



Fostering scholarly infrastructure through PIDs to mobilize the science as an open enterprise: lessons from data-repositories in India

Jiban Krishna Pal 
Indian Statistical Institute

Dorothea Strecker 
Humboldt University of Berlin



Co-funded by
the European Union



Fostering scholarly infrastructure through PIDs to mobilize the science as an open enterprise: lessons from data-repositories in India

Abstract: The doctrine of open science invariably fosters the availability of scientific data for sharing and reuse, which in turn improves the quality of research with greater insights, and leads to better science. Modern data management has set a new paradigm that integrates the FAIR data principles for making data *Findable*, *Accessible*, *Interoperable*, and *Reusable*. As a first instance, this presentation aims to deliver a realistic view of implementing the FAIR guiding principles in the development of the Indian data-repository landscape. So an assessment of the state of FAIRness of Indian data repositories has been conducted by augmenting its' attributes in different dimensions.

The study reveals that most of the data repositories in the country do not comply with the FAIR data principles in terms of metadata generation, persistent identification, and many other means of practice. Almost all the Indian repositories provide access to their digital resources without having persistent identifiers (PIDs), and they rely on traditional practice via OpenURL where the hyperlinks are either changed frequently or their location paths are broken (no longer exist).

Hence this presentation entails how to develop scholarly infrastructure through persistent identifiers (PIDs) with the application of relevant schemas viz. Digital Object Identifiers (DOIs), Persistent Uniform Resource Locators (PURLs), National Bibliography Numbers (NBNs), Archival Resource Keys (ARKs), and CNRI Handles. It also emphasizes the extensive uses of each schema with their purpose and possible strategies for implementing PIDs with the need for organizational support.

Eventually, it demonstrates an approach of FAIRness through global infrastructure by linking digital objects with DataCite services, researchers with ORCID, and organizations with ROR to discover the resources persistently with sustainability. So this presentation (focusing on a prototype investigation) illustrates how PIDs are being practiced toward maintaining the FAIRness of data-repositories in India and mobilizing the science as an open enterprise.



Email: jiban@isical.ac.in

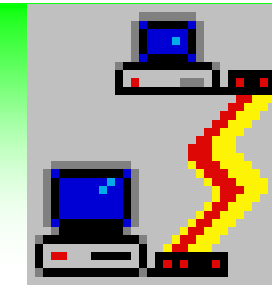
<https://orcid.org/0000-0002-2870-9180>

dorothea.strecker@hu-berlin.de

<https://orcid.org/0000-0002-9754-3807>



The Changing Scenario !



- Approaches to organizing information have been changed over time
 - Catalogue card to Bibliographic record, Lib-OPAC to Info-Gateway, Digital archives and ...?
 - Traditional Cataloguers are gradually designated as Metadata Librarian or Specialist ...?
 - Professionals are now creating metadata for DLs, harvesting metadata for IRs ...?
 - Even responsible to select metadata standards for Res-data repositories ...?
 - The Century has witnessed an unprecedented change world over
 - The concept of ownership has turned into access mechanism ...!
 - Mandates for sharing publicly funded research outputs ...!
 - Ensuring the availability of scientific data for reuse ...!
 - Securing progress of global repository landscape...!
 - Stakeholders are mobilizing the science as an open enterprise
 - Professional communities are constantly enhancing the open-science practices ...*
 - Researchers are progressively making their presence in open repositories ...*
 - Science organizers are grappling with the open scholarly infrastructure ...*
 - Scientific endeavours (researches) are becoming increasingly global ...*
- ❖ *Changes are inevitable, and we have to keep on changing to ensure connectivity among entities in the research ecosystem - **PIDs are here to stay and evolve.***

Persistent Identifiers – *An Obvious Necessity*

A PID is a long-lasting reference to an entity that may refer to researcher, organization, and digital objects including publications, datasets, grants, projects, software, reports, among others.

- To improve visibility & discoverability
- To obtain transparency & FAIRness
- To integrate services and metadata
- To create long-standing references
- To nurture Open Science practices
- To promote global connectivity



<https://orcid.org/0000-0001-6622-4910>



<https://ror.org/01y2jtd41>



<https://doi.org/10.5061/dryad.708gr>



❖ *Varieties of PIDs are:* ROR, ORCID, DOI, PURL, NBN, ARK, CNRI-Hdl, ISBN, ISSN, VIAF, ISNI, URN, XRI, etc.

DataCite DOI Covers a Wide Range of Resources

viz. Journal articles, Conference papers, Book-chapters, Research datasets, Software, Images, and even Gray literature like Theses, Master’s dissertations, Project reports, Preprints, among others.

DataCite Registry

Dataset	16,200,197	☐ Dissertation	120,391
Physical Object	14,098,123	☐ Report	115,543
Text	11,966,471	☐ Project	101,985
Image	4,137,380	☐ Conference Paper	87,841
Other	2,396,391	☐ Book	83,956
Journal Article	1,125,328	☐ Sound	52,004
Preprint	1,115,782	☐ Book Chapter	34,323
Collection	977,478	☐ Event	19,157
Software	451,102	☐ Data Paper	15,632
Audiovisual	346,942	☐ Model	14,638
Interactive Resource	127,466	☐ Workflow	5,931
		☐ Output Management Plan	3,391

ROR-ID for the Indian Statistical Institute



PiD
fest
CONFERENCE
ON PERSISTENT
IDENTIFIERS
2024

 <https://ror.org/00q2w1j53>

Indian Statistical Institute

ORGANIZATION TYPES

Education, Funder

OTHER NAMES

Acronyms

ISI

Labels

भारतीय सांख्यिकीय संस्थान (hi)

LOCATIONS

Kolkata (GeoNames ID 1275004), India

WEBSITE

<http://www.isical.ac.in/>

OTHER IDENTIFIERS

GRID grid.39953.35

ISNI 0000 0001 2157 0617

Crossref Funder ID 501100013357

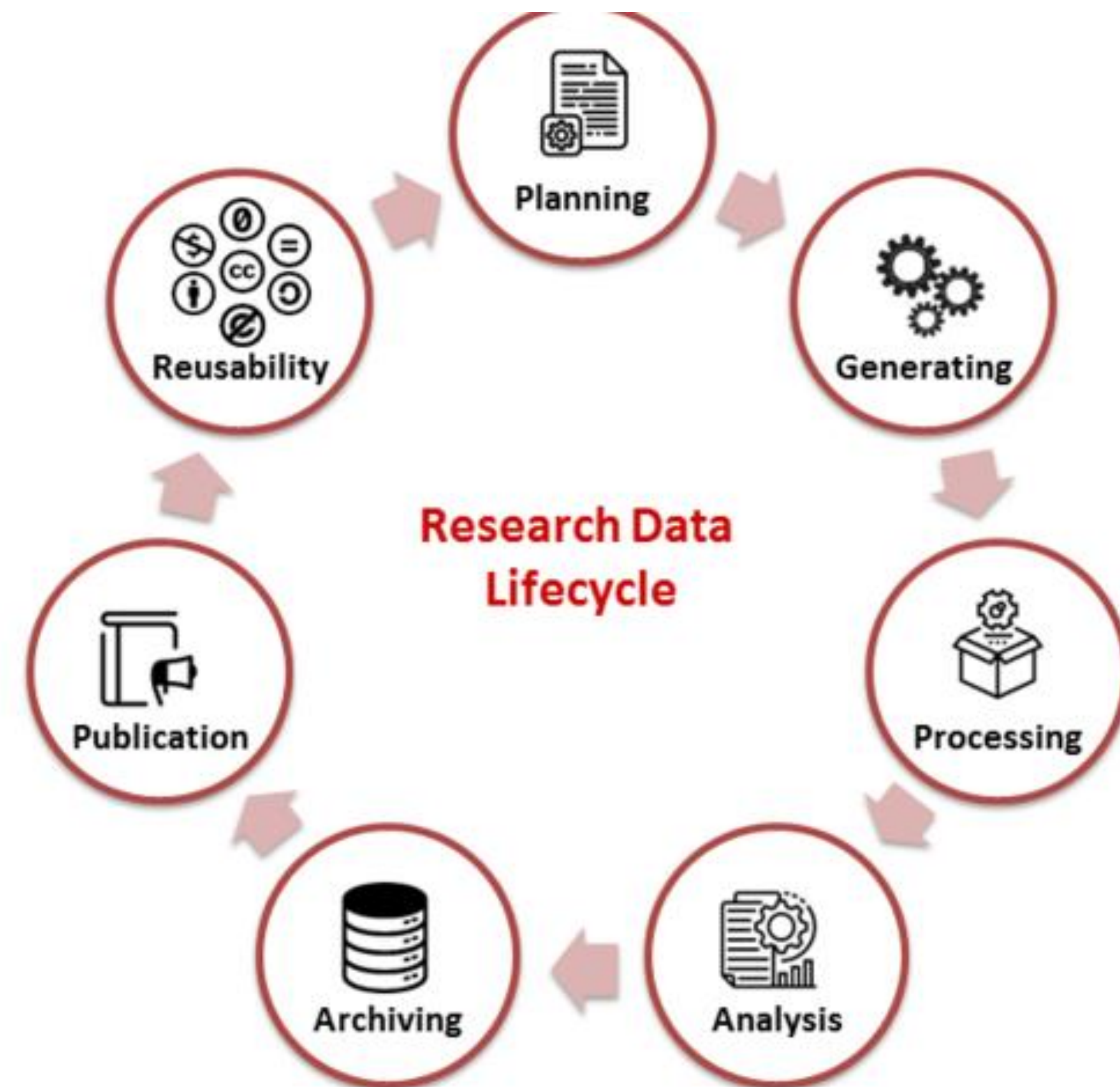
Wikidata Q3105568

Some record data is not displayed in this view. [See JSON view for full record data](#)

Record last modified 2024-05-13. Is there an issue with the data on this record? [Suggest a change](#)

Research data?

- Research data refers to the entities or objects (attributes) that has been observed, generated, and collected to validate the findings during a research project.
- It can be lab-notebooks, field-books, diaries, transcripts, soft-codes, photos, videos, audiotapes, films, and even a text-corpora would definitely be considered as research-data.
- Entire text of the novels written by certain authors or produced during a period can be examined in different directions (say, word frequencies in computational linguistics).



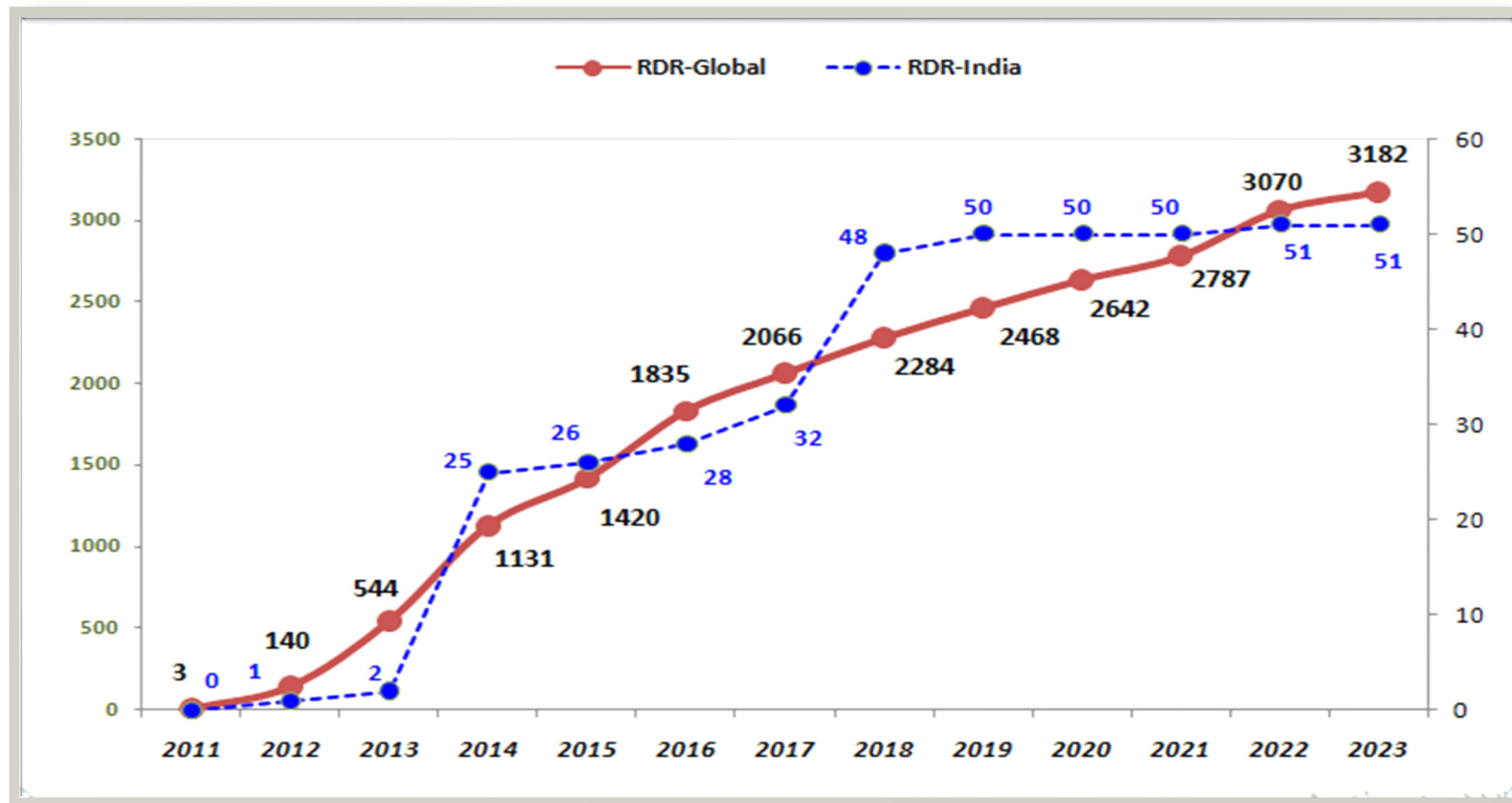
CC BY license: <https://creativecommons.org/licenses/by/4.0/>

Research data Repositories

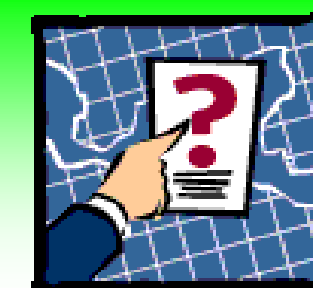
- Data repositories are specialized infrastructures for depositing datasets (along with metadata) in order to enable long-term access and management following open-science dynamics.
- Re3data is a global registry of research data repositories that covers all the disciplines and repository types.
- The ‘registry’ has gone live in autumn 2012 and has been funded by the German Research Foundation (DFG).
- This unique platform provides extendable repository descriptions including access policy, licenses, use of PIDs, and many other aspects of repository management.
- It is currently the most comprehensive (n=3180) source of information on RDRs, which typically disposes its metadata very openly with reliability.



Development of data-repositories (Global vs India)



Obvious queries are...?

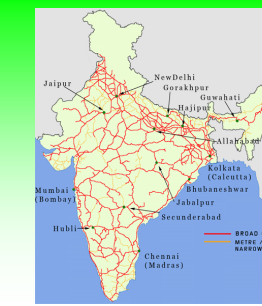


- What's the current status of data-repositories in India?
- Whether these repos are being managed properly?
- Who are the leading players in the country?
- Are they using OAI compliant software?
- Following any metadata standard?
- **Using any persistent identifier?**
- Having their access policy?
- Citation guidelines?
- Usage license?

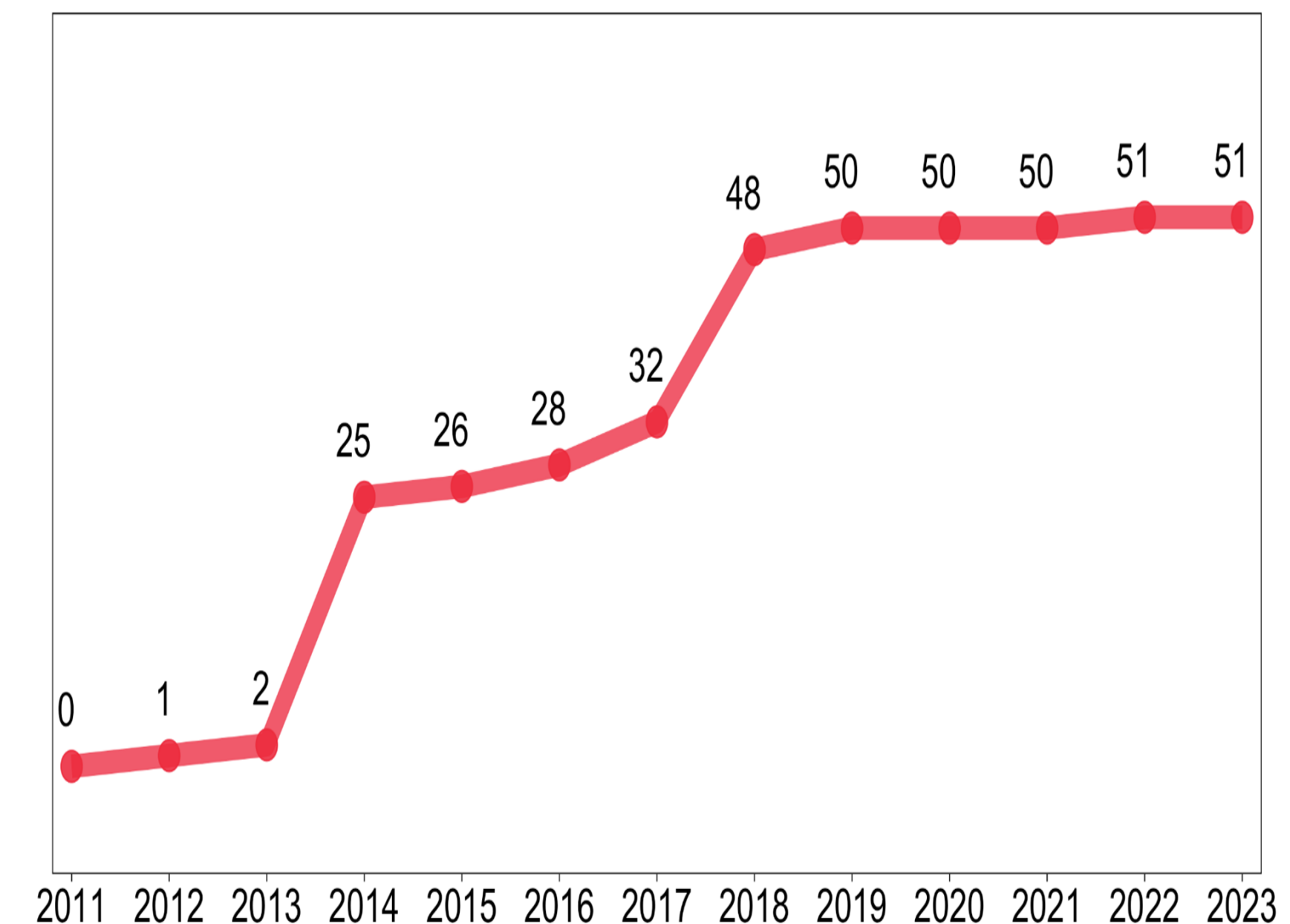


- **How FAIR are data-repositories in India?**
 - Steps to be taken to promote the FAIRness?
 - **What progress has been made yet?**
 - Queries are manifold

Glimpses of data-repositories in India



- There are 51 data-repositories in India as indexed in the Re3data registry (started in 2012); but most the repositories are unstructured with mixed contents and without persistent identifiers for the datasets.
- Data-repositories in the country do not have compliancy with the FAIR data principle – in terms of metadata generation, persistent identification, licensing strategies, citation guidelines, and vocabulary standard.
- However, a few positive signs of adhering certain principles are being observed in Indian scenario; 12 repos are attempted to use a standard metadata schema, 5 repos are using PIDs (DOI=3, Handle-2), 7 repos have CC-licenses, and 6 others have public domain licensing. Two repos are using Dataverse.
- Compared to Asian and BRICS countries, Institutions from India operate a large number of research data repositories.



Citing an example of Indian data-repository

re3data.org

Search

Browse ▾

Suggest

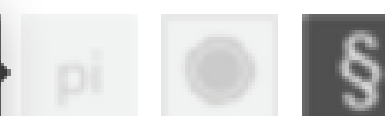
Resources ▾

Contact

Open Government Data Platform India

OGD

The research data repository does not use a persistent identifier system.



Subject(s)

Medicine Geography Economics Geosciences (including Geography)
Natural Sciences Life Sciences Social and Behavioural Sciences
Humanities and Social Sciences

Content type(s)

Scientific and statistical data formats Structured text Standard office documents
Databases Structured graphics

Country

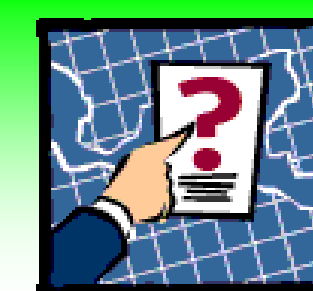
India

The Portal is intended to be used as catalog of datasets published by ministries/ department/ organizations of Government of India for public use, in order to enhance transparency in the functioning of the Government as

Activate Windows
Go to PC settings to activate Windows

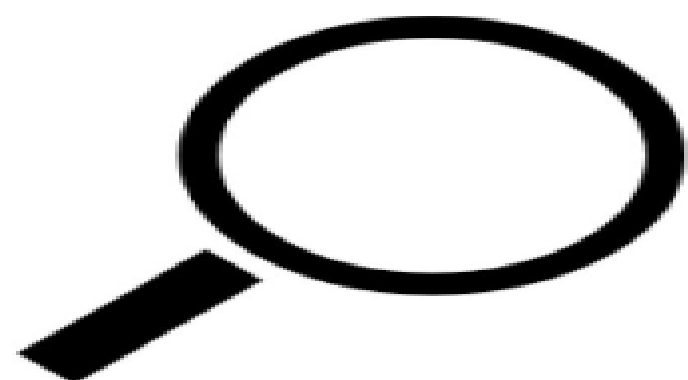
FAIR Data...?

FAIR is an inclusive construct that combines a set of guiding principles

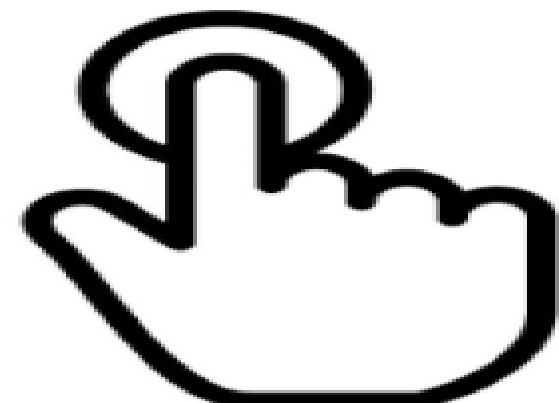


PiD
fest
CONFERENCE
ON PERSISTENT
IDENTIFIERS
2024

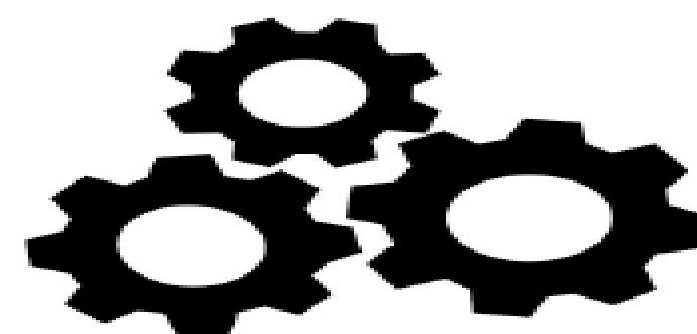
F_{indable}



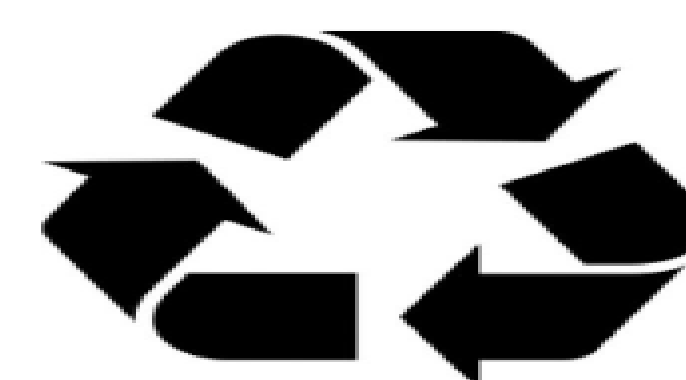
A_{ccessible}



I_{nteroperable}



R_{eusable}



As open as possible, as closed as necessary : FAIR ≠ Open Access

European Union. (2018, November 26). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. Retrieved November 6, 2019, from <https://op.europa.eu/s/npSYd>

:: TO BE FINDABLE ::

- ✓ Whether data-repositories are assigned globally unique and persistent identifier
- ✓ Whether data-repositories are described with rich metadata to support discoverability
- ✓ Whether data-repository records are clearly and explicitly included the identifier of datasets

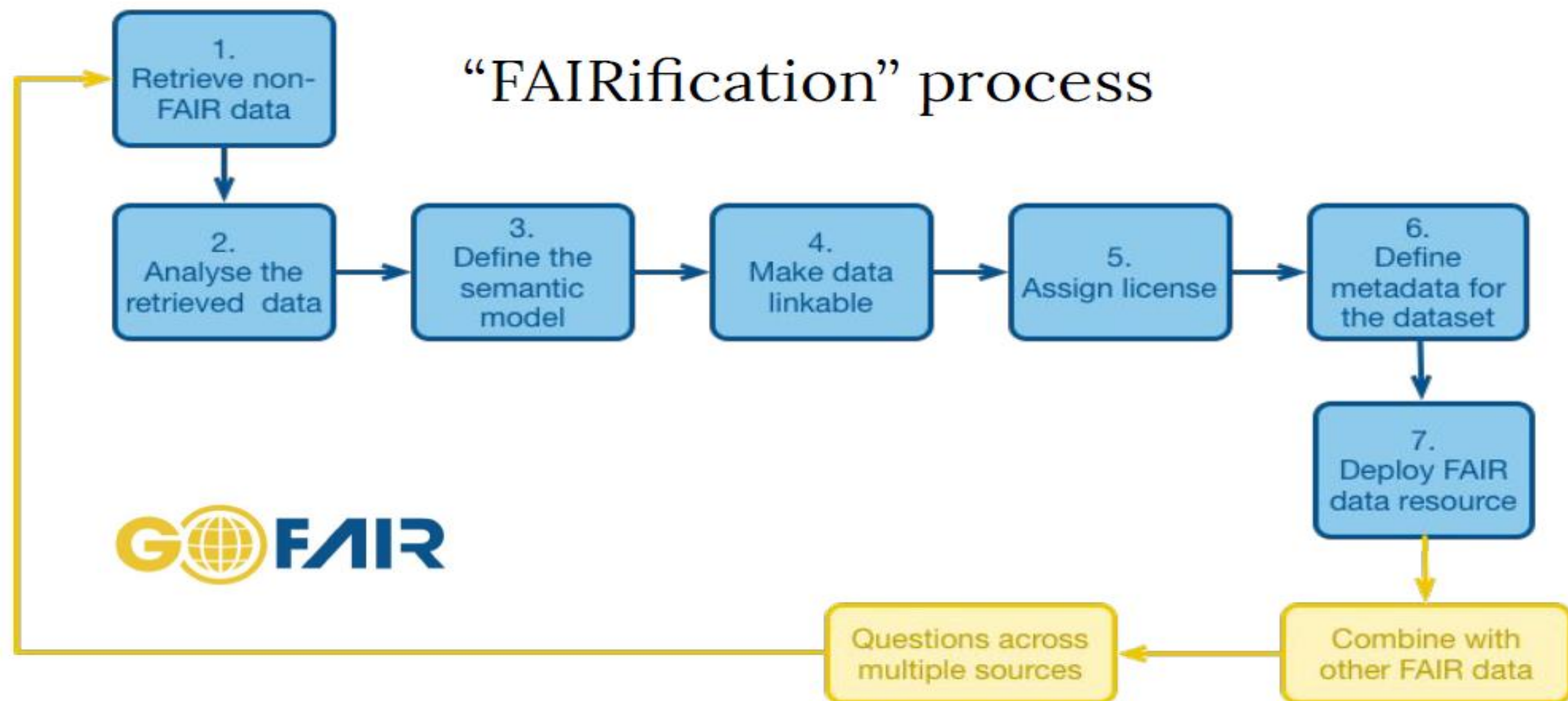
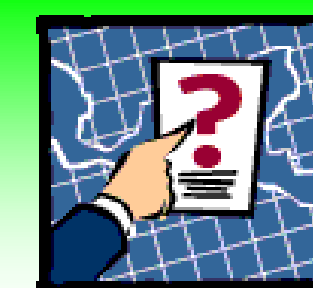
NO

MAY BE

YES

How to make your data FAIR...?

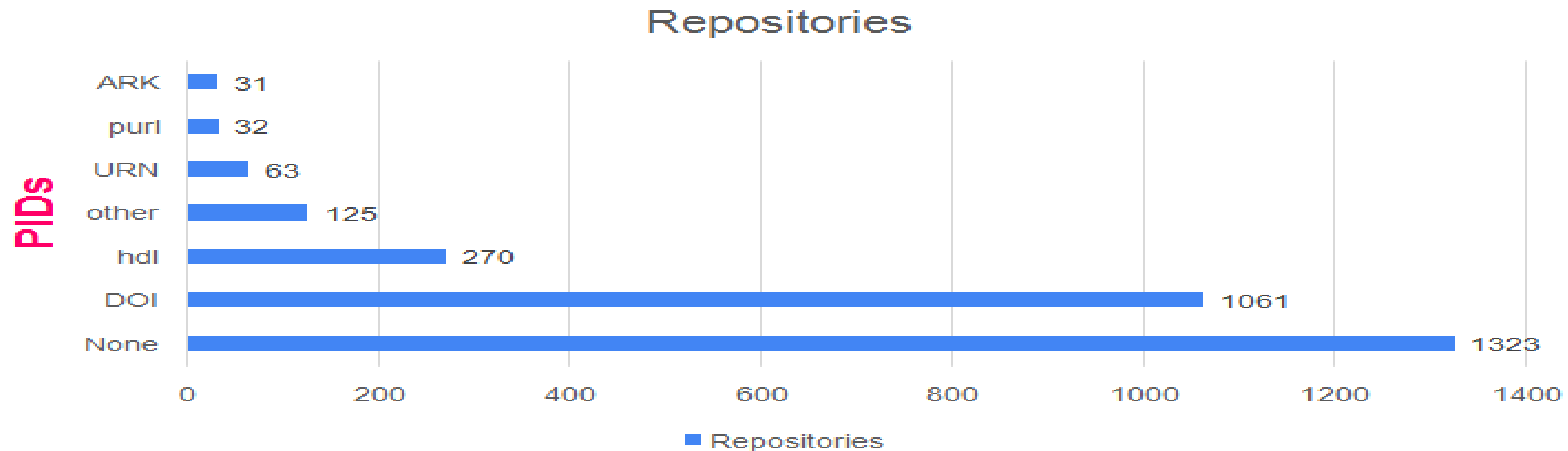
FAIR is an inclusive construct that combines a set of guiding principles:



<https://www.go-fair.org/fair-principles/fairification-process>

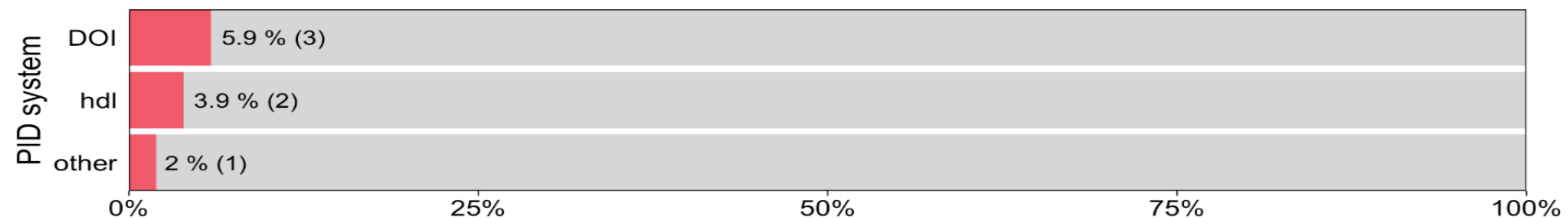
Ac
Go

PIDs in Research Data Repositories



<https://www.re3data.org/metrics/pidSystems>

A total of **43% repositories** uses **PID** system in the global repository landscape and **only 12%** repositories associated with Indian institutions use PID system. Only 6 repositories are using PIDs (Source: *Re3data, May 2024*)



Roadmap for Building National PID Strategies

- Reaching out to major players and key drivers (including service providers and funders) to raise their awareness of the potential benefits of PIDs in advancing scholarly infrastructure
 - Spreading of relevant words on PIDs through open forums and mailing lists
 - Bringing stakeholders together to realize the potential benefits of PIDs
 - Initiating the Strategy discussions and coordinating events
- Defining the collective needs of different stakeholders with a view to developing scholarly infrastructure within the country
 - Reduces burdens of individual decision making avoiding local directives
 - Optimally considers the views of all stakeholders in the country
 - Coordinates input mechanism and drafting of the Strategy
- Refining the value propositions aligning with the strategies adopted by other countries
 - Facilitates further discussions with community consultations
 - Promotes engagement with different working groups
 - Recognizes practical development pathways
- Engaging stakeholders with National PID Strategies to ensure connectivity among the entities of scholarly resources in the research ecosystem
 - Provides unique identifier and fixed location path to connect researchers, objects, and organizations
 - Creates a long-standing reference to a digital object with improved metadata and discovery
 - Better tracking of research entities with interoperability and reproducibility

Concluding remarks



PiD fest
CONFERENCE
ON PERSISTENT
IDENTIFIERS
2024

- This study suggests that data-repositories in India need to be more compliant with the FAIRer world by means of adopting PIDs through close collaboration of diverse stakeholders in different scales.
- There is a pressing need to empower different stakeholders (including researchers, funding bodies, and service providers) to hasten the advancement of scholarly infrastructure in developing nations like India, aligning with open-science practices.
- Indeed this study can be seen as a blueprint for developing National PID Strategies in India aligning with the preset criteria made by other countries so as to ensure global connectivity through extensive use of relevant PIDs throughout the research ecosystem.
- Finally it demonstrates an approach of FAIRness through global infrastructure by integrating digital objects with DataCite services, connecting researchers with ORCiD, and linking organizations with ROR to discover the resources persistently and sustainably.

Challenges for developing countries



PiDfest
CONFERENCE
ON PERSISTENT
IDENTIFIERS
2024

- **Limited awareness of the value of scholarly infrastructure and PID adoption benefits. Community building and collaboration with the international PID communities would have a far reaching impact on developing National PID Strategies in India.**
- **Lack of underlying infrastructure, capacity gaps, and digital divide pose significant challenges in India for implementing PIDs in terms of FAIR governance principles.**
- **An alternative approach is essential for greater inclusiveness and widespread adoption of PIDs through ongoing collaboration between the stakeholders, including service providers and funding agencies.**
- **Financial constraints impact PID infrastructure development and obviously create hurdles for minting high cost DOIs. Increased funding, coupled with policy mandates for PID adoption in research workflows, could accelerate progress.**

Global engagements

- Since 2023, DataCite has been raising awareness through its Global Access Program thereby promoting indian stakeholders toward an equitable and inclusive research ecosystem. This UGC-INFLIBNET project aims to integrate DOIs with Research Data Repositories.
- CORE and INFLIBNET have agreed to collaborate on promoting open science practices with a focus on digital repositories in India.



Never stop questioning...

Never lose your holy curiosity...

(Albert Einstein)

■ Q1 -

■ Q2 -

■ Q3 -

■ Q4 -



References

- Bhoi, N. K., Patel, J., & Dutta, B. (2023). State of Research Data Management Practices in the Top-ranked Higher Education Institutions in India. *International Information & Library Review*, 1–19. <https://doi.org/10.1080/10572317.2023.2167051>
- Cho, J. (2019). Study of Asian RDR based on re3data. *The Electronic Library*, 37(2), 302–313. <https://doi.org/10.1108/EL-01-2019-0016>
- Meadows, A., Haak, L., & Brown, J. (2019). Persistent identifiers: the building blocks of the research information infrastructure, UKSG Insights, 32(9), <https://doi.org/10.1629/uksg.457>
- Misgar, S. M., Bhat, A., & Wani, Z. A. (2020). A study of Open Access research data repositories developed by BRICS countries. *Digital Library Perspectives*, 38(1), 45–54. <https://doi.org/10.1108/DLP-02-2020-0012>
- Simons, N., Brown, C., Bangert, D., & Sadler, S. (2023). RDA National PID Strategies Guide and Checklist (RDA-National PID Strategies WG recommendations). <https://doi.org/10.15497/RDA/00091>
- Singh, R. K., Bharti, S., & Madalli, D. P. (2022). Evaluation of Research Data Management (RDM) services in academic libraries of India: A triangulation approach. *The Journal of Academic Librarianship*, 48(6), 102586. <https://doi.org/10.1016/j.acalib.2022.102586>.

Contributors



Jiban K. Pal, Library Documentation and Information Science Division, Indian Statistical Institute, Kolkata. Email: jiban@isical.ac.in
(<https://orcid.org/0000-0002-2870-9180>)



Dorothea Strecker, Berlin School of Library and Information Science, Humboldt-Universität zu Berlin, Email: dorothea.strecker@hu-berlin.de
(<https://orcid.org/0000-0002-9754-3807>)