisions driven by relevant data.

With free access to data, overlapping data generated from different arms of academic and research institutions can be easily checked. Using suitable implementation tools, free data availability will allow like-minded individuals and research institutes working on the same theme to collaborate and share resources. The complex and interconnected global economic, environmental and social challenges need urgent interventions from science, technology, and innovation (STI). The success of addressing such overarching issues requires transdisciplinary research,

combined knowledge from different domains, and access to research data from various disciplines. Global research initiatives in these directions can be possible through collaborative research and opening up research data through accessible data

13. UK Research and Innovation (UKRI) open access Policy; <https://www.open.ac.uk/library-research-support/open-access-publishing/rcuk-open-access-policy>

14. European Community open access Policy; <https://ec.europa.eu/research/openscience/index.cfm?pg=openaccess>

15. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., The FAIR Guiding Principles for scientific data management and stewardship. In *Scientific Data*, 2016, 3, 160018.

16. National Data Sharing and Accessibility Policy-2012 (NDSAP-2012), <https://data.gov.in/sites/default/files/NDSAP.pdf>

portals. The successful implementation of such an Open Data Portal will require trained human resources, infrastructure, and governance. To avoid legal complications associated with websites like Sci-hub, guidelines have to be provided in Open Data Sharing Policies. Various factors like data standardization, the technological requirement for managing and sharing data, ethical regulations, roles of various institutes, and detailed policy-level guidelines needs to be addressed. Thus, a sustainable data portal design must address the challenges of financing, governance, architecture, and operational models.

## Recommended policy actions in India

### *Research data repository*

Establishing a research-data repository in each field of research connected to academic and R&D institutions can help with the efficient management of data. The advancements in digital technologies, various software for data processing, and cloud computing have opened up new dimensions in reusing research data and driving discoveries.

### *Governance and management*

ORD governance will be a critical part of the ORD policy implementation. Making data openly available cannot be sufficient unless protocols are maintained for its quality check. Systematic data storage, management, and funds require an efficient governance mechanism. Government can standardize the format and file types to upload and keep data by following

guidelines given in different standards. For instance, ISO 15489 standard defines the method of data generation and record keeping. Similar to the United States, where the office of Management and Budget manages the open data policy<sup>17</sup>, there is a need to assign the responsibility to any suitable organization to implement the policy by all research institutes. Also, creating an organization and task force can help establish an open data policy for the country.

### *Legal framework*

A legal basis for quality, frugality, sharing, and use of research data, especially for data of sensitive nature, and adoption of data-management plans and data ownerships should be established. For instance, in the United States, the Information Quality Act asks data publishers to issue their guidelines about quality, utility, and ownership of information<sup>18</sup>.

### *Ethics and clinical data guidelines*

A critical component of research in the pharmaceutical and medicinal field is the requirement of clinical trials. Some medical investigations involve the personal information of the subjects involved. Establishing guidelines about data access for the research involving Human Subjects, informing the Subject about the likely outcome of this activity, and ethical considerations needs can assist in Open Data Access in these sensitive areas.

### *Incentivizing research data sharing*

Researchers have no incentives to take extra

17. Report of Department of Biotechnology on Biological data storage, access and sharing policy of India, 2019; [https://www.nhp.gov.in/NHPfiles/Draft1-Biological\\_Data\\_Policy.pdf](https://www.nhp.gov.in/NHPfiles/Draft1-Biological_Data_Policy.pdf)

18. Report of Office of Management And Budget on Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies, The\_White\_House, available at: [https://obamawhitehouse.archives.gov/omb/fedreg\\_final\\_information\\_quality\\_guidelines/](https://obamawhitehouse.archives.gov/omb/fedreg_final_information_quality_guidelines/)

initiatives to share all the data, code, etc., from a project or data repositories. Rather, fears of scooping, data use without acknowledgment and attribution, losing out to competing research groups, etc., act as deterrents. Webinars and conferences about making data open access can make researchers aware of its benefits and encourage them to volunteer for the process. The government may give credit for data sharing in the government repository weightage equivalent to the journal publications and presentations for job application and promotion. A mechanism to incentivize public-funded research data open can encourage research data sharing while addressing concerns.

## Conclusion

Science is incremental and grows on the shoulders of the already existing knowledge and data. Access to data can accelerate the growth of science and discoveries. The data

generated from research is valuable and should not be locked under the paywalls of commercial journals, especially when the research is publicly funded. The adoption of data sharing culture can increase the transparency and trustworthiness of any published results. Open research data practices can help make public-funded data available to the whole ecosystem of scientists, businesses, entrepreneurs, and citizens. This will assist in educating and training the researchers to enhance their data skills and expertise. The policy will be a stepping stone towards establishing rules and regulations for availability, maintenance of data, adoption of data-management plans, and data ownership. Many innovation and commercialization activities inspired by shared research data could be undertaken in the future. Using suitable policy formulation, implementation, and maintenance procedures, open research data policy can strengthen the collaboration among sciences and social sciences ecosystems.

DST-CENTRE FOR POLICY RESEARCH



**Contact information:**

e-mail: [coordinator.cpr@iisc.ac.in](mailto:coordinator.cpr@iisc.ac.in)

Address: DST Centre for Policy Research  
Indian Institute of Science (IISc),  
Bengaluru 560012, India



[dstcpriisc.org](http://dstcpriisc.org)



[@DSTCPRIISc](https://twitter.com/DSTCPRIISc)



[@DSTCPRIISc](https://www.facebook.com/DSTCPRIISc)



[tinyurl.com/inDSTCPRIISc](https://www.linkedin.com/company/tinyurl.com/inDSTCPRIISc)



[tinyurl.com/YTDSTCPRIISc](https://www.youtube.com/channel/UCYTDSTCPRIISc)