Knowledge Flow Network: Can it bridge the gaps for the Open Science ecosystem?

Sudeepa Nandi[#], Ishita Goyal[#], Vineet Sharma[#]

[#,*]Contributed equally

Introduction

The world is driven by the power of scientific knowledge and moving towards Knowledge-based economies. Empowering the knowledge creation process is even more critical for Lower Middle Income countries (LMICs) to avoid the middle-income trap and address increasing income disparities. However, the mode of action to push knowledge-based economies will differ from country to country. India being a youthful country, its strength lies in human capital and this resource should be harnessed for scientific innovation. They should be trained in specialised skills to create, disseminate, and apply knowledge for growth. Seamless access to knowledge is imperative in building human capital and hence, the country needs an open science environment. Open Science is a way to promote universal and equitable access to knowledge and quality scientific data. Access to scientific data is a vital enabler of the knowledge economy and promotes reproducible and trustworthy science, which is essential for progress. At this point, the country needs robust and coordinated government policies to access research data.

The significant loopholes in the current state of data sharing and access in India lie in the scattered nature of the available resources, the lack of structured policies, and the stakeholders' lack of awareness regarding the existing policies. Currently, there is no dedicated policy in India to promote open access to research data. Moreover, there are inconsistencies between different policies and executive initiatives like the National Data Sharing Accessibility Policy (NDSAP, 2012), Shodhganga, National Digital Library of India, Biological Data Storage, Access and Sharing Policy of India, and National Knowledge Network (NKN). All these policies and initiatives are seeking to address different aspects of open access to Indian research. But in reality, these have led to redundancy and confusion among the researchers. Hence, it is essential to integrate these resources and platforms and have similar data-sharing practices. Enhanced access to

Importance of a Research Data Sharing Policy for Stakeholders

- A researcher may benefit from the specific information that should be known ahead of time to comply.
- Institutes/ Universities may benefit by creating & facilitating the process for researchers and generating quality output from their Institutes.
- The public may know exactly what research data can be reused by them and how?
- Funding/ Managing bodies may be able to advance the process of developing national open-access policies and infrastructures for access to research data.

* The authors participated in the 'Participatory Science Advice for Policy Making' workshop as part of the Grassroots Science Advice Promotion Awards 2021 by the INGSA Asia to Dr Jenice Jean Goveas and Dr Moumita Koley.

data, information and knowledge would immensely help not only researchers but the whole society. At this point, a comprehensive open research data-sharing policy — dedicated to creating a collaborative and accessible research ecosystem in India is a prerequisite.

The draft Science, Technology, and Innovation Policy (STIP) 2020 intends to create a streamlined research ecosystem for which it envisions a central repository of all the data generated in the STI ecosystem. The idea of a central, accessible research data platform in India is not new, and we can find it spread across Indian policies and laws like NDSAP and STIP draft 2020.

The collaborative and open data sharing of COVID-19-related research is a recent example of how the cooperation and participation of a large pool of researchers helped to generate giant strides in science and technology. Various other projects like Computer vision data, Human genome project, neurological diseases, etc. also have revealed how open access to data and the underlying cooperative spirit can enrich research.

Indian researchers annually produce a huge amount of data. Much of the data holding academic and research value simply languishes with labs and researchers due to lack of accessibility. This necessitates a common platform for research data that is accessible, collaborative and scalable for nurturing more robust evidence-based publicly-funded research ecosystems.

The following elements are essential strategies to consider while formulating an open research data sharing policy and knowledge flow network.

Specifics of the Policy	
Features	**Scope**
Research data	Research data covered by the policy may include data from institutional researchers, independent researchers, NGOs, analysts, metadata and raw data. Bio-medical data may be anonymised and de-identified to avoid ethical concerns.
Exceptions to research data	Type of data may be stored privately and should not be made publicly available. Provisions must be there to classify such sensitive data such as health records, data related to defense & security, etc.
Open access to research data	Open access involves various activities aiming to remove the barriers to sharing research data. Research data should be as open as possible, human and machine-readable, without violating the privacy & confidentiality of private and sensitive data.
Infrastructure for data storing	Infrastructure for data storing must state the position on the use of data storing platforms. Data storing platforms are the means for sharing data with the community. Discipline-specific platforms with tags for interdisciplinary or multidisciplinary approaches are preferred to general repositories whenever available. Data storing platforms may include Digital data repositories, Physical data warehouses, metadata repositories, Data-lakes and Data-marts. National repository networks connecting localised/ institutional platforms provide research data and services.
Incentives for researchers	Innovative incentives that distinguish open practices over the existing system. The current system is biased towards the number of publications in prestigious journals, which does not inspire researchers to provide early and open access to their research data.
Data sharing/ Management plan	Open services and infrastructure should use standards and best practices to ensure international interoperability in repositories so that the data can be assessed, discovered and reused by humans and machines. Provisions must be added to spruce the data periodically to eliminate redundant and sensitive data. Investments are needed, not only to build a system but to maintain it.

Mandates of data sharing	Precise instructions about specific data sets with a graded plan on using the data with indexing and searchability on the web are essential. Information such as *Private sharing* of Sensitive data and Personal data with anonymity and informed consent where Data Use Agreements may be considered; *Suggestive sharing* for Independent Researchers, NGOs, Analysts; *Mandatory sharing* for Institutional researchers. Interoperable standards and best practices must be enforced at the Institute and National levels as per the international standards. A researcher has a right to use their data first-hand and share the extent of their data at their will, but with a rational time frame of restrictions.
Process of data authentication	Statement on the need for data authentication, where a peer review of data at the Institutional level is anticipated, along with the expectations of peer reviewers in their assessment of data files. A step of data authentication may help build trust in the data produced and help build a reputation for the Institute to produce relevant research.
Awareness programs	Research data sharing remains a new concept, and common questions can be answered by Programs and Initiatives among researchers and citizens to make them aware about the role of open access to research data, infrastructure, mandates and the sharing process.

Recommendations

The policies should focus on FAIR (Findable, Accessible, Interoperable, and Reusable) guidelines for data sharing and ethical and inspiring usage of accessible data to sustainably continue the research data flow system. The policies should facilitate building tools, infrastructures, and services for open data access, sharing and supporting legal and governance structures to effectively cater to all the stakeholders. Only then can the purpose of open research data sharing and access achieve its full potential. Following are some recommendations for creating a knowledge flow network, amalgamating structures based on the needs of all those involved in the research ecosystem, including researchers, authors, editors, referees, funders, citizens, etc.:

1. To better understand **the concerns and obstacles of the stakeholders, a survey** might be a good starting point in any country, as was done by Bezuidenhout et al. (2016) in Africa[1] or by Springer Nature and figshare in collaboration with Digital Science.[2]
2. ***Interconnected Data and Regulatory Framework***: Currently, in India, some domain-specific open data repositories and digital platforms, and institutional library collections are already available. However, they are neither well-organised nor used widely. There is also a lack of associated structured policies for their usage.[3] As a first step to building a solid foundation for a successful exchange of research knowledge and ideas, a thoroughly-strategised network should be **sculpted by interlinking the repositories and other scientific platforms** (as shown in Fig. 1).
3. ***Research Institutional Libraries with Data storage facilities***:

   3.1 To avoid overloading the scientists/researchers and remove the infrastructural burdens, the institutional libraries can have Data Management Bodies (DMBs), who would be the field experts on data sharing architecture. It will also help to **bridge the digital gap for scientists**.

   3.2 Institutions can make raw data and metadata submission to the DMBs mandatory post-publication of an article or completion of a thesis project. Researchers should **sign data availability statements** after publishing a work/thesis, including the reason for not sharing anything if there are such occurrences (depending on the nature of data, viz. sensitive/private/legal or any other specific reason).

   3.3 Although Institutional DMBs will have full access to raw data and metadata, they may apply a **case-specific low-barrier mechanism** before sharing and making them openly accessible hereafter. The filters can be based on the negative list defined by a consensus between the researchers, organisations, and the government (on privacy, national security, and Intellectual Property Rights (IPR)).

   3.4 ***Discipline-specific data sharing standards*** should be well-defined nationwide: Since research data from diverse fields have different formats and time-dependent importance, discipline-specific

standards should be followed. For example, in the case of clinical research data, necessary deidentification of candidates/patients may be set as a norm, or research data of varying subjects and natures may have different embargo times before being openly shared.

3.5 Since these libraries will be the first point of submission of research data and metadata, appropriate **copyright licenses/ persistent and unique identifiers** (viz. DOI, Handle) for citations should be provided as done in Harvard Dataverse, Zenodo - CERN and the University of Michigan Library.

3.6 In the long run, there could be further modifications such as tracking the user and usage of data and metadata.

4. ***Discipline-specific repositories***: Following the submission to local Institutional libraries, shareable data should be additionally linked to discipline-specific repositories based on subject-specific defined standards. **Funding agencies can make this a mandate.** Likewise, data and metadata from independent researchers, NGOs and citizen science should be directly deposited here. A registry/archive portal (like re3data or Indian Science and Technology Archive of Research (INDSTA; as suggested in draft STIP 2020)) can function as an aggregator for all these data from different subject-based repositories to be easily accessible to everyone. Independent researchers or common masses can send out their queries for additional data through this portal.
5. As a long-term goal, a ***Central directory***, with bulletin boards and popular science articles, science news, etc., appended to the relevant research articles, data, and metadata from the National Archive, will be hugely beneficial. It would create an interdisciplinary platform fostering more collaborative work. Such Central directories can be the Global Knowledge Exchange Point (Fig. 1).
6. This system of knowledge flow from institutions to the Central directory could be **dynamically monitored, for gap recognition and necessary interventions**, by a National STI Observatory (to be established as a recommendation in the draft STIP 2020). Hence, there should be flexibility in the up-gradation of policies based upon the experience of multiple institutions and repositories. It should also include methods/ protocols for data management and security plans to keep up with the new formats of data storage, etc., and also be able to quickly train DMBs. There could be help from private organisations for these purposes and joining forces for continuous maintenance.
7. ***Acquiring and allocating resources to build and maintain the framework*:** For constructing and maintaining an elaborate and well-weaved network, policies should also focus on the economic factors and seek external investments. Fortunately, as mentioned above, many domain-specific repositories and platforms (e.g., India Biodiversity Portal, Clinical Trials Registry, India Environment Portal, ICAR Research Data Repository) already exist in our country, which if clubbed together as parts of this network will reduce the cost significantly. Besides, private organisations will also benefit from the openly shared government funded data. Hence, there should be policies where the government can support their Corporate Social Responsibility funds in the open science framework building and maintenance.
8. ***Financially and technologically support storage of ever-expanding data***: Additionally, the system would need a mechanism to safeguard data in the physical format and the repositories.
9. ***Easier information extraction***: It would require creating digital platforms non-technical, interactive (searchable and indexable), and inclusive so that they are easily usable and understandable by people outside the research niche. It would promote citizen science and multidisciplinary projects with greater involvement of all stakeholders, even if they are not field experts.
10. ***Educating stakeholders about the available research knowledge flow system***: Awareness regarding the whole knowledge flow cycle has to be increased amongst the stakeholders. Active policies should support familiarising researchers and citizens with the two-way routes (participatory and institutional approaches) in the research knowledge flow cycle (Fig. 1).
11. ***Rewards and recognition systems to nurture openness***. If researchers' value, employability and funding are still judged based on the number of publications in high-impact journals, then scientists will remain hesitant in sharing data openly due to the fear of their data being misused without getting the due credit. Hence, it is crucial to revisit the researchers' evaluation policies- putting more weight on the

open sharing of data and collaborations[4] (as recommended by DORA and also adopted by the European Research Council). Simultaneously, all National Journals could have alternative scooping protection plans for scooped projects due to open data sharing (as currently with PLOS Journals Complementary Research Policy). This is especially important for developing countries that are adopting open data sharing policies, yet lacking in research infrastructure compared to developed nations, whose research progression speed is faster.

12. ***Incentivising citizens*** who ideate new work from the openly accessible data would encourage participation by citizens, further strengthening the top-down knowledge flow framework as a positive feedback loop.

To improve the productivity and dynamicity of the ecosystem, we suggest the following:

1. Building a well-knit knowledge flow network (as in Fig. 1) interconnecting different repositories with research articles/thesis with their data, metadata, codes, and related general articles;
2. Defining specific standards of shareable data in a more domain-specific manner;
3. To incorporate newer policies to raise awareness and openness amongst the stakeholders;
4. To bring in supportive legal structures for open sharing of data to avoid misuse;
5. To join hands with private organisations to maintain different platforms and repositories;
6. Investing in infrastructure, training, education, digital literacy and capacity building;
7. To encourage innovative approaches and methodologies to strengthen the open science research ecosystem.

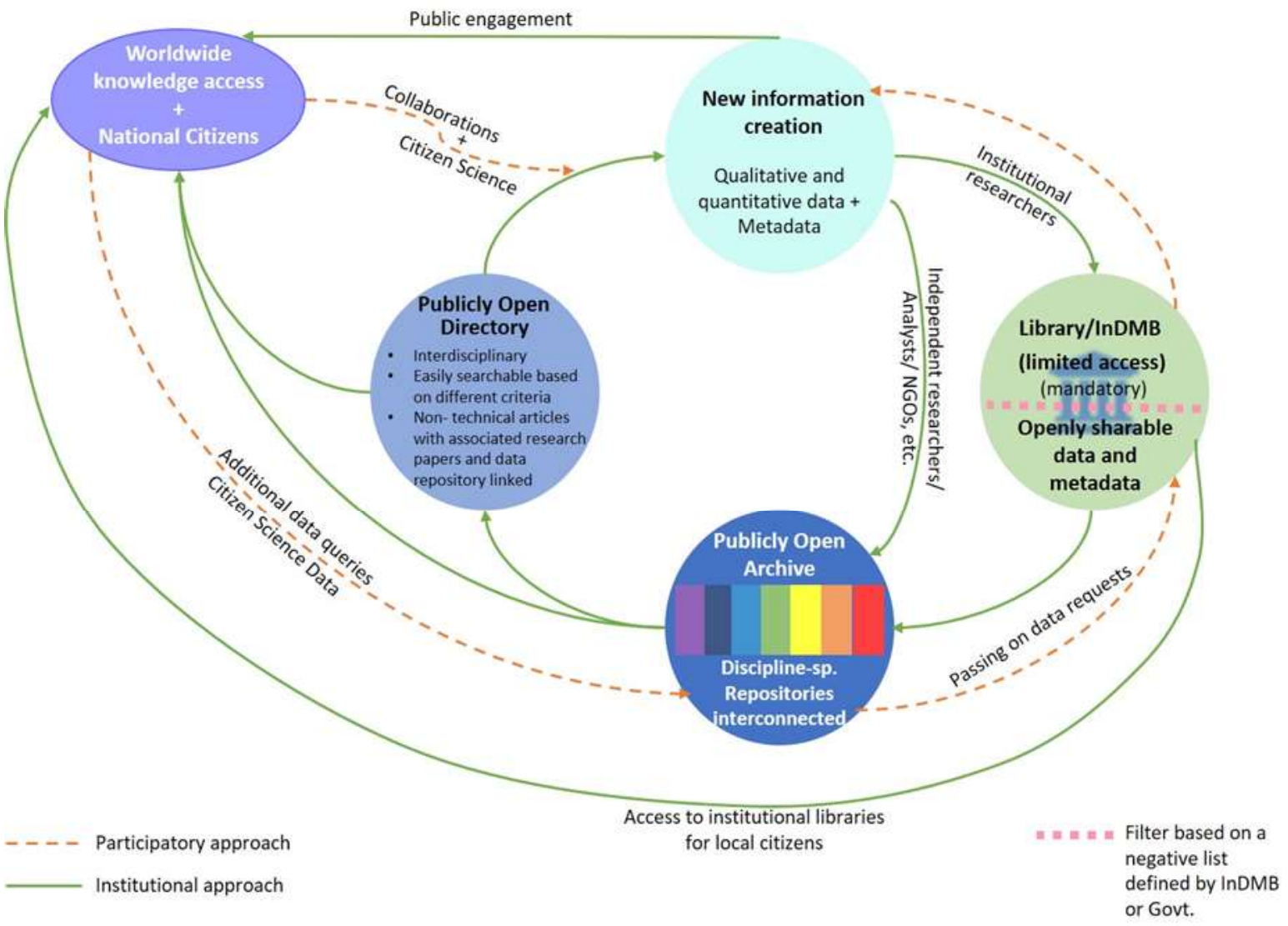


Figure 1. Structural framework for open data sharing of government funded research work

Conclusion

The future success of this open synergistic system will largely depend on its ability to respond to the diverse needs of its stakeholders efficiently. By implementing these changes, we hope to imbibe a culture of research and innovation, with a balance between open science mindset and Intellectual property rights, which is crucial in today's fast-moving world. It would demonstrate the benefits of sharing and accessing open data, as some of these changes have been beneficial to the European Union and other developed countries. Recently, the National Institute for Health and Care Research, United Kingdom, launched an open-access publishing platform to allow researchers to "rapidly" publish any research information, including incremental findings, case reports and even negative findings. The platform, called NIHR Open Research, would also include links to all supporting data, allowing reanalyses, replication and reuse, as well as making it easier for researchers to collaborate. India should also make concerted efforts, as suggested in this policy brief, to develop and boost its open science ecosystem. Since the South Asian countries share similar socioeconomic backgrounds, even though here we have focussed on Indian Open Science policy changes, such a data-sharing network can be expanded and useful for the entire South Asian research ecosystem like the Research data alliance, an intergovernmental data network sharing between the European Union, US and Australia.

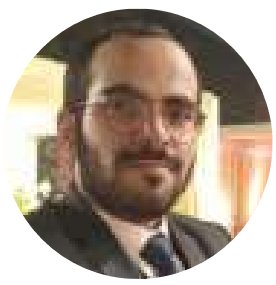
References

1. Bezuidenhout L, Rappert B, Leonelli S, Kelly AH (2016) Beyond the Digital Divide: Sharing Research Data across Developing and Developed Countries. *figshare*. https://doi.org/10.6084/m9.figshare.3203809.v1.
2. Science D, Simons N, Goodey G, Hardeman M, Clare C, et al. (2021)The State of Open Data 2021. Digital Science. Report. https://doi.org/10.6084/m9.figshare.17061347.v1
3. Anilkumar N (2018) Research Data Management in India: A Pilot Study. *EPJ Web of Conferences*, 186: 03002. https://doi.org/10.1051/epjconf/201818603002
4. Bonn NA, Pinxten W (2021) Advancing science or advancing careers? Researchers' opinions on success indicators. *PLoS ONE* 16(2): e0243664. https://doi.org/10.1371/journal.pone.0243664

About the authors



Sudeepa Nandi
Research Scholar, Department of Biological Sciences, Tata Institute of Fundamental Research, Mumbai, India



Vineet Sharma
Research Scholar, Department of Humanities and Social Sciences, IISER Pune, Pune, India



Ishita Goyal
Research Scholar, CSIR - National Institute of Science Communication & Policy Research, KS Krishnan Marg, Delhi, India