

Gandhinagar, Gujarat, India

26-28 October 2023

# Landscape of Metadata Schemas for Research Data Repositories

## Fairsharing Analysis

Manu T R<sup>1\*</sup> and Nabi Hasan<sup>1</sup>

<sup>1</sup>Central Library, Indian Institute of Technology Delhi

Corresponding mail: [manutr@library.iitd.ac.in](mailto:manutr@library.iitd.ac.in)

26<sup>th</sup> International Symposium on Electronic Theses and Dissertations ETD 2023, Gujarat, India 26-28 October 2023

# Coverage

- Introduction
- Best practices of metadata for research data
- Objectives of the Study
- Methodology
- FAIRsharing
- Research results
- Conclusion
- References

# Introduction

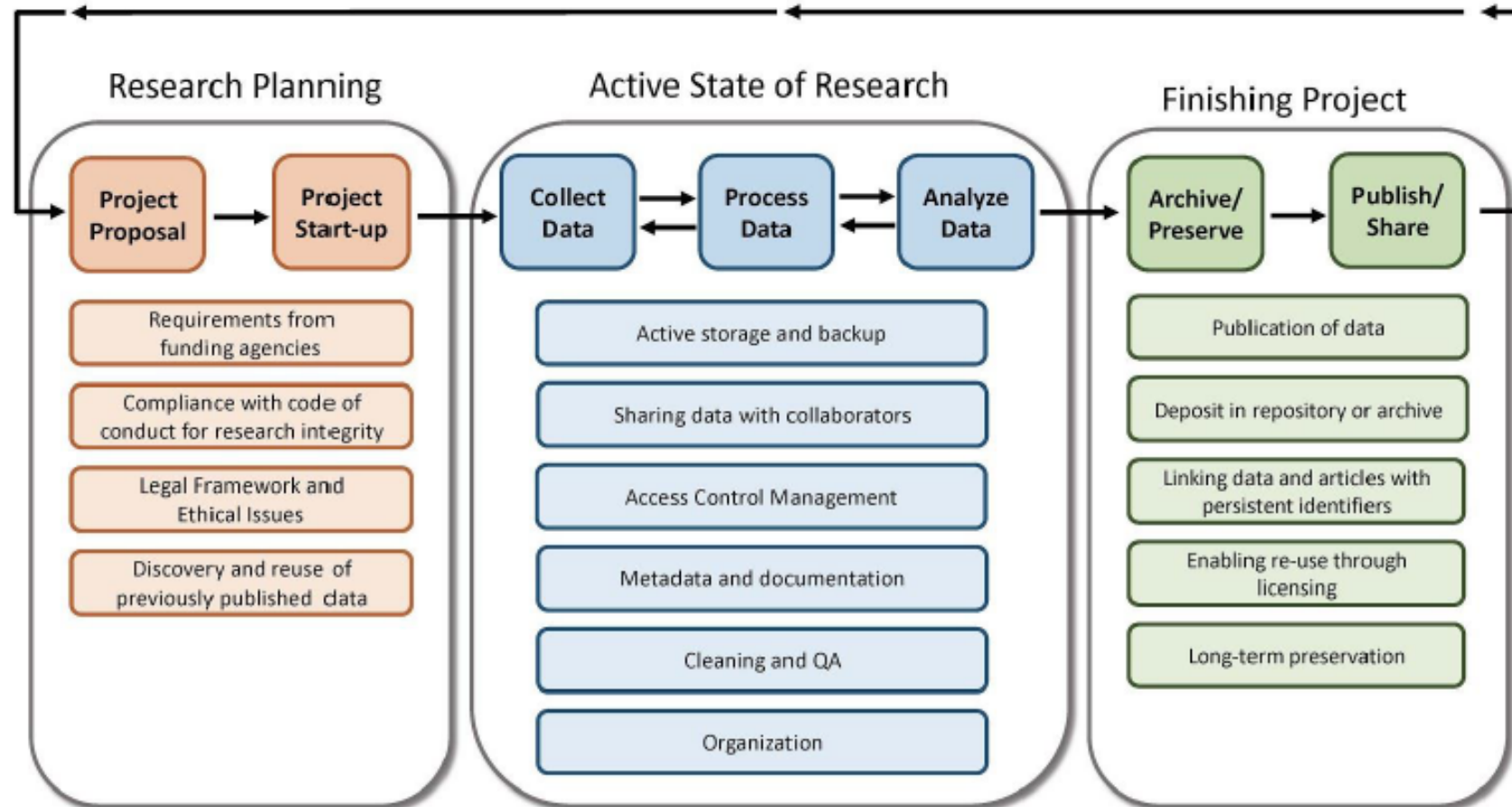
- Research data is the recorded information generated while conducting research, writing an article, theses and dissertations, and other research processes.
- Providing access to research data would be a challenge
- Several metadata schemas are available for describing the research data repository
- Need to study the metadata standards to facilitate the selection of appropriate metadata standards
- RDA Metadata Standards Directory, [re3data.org](https://re3data.org), and FAIRsharing
- Metadata schemas can be used to generate metadata for research datasets, which differ in types of data, data elements and content.
- A decision about selecting a metadata schema for a research repository has implications for the quality of analysis.

# Best Practices of Metadata for Research Data

The best practices of metadata for research data are to create

- Data dictionary
- Create
- Manage
- Documentation
- Data storage systems
- Describe the contents of data files
- Document taxonomic information
- Maintain consistent data typing
- Separate data values from annotations
- Understand the geospatial parameters of multiple data sources

# Metadata and the Research Data Lifecycle



## Objectives of the study

The present study is being undertaken with the wide objective of analyzing the landscape of metadata schemas available for research data repositories. The focused objectives are as follows:

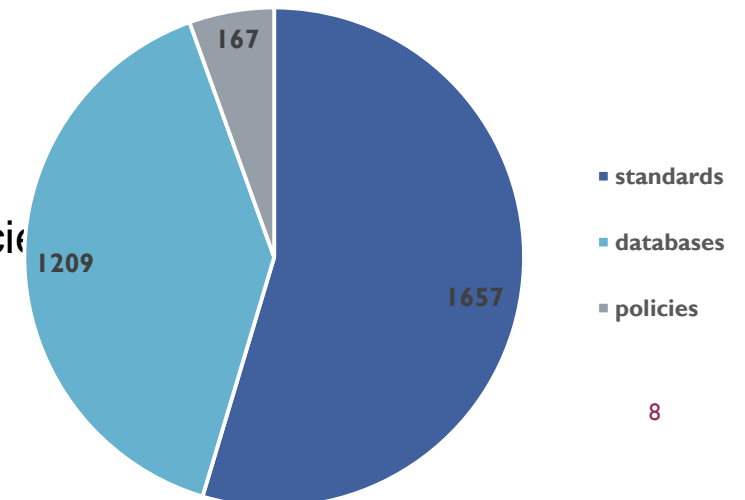
- To identify the list of metadata schemas, available for research data repositories.
- To study the growth and development of metadata schemas
- To study and distinguish the disciplinary metadata schemas.
- To identify the country-wise contributions, organizations, government bodies and institutional contributions to developing and actively maintaining metadata schemas.
- To analyse metadata schemas by active and readily available for use, supported record types and domains, and recommended license attributions, etc.

# Methodology

- The researchers have used the Fairsharing directory as a data source for the study.
- Required data retrieved through RESTAPI as on April 25, 2023.
- Data including general data, metadata details, grant data, and publication data of the metadata schemas.
- The extracted data were cleaned and analyzed using online visualization tools (RAWGraphs) and advanced Excel features.
- The result presented in the graphical and tabular formats fulfills the objectives of the study.
- The researchers do not consider the following two directories as data sources.
  - RDA Metadata Standards Directory.
  - re3data.org

# FAIRsharing

- Fairsharing is a community-driven platform to provide information and resources on metadata standards, data policies, databases and repositories.
- The community works together to enable FAIR principles by promoting the value and use of the standards, databases and policies for research data.
- It covers the standards, databases, repositories, and data policies, accelerating the discovery, selection, and use of these resources.
- As of August 2023, FAIRshaing has over 3,868 records which includes.
  - 1657 standards
  - 1209 databases
  - 167 policies (of which 87 are from journals, 33 from funders, 15 from social media, 14 from projects, 13 from journal publishers and 6 from institutions).





STANDARDS

DATABASES

POLICIES

COLLECTIONS

ORGANISATIONS

ADD CONTENT

STATS

A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.

Guides **consumers** to discover, select and use these resources with confidence.

Helps **producers** to make their resources more visible, more widely adopted and cited.

Provides **humans** and **tools** with access to trustworthy content to enable data management tasks.

RESEARCHERS

DEVELOPERS & CURATORS

JOURNAL PUBLISHERS

LIBRARIANS & TRAINERS

SOCIETIES & ALLIANCES

FUNDERS

## Journal publishers & organisations with data policies

Create and maintain an interrelated list of citable standards, databases and repositories to recommend to your authors, users or their community, and revise this recommendation over time...

[read more](#)



1678 Standards

Terminology Artifact	798
Model/Format	582



2055 Databases

Repositories	1060
Knowledgebases	841



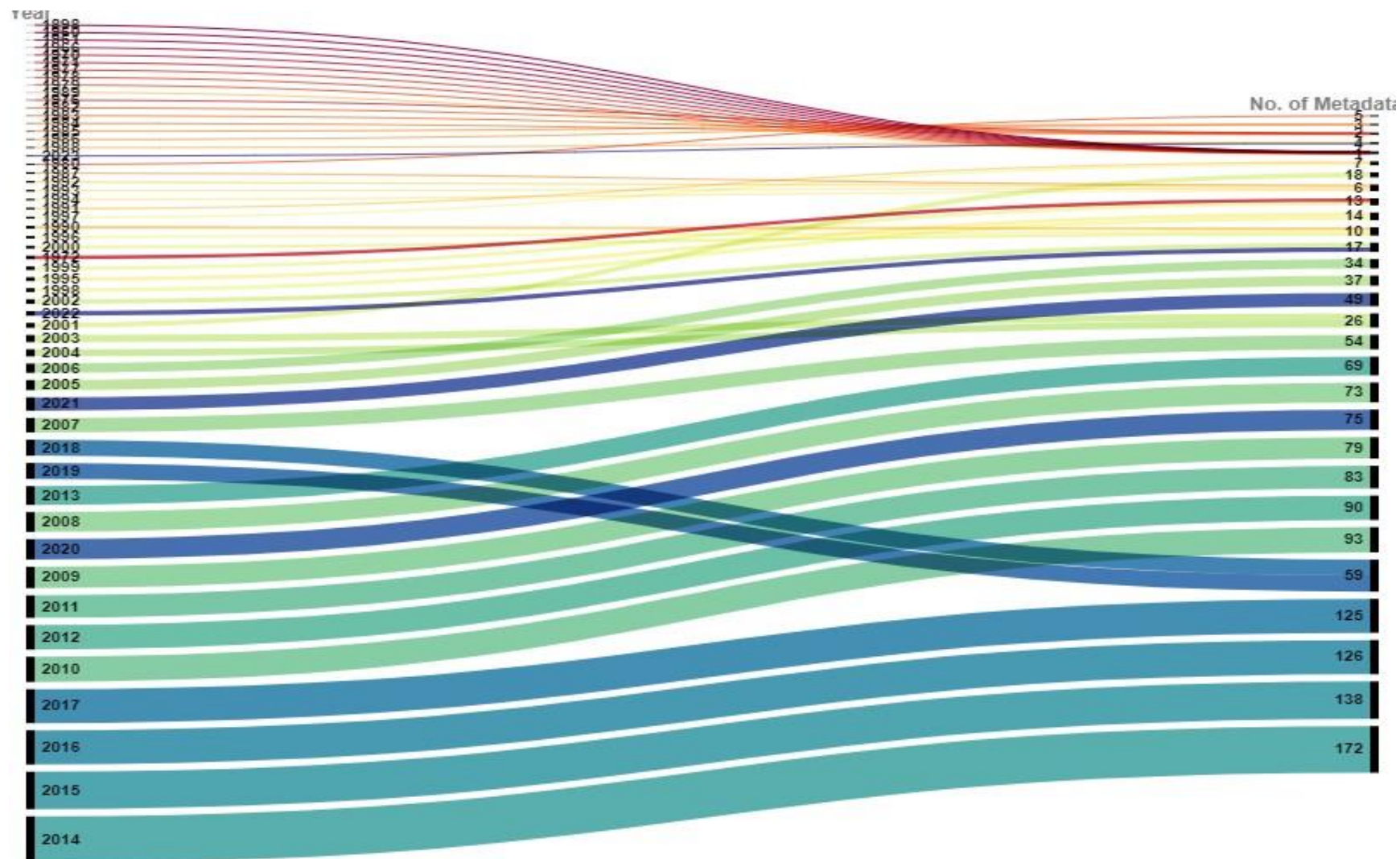
168 Policies

Journal	87
Funder	33

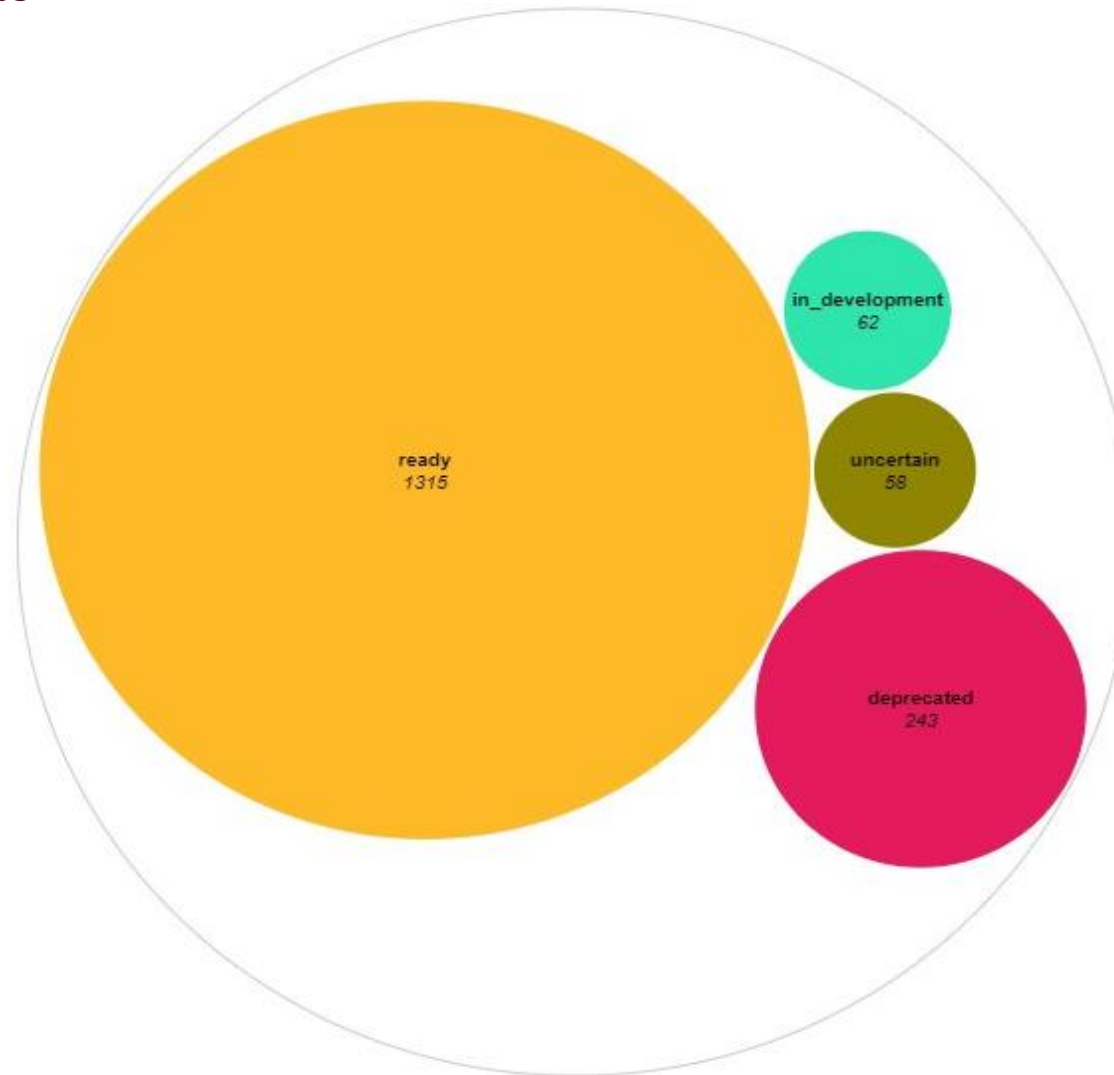
## Research results

- Over 1657 metadata schemas for research data have been developed since the 1900s.
- These are at various stages like:
  - 1309 metadata schemas are active and ready for use
  - 184 metadata schemas have been deprecated
  - 58 are unsure of their status
  - 62 are currently in the development stage and not ready for use
- Out of 1678 metadata schemas
  - only 99 (5.89%) have been recommended by a data policy from a journal, publisher, or funder
  - the remaining 94.10% (1579 schemas) have not been recommended due to various restrictions.

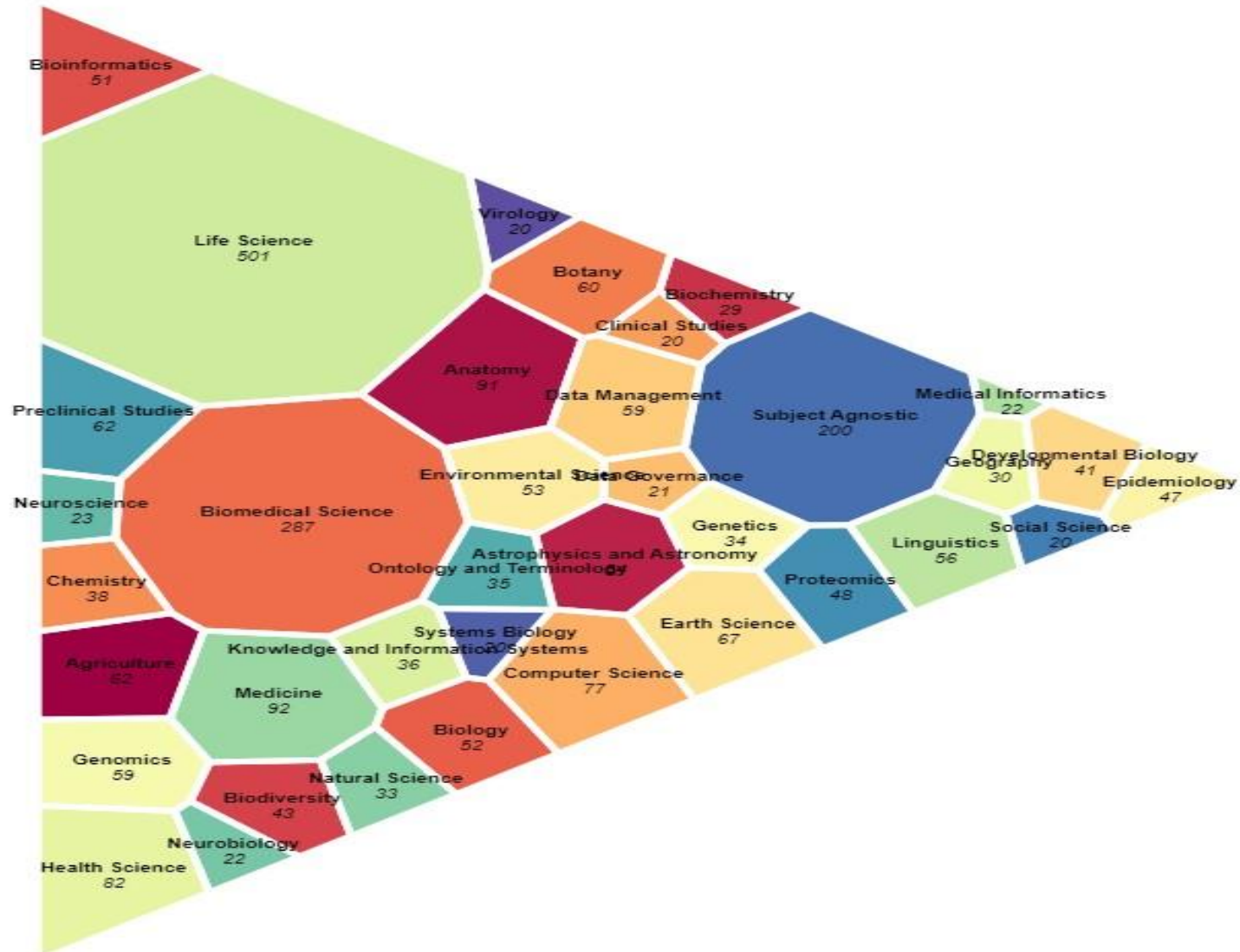
# Growth and development of metadata schemas



## Status of metadata schemas

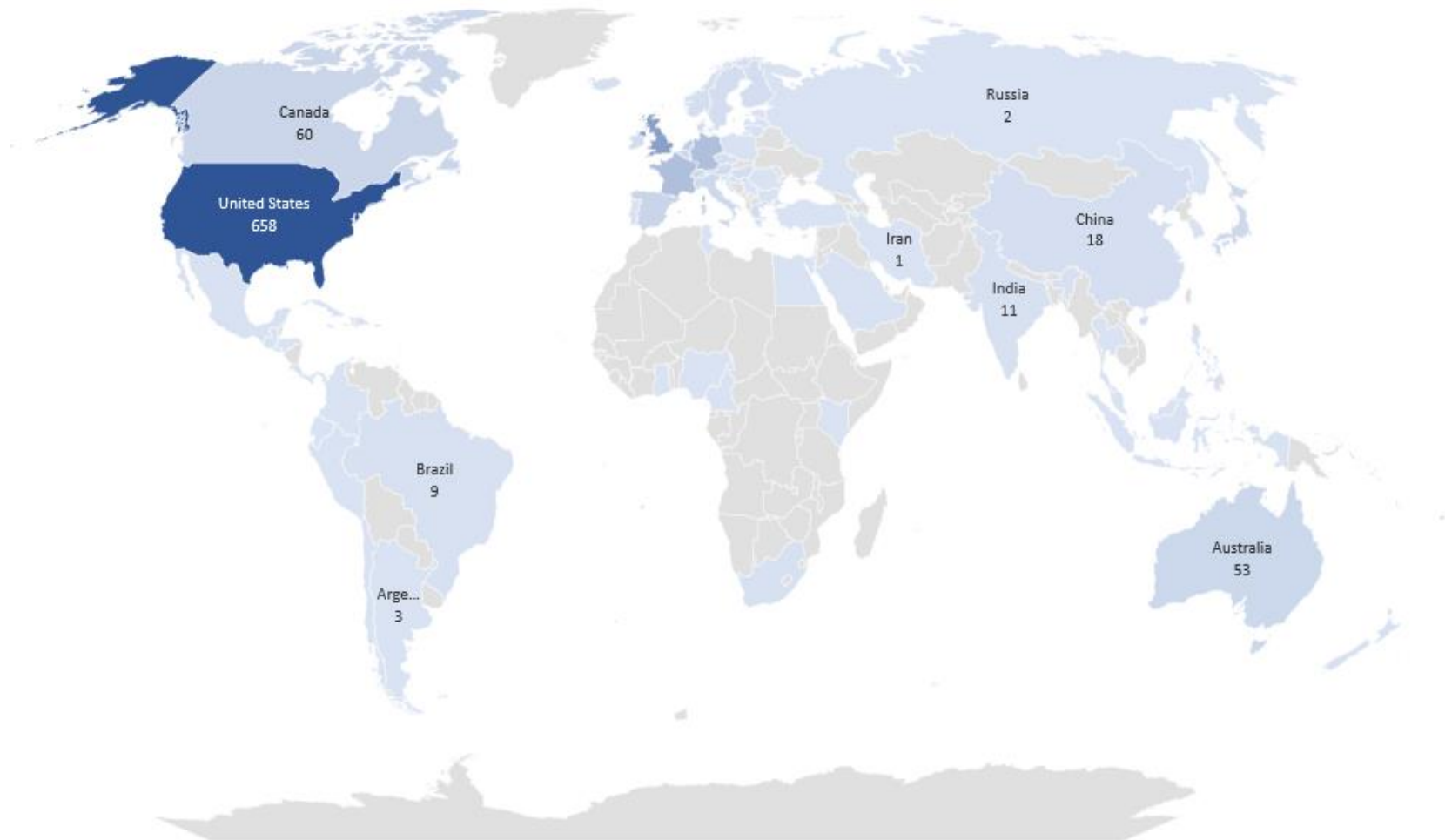


## Subject-wise metadata schemas



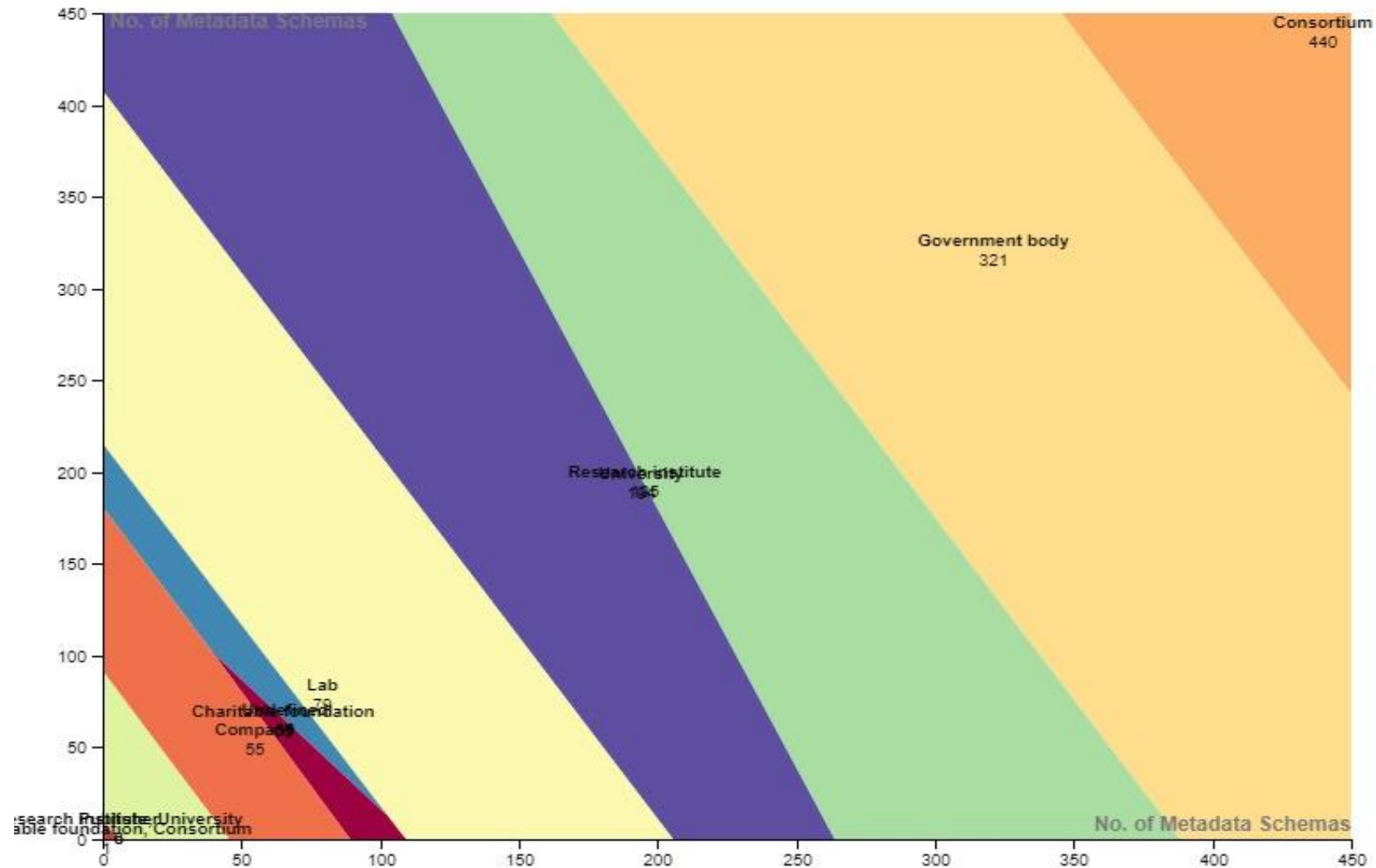


## Country-wise contribution on metadata schemas development

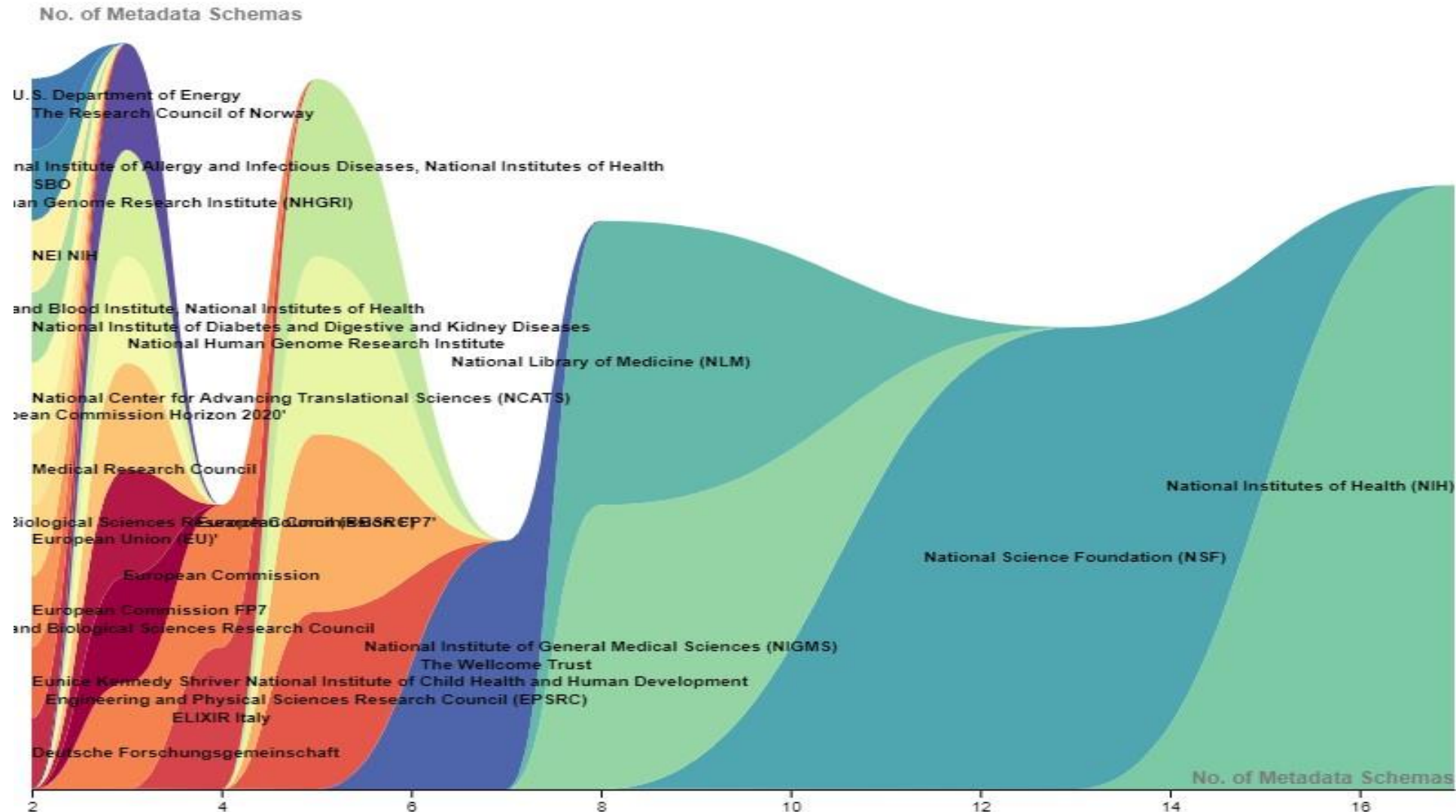


Powered by Bing  
© Australian Bureau of Statistics, GeoNames, Microsoft, Navinfo, OpenStreetMap, TomTom, Zenrin

## Types of Organizations maintaining metadata schemas



# Organizations funding metadata schemas





## Licenses developing metadata schemas



# Summary

- The present study gives a clear impression of metadata schemas available for research data repositories.
- It found over 1500+ schemas with interdisciplinary subjects, but 78% were active and ready for use.
- It ensured that the schema used is stable, applicable for all kinds of data, recommended by the data policies of journals, publishers and, funders etc.
- The study also helps professionals, decision-makers and government bodies to know:
  - Countries are majorly contributing to schemas development
  - Funding agencies or government bodies have been funding, and
  - Organizations or institutions are maintaining the metadata schemas available for research repositories.
- It also highlights the metadata schemas which recommended using various license attributes.
- The National Institutes of Health (NIH), National Science Foundation (NSF), National Library of Medicine (NLM), European Bioinformatics Institute, etc. are major top organizations
- It also covered the major top research and academic institutions from India.

## References (Selected)

- Basoni, A., Menegon, S. & Sarretta, A. (2015). Sailing towards open marine data: the RItMARE Data Policy. *ERCIM News* (100), pp. 22-23.
- Carlson, J. & Stowell-Bracke, M. (2013). Data management and sharing from the perspective of graduate students: an examination of the culture and practice at the water quality field station. *portal: Libraries and the Academy*, 13(4), 343-361. <https://dx.doi.org/10.1353/pla.2013.0034>
- Cudre-Mauroux, P. (2020). *Design considerations on SWITCH's connectome vision*. Zorich: A SWITCH Innovation Lab.
- DataOne. (2023, 05 18). *Metadata*. Retrieved from Data Observation Network for Earth: <https://www.dataone.org/best-practices/metadata>
- Douglass, K., Allard, S., Tenopir, C., Wu, L. & Frame, M. (2014). Managing scientific data as public assets: data sharing practices and policies among full-time government employees. *Journal of the Association for Information Science and Technology (JASIST)*, 65(2), 251–262. <https://dx.doi.org/10.1002/asi.22988>
- Ensom, T. & Corti, L. (2012). *Research data management at the University of Essex findings from a pilot study*. ESSEX: UK Data Archive.
- Eynden, V. V., Corti, L., Woollard, M., Bishop, L. & Horton, L. (2009). *Managing and sharing data*. Colchester: UK Data Archive.
- FAIRsharing (2023, 05, 29) *FAIRsharing: standards, databases, policies*. <https://fairsharing.org/>
- GEO. (2015). Data management principles implementation guidelines. *Group of Earth Observations*. GEO-XII.
- Jeffery, K. G., & Koskela, R. (2015). RDA: the Importance of Metadata. *ERCIM NEWS*, ( 100), pp. 23-24.
- Kethers, S., Shen, X., Treloar, A.E. & Wilkinson, R.G., (2010). Discovering Australia's research data. *Proceedings of the 2010 Joint International Conference on Digital Libraries, JCDL 2010, Queensland*. <https://dx.doi.org/10.1145/1816123.1816175>
- Kosinov, A., Erkimbaev, A., Kobzey, G. & Vladimir, Z. (2019). Data curation approach to management of research data, use cases for



# Thank You