

# INDIAN CONTRIBUTION TO MEDICINE DATASETS: AN ANALYSIS

Nayanthara S

Doctoral Research Scholar

Central University of Tamil Nadu

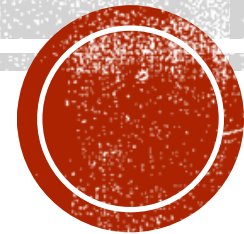
Thiruvarur

Dr. Anila Sulochana

Asst. Professor

Central University of Tamil Nadu

Thiruvarur



# WHY MEDICINE DATASETS?

POPULATION  
GROWTH

AGEING  
POPULATION

MULTIPLE  
CONDITIONS

Learn more  
from health data



Data  
Saves  
Lives

NEW  
MEDICINES  
AND DEVICES

PERSONALISED  
TREATMENTS

BETTER  
INTEGRATED  
CARE

ALGORITHMS  
FOR SAFER  
CARE

# INDIA: HEALTH DIVERSITY





# RESEARCH DATA



- Research data is a well-identified set of data that has been produced (collected, processed, analyzed, shared, and disseminated) by a (again, well-identified) research team.
- The data has been collected, processed, and analyzed to produce a result published or disseminated in some article or scientific contribution.
- Each research data encloses a set (of files) that contains the dataset may be organized as a database, and it can also include other elements as documentation, specifications, use cases, and any other useful material as provenance information, instrument information, etc. (Gomez-Diaz & Recio, 2022).



# OBJECTIVES

- To figure out India's contribution to the datasets that belong to Medical Sciences in re3data.org.
- To examine the nature and various aspects of the Indian Medicine datasets in re3data.org.
- To understand the collaboration pattern of medicine datasets in re3data.org.



# RE3DATA.ORG

- A global registry of research data repositories that covers research data repositories from different academic disciplines.
- Includes repositories that enable permanent storage of and access to data sets to researchers, funding bodies, publishers, and scholarly institutions. re3data promotes a culture of sharing, increased access, and better visibility of research data.
- The registry went live in autumn 2012 and has been funded by the German Research Foundation (DFG).



# METHOD

- The Indian Medicine raw datasets are listed using various filters from re3data.org and then analyzed based on the following aspects:
  1. Do these datasets follow fair data principles?
  2. Are these datasets citable?
  3. Is the dataset source clearly stated, corroborating the authenticity?
  4. Who are the major contributors to these datasets?
  5. Is the privacy policy stated, whether it is of patients or medical data?
  6. What category of access is provided to these datasets?
  7. What kind of licensing is used for these datasets?
  8. What are the keywords used within these datasets?



# **ATTRIBUTES AND CONDITIONS: FAIRSHARING.ORG**

- Data Curation
- Data Access Condition
- Resource Sustainability
- Data Preservation Policy
- Data Deposition Condition
- Data Versioning
- Data Contact Information
- Citation To Related Publications
- Data Access for Publication Review





# METADATA AND DATA STANDARDS FOR HEALTH DOMAIN IN INDIA

- The “Metadata and Data Standards” initiative by the Ministry of Communication and Technologies under the National e-Governance Plan (NeGP).
- Aims to promote the growth of e-Governance within the country by establishing interoperability across e-Governance applications for the seamless sharing of data and services.
- Interoperability at the institutional level would necessitate conversations between public health organizations in order to understand information needs and barriers to better information quality and use--much of which relates to the information collection and recording, patterns of flow and aggregation, and contexts of use of information rather than semantic or technical considerations.
- The datasets indexed in re3data.org from India are not exclusively dedicated to medicine. So, the MDDS standard is not applied.



# FINDINGS

1. Of the 197 medicine databases with raw data sets, six are from India (3.04%).

1. *Maharashtra State Data Bank*
2. *Open Government Data Portal of Odisha*
3. *Open Government Data Portal of Surat City*
4. *Human Proteinpedia*
5. *Open Government Data Portal of Tamil Nadu*
6. *Indian Genetic Disease Database*



# FINDINGS

2. The major contributors to these Indian datasets are Non-Profit Organizations collaborating with the state government.
3. All six of these databases follow fair data principles, though not to a complete extent.
4. The source of data is authentic in all six data sets as various state governments and the assigned departments or legitimate non-profit organizations carefully curate all these datasets.
5. All six of these data sets are open.
6. All of them use either any of the various copyrights or Open Government Licenses.
7. All six databases are citable, and the citation is given at each repository's profiles with DataCite citations.
8. The Major keywords used are Public Health, Medicine, Biology, Social Medicine, Public Health Research, Life Sciences, etc.



# CONCLUSION

- The availability and accessibility of research data will aid in discovering new knowledge and developing new approaches for improving research quality.
- To advance knowledge, researchers affiliated with institutions and organizations should be motivated to deposit and incorporate data gathered through their research into institutional research data repositories or other discipline repositories.
- India has not yet become a significant global contributor to medical datasets. The six datasets currently indexed in re3data.org encompass a range of subjects beyond medicine.
- Given India's vast subcontinental expanse and its diverse genetic and health profiles, there is considerable potential for further advancements. The absence of standardized data formats, coding systems, and data exchange protocols.
- The availability of high-quality medical datasets is crucial for medical research, drug discovery, and treatment development. The lack of standardized datasets can hinder the progress of research and innovation in the healthcare sector.



**Thank You**

