Scientometrics (2022) 127:4227–4236
https://doi.org/10.1007/s11192-022-04395-6

Indian Science Reports: a web-based scientometric portal for mapping Indian research competencies at overall and institutional levels

Vivek Kumar Singh[1] · Abhirup Nandy[1] · Prashasti Singh[1] · Mousumi Karmakar[1] · Aakash Singh[1] · Hiran H. Lathabai[1] · Satya Swarup Srichandan[1] · Anurag Kanaujia[1]

Received: 3 February 2022 / Accepted: 26 April 2022 / Published online: 7 June 2022
© Akadémiai Kiadó, Budapest, Hungary 2022

Abstract
The article presents an introduction to a newly created scientometric portal called Indian Science Reports, available at www.indianscience.net. The portal is designed to fulfil the need for a single integrated resource for analytical data about research competencies of India at an overall level as well as Indian institutions at an individual level. India's research performance in terms of research output, citations, highly cited papers, international collaboration, open access levels, gender distribution and social media visibility etc. are computationally analysed using publication metadata collected from Dimensions database. The portal also provides a mechanism to look for research performance of all major Indian higher education institutions, on various standard parameters, through an institutional search. Further, a concept-based search is integrated to identify top performing Indian institutions on a given research topic. The portal, thus, provides an invaluable resource of Indian scientific research data and information, which can be used for various purposes ranging from scientometric evaluation to thrust area-based funding decisions.

Keywords Indian science · Research competitiveness · Research evaluation · Scientometrics

Introduction

Science Technology and Innovation (STI) Data plays crucial role for the policymakers across the world to make informed and actionable policy decisions. The requirement for measurement of research output is higher than ever before as the government, and funding organisations heavily depend upon these measurements when deciding upon award of research grants, performance-based funding, etc. Cataloguing and measurement of STI

✉ Vivek Kumar Singh
vivek@bhu.ac.in

✉ Anurag Kanaujia
anuragkanaujia01@gmail.com

[1] Department of Computer Science, Banaras Hindu University, Varanasi 221005, India

data is an important exercise which has been done since the adoption of scientific method. Traditionally, this STI data is collected by international agencies, such as OECD, World Bank, United Nations etc. As these organisations focus on national level indicators, the data available through them does not provide detailed insights at micro scale (Prathap, 2018). The role of Scientometrics as a methodology for drawing actionable inferences is important in this regard. It provides a set of tools that can be used with a significant level of reliability to compare different disciplines which would otherwise be a difficult exercise (Sooryamoorthy, 2020). The field of scientometrics has developed from a cataloguing and description activity into an analytical science. It is now being used not only for mapping of research in any field and measurement of scientific output but is also being employed to evaluate national R&D performances, check for patterns of activity in various disciplines and even to reduce the time and cost for review of submitted manuscripts before publication by journals (Khokhlov, 2020).

Before presenting the motivation and objectives for the Indian Science Reports portal, we look at some of the popular STI data platforms and their key characteristics. OECD's STI statistics (https://www.oecd.org/innovation/inno/stistatistics.htm), provides country specific information on the STI indicators related to Human and financial resources devoted to R&D. These indicators are provided at a country level and do not provide the finer details for each country i.e., at the level of institutions, or even the research focus and activity areas of a country. Another popular platform is a web application called Excellence Maps (https://www.excellencemapping.net/), which is developed with the aim to visualise the performance of universities and research institutions on an overall level and in some selected subject areas. This application uses publication data available from Scopus in order to rank the institutions on the basis of their performance. The platform, however, uses four variables only, including proportion of papers from one institution and three non S&T indicators (Bornmann, et al., 2015). The research done at the national level is largely left out of the assessment criteria in both the platforms. Consequently, regional resource databases such as Airiti (http://www.airiti.com/) (China), African Journals OnLine (AJOL) (https://www.ajol.info/) (Africa), Indian Citation Index (ICI) (http://www.indiancitationindex.com/) (India) etc. provide access to country specific research publications. Ranking platforms such as the Global Innovativeness Index[1] (https://www.wipo.int/global_innovation_index/en/), Times Higher Education (THE) University rankings (https://www.timeshighereducation.com/), Academic Ranking of World Universities and Quacquarelli Symonds (QS) World University Rankings (https://www.topuniversities.com/), Centre for Science and Technology Studies (CWTS) Leiden Rankings (https://www.leidenranking.com/) have also been using various strategies for measurement of organisational performance of universities, research institutions and other higher education institutions. However, most of these platforms/ portals provide limited data analytics and often suffer from underrepresentation of institutions in developing countries like India.

The scientific community in India is rapidly growing both in terms of number of individuals and its contribution to the global scientific research and development activities. Therefore, suitable repositories and analytical platforms are needed to report on Indian scientific research. Two main repositories of information are relevant to mention here, namely, National Institutional Ranking Framework (NIRF) (https://www.nirfindia.org/) and the Department of Science and Technology (DST) portal (https://www.indiascienceand

[1] GII provides performance measures and ranks 132 economies on their innovation ecosystems.

technology.gov.in/). However, the overall context and intended purpose of these portals are significantly different, when compared to the Indian Science Reports portal as highlighted below.

The NIRF is a ranking framework for Indian Institutions, which focuses on five parameters namely, Teaching learning and resources; Research and professional practice; Graduation outcomes; Output and inclusivity; and Perception about the institutions (https://www.nirfindia.org/Docs/Ranking_Methodology_And_Metrics_2017.pdf). It provides 30% weightage to the research outputs and hence is not solely a research-based ranking framework. Some previous studies have also highlighted that the weightage to research output under NIRF is reduced because the evaluation methodology combines various dimensions of institutional performance into a single score, making it less suitable for the purpose of scientometrics and for use by researchers and policy makers (Mukherjee, 2019; Prathap, 2017). Further, NIRF does not provide any analysis on the parameters of collaboration, open access, gender distribution, research grants etc.

The DST portal, on the other hand, is mainly a repository of information about funding and institutional systems involved in science. It provides data about research projects funded by DST in various thematic areas, research funding schemes available, listing of organisations, various government programmes for S&T development, fellowships for students and professionals, start-up ecosystem and list of innovations across the country etc. Thus, it is more focused on the input side (involving funding, manpower, institutions, enabling infrastructure etc.) of the scientific ecosystem. It does not provide any scientometric analysis of the research output from India and Indian institutions. This is a major gap which served as the motivation for development of a portal which can represent the details of research output for India at various levels of granularities. The Indian Science Reports Portal is an attempt to develop a framework for assessment of these activities by using scientometric tools and presenting them in an organised, simple and easily accessible format. The following sections provide a detailed introduction about the portal and its salient features that can be useful to the intended audience for both research and information purposes.

The Indian Science Reports portal

The Indian Science Reports Portal is developed as a platform to present indicators on research performance of Indian universities, research institutions and other organisations, along with a national S&T output and analytics reporting system. This platform is developed by leveraging the standard scientometric and data analytics techniques and utilises the research publication listing available on Dimensions (https://app.dimensions.ai/) to develop detailed insights into the research performance. It also provides analytical insights through the application of novel indicators such as x-index and x(g)-index (Lathabai et al. (2021a, b). The platform/portal is designed with multiple levels of information richness. It provides broad outlook in terms of comparison between the research performances of India with selected high performing countries across the globe; as well as allows the user to zoom in to see the research performance of individual institutions by looking at indicators such as total publications, total publications in different disciplines, international collaboration, gender distribution, open access etc. It also includes information on indicators which have developed relatively recently and do not feature on other ranking and analysis platforms. These include social media visibility of the research, research on Sustainable Development

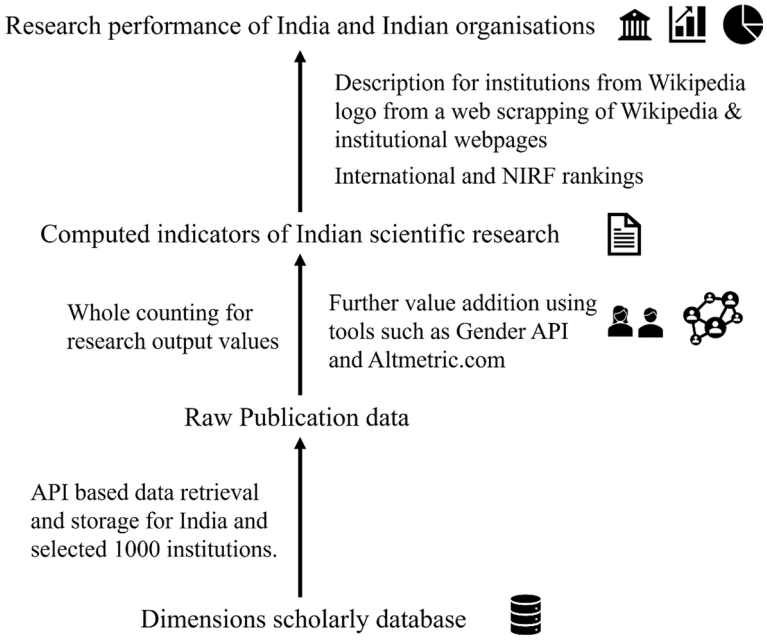

Fig. 1 The schematic representation of the process of data retrieval, processing and presentation on the Indian Science Reports portal

Goals (SDGs), and competence areas of the institutions based on their thematic research strengths.

Data and methodology

The data for the development of this portal was taken from Dimensions database through the subscription access provided by Dimensions API (Application Programming Interface). The period of analysis was 2010–2019 and the document types considered were "article" and "proceedings" due to the majority of research reported under these types. Full counting method (Vinkler, 2010) was preferred for results for the reasons of simplicity. In order to obtain the social media presence of publications, the subscription-based access of Altmetric.com was used. The Altmetric.com provides the mentions of publication on various social media platforms such as Twitter, Facebook, Wikipedia, Blogs etc. The gender of the authors was determined using Gender API,[2] a subscription-based platform which determines the gender of a given name using the first name and the country of the author. The first author's first name and country of affiliation were extracted and passed through Gender API platform. The gender value for an author which had an accuracy value of more than or equal to 70% were selected. The collected analysis was used to develop visual representations for selected indicators using Python and JavaScript, and incorporated on

[2] https://gender-api.com/en/api-docs.

Table 1 List of areas and indicators covered in the Indian Science Reports Portal

National research output	Institutional reports (for Indian institutions)
• Research output	• Short introduction
• Citations	• Key indicators
• Collaboration	• Research output and citations
• Gender distribution	• Authorship and collaboration
• Open access	• Research grants
• Social media visibility	• Gender distribution
• Grants	• Open access availability
• SDG related research	• Social media visibility
• Major institutions	• UN SDG related research
	• Thematic structure of research output
	• External data (Year 2021)
	• Ranking from five major ranking platforms

the web-based user interface. Figure 1 presents a schematic diagram of the process and methodology.

Functionality and unique characteristics of the portal

The Indian Science Reports portal has been developed by utilising available scientometric data for India as well as data for selected 1000 Institutions in India (selected in descending order of total publications). The indicators used have either been taken from Dimensions database or developed and calculated through programs by analysing the data for the period between 2010 and 2019. It uses indicators about research output, citations, authorship, international collaboration, gender distribution, open access, research grants and social media visibility etc. for the Indian research output during 2010-19 (Table 1). A total of 50 + indicators are considered in this academic exercise and represented visually over the 10 year period.

National level indicators

The metadata collected from the dataset is used to develop and compute indicators for evaluation of scientific output at the national level. This has helped in presenting the information on the comparative position of India with respect to the other highly performing countries. This is presented on the portal in sections namely, Research Outputs, Collaboration Patterns, Gender Distribution, Open Access, Social Media Visibility, Research grants (from both government and non-government agencies), Publications on Sustainable Development Goals (SDGs), and a comparison of indicators for Major Indian Institutions. These indicators provide a holistic picture of the national landscape in scientific research and development activity (see Table 1).

Institution level indicators

On a much more detailed level, the information about individual institutions is also provided on the portal. A part of this information may also be found on university ranking

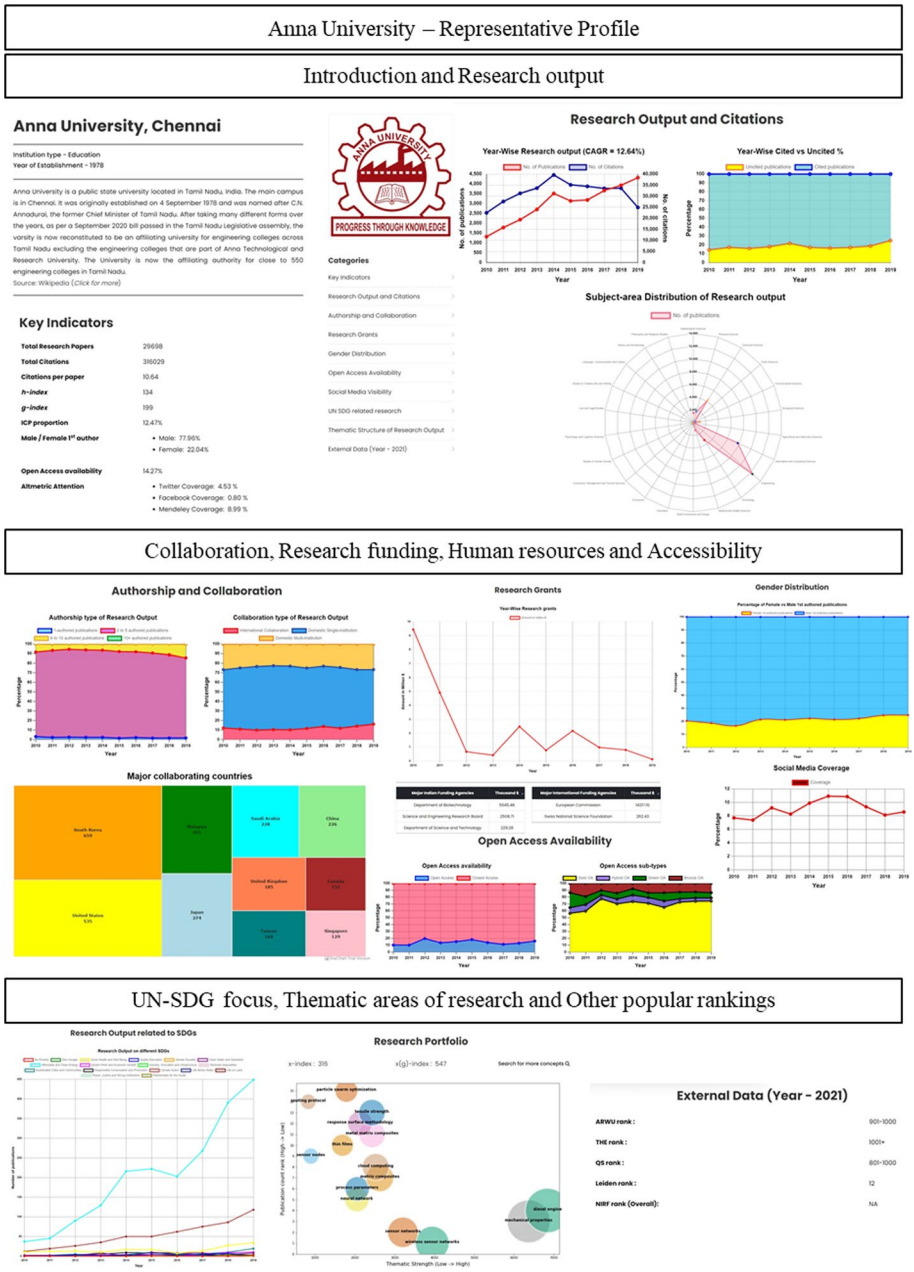

Fig. 2 Representative research portfolio of one Indian University (Anna University)

platforms as mentioned in the introduction section. However, neither they are as comprehensive nor they have this scale of coverage. For instance, the QS world university rankings and the university rankings for 2022 feature only 22 and 35 Indian universities and institutions, respectively, in their top 1000 list. By focusing only on Indian universities and

research institutions, the Indian Science Reports portal provides a comprehensive list of indicators to evaluate the performance and activity of these institutions. It provides details about *Organisational Profiles* (The platform collates information from various sources to provide a comprehensive profile of the institutions), and *Standard scientometric indicators* (such as CAGR and h-type indices are computed as per their standard definitions). It also uses x-index and x(g) index, as per the idea proposed in Lathabai et al. (2021a, b), to develop *Research Portfolios* i.e., thematic competence mapping for institutions (Fig. 2).

Some examples of useful indicators and analysis included

Publications and citations

India's overall research output, its growth over time, research output rank and global share, the citations to Indian research output, relative citation ratio and India's share in highly cited research output of the world is provided. It can be seen that India has a high CAGR value of 9.46%, which is lesser than only three countries—Russia (11.43%), Iran (10.56%) and China (9.50%). In terms of global share, India accounts for 2.95% of the total research output of the world during 2010–2019 period.

Comparison with other major countries

For a better understanding of India's research performance, the values of research output volume, CAGR, and global share of 20 major countries are compared, as shown in different tables in the portal. It can be seen that since the year 2015, India is the sixth largest producer of S&T research output and ranks at ninth place in total citations in the year 2019.

Subject area distribution of research output

Using a spider web chart, the number of publications in each (of the selected 22) subject area is represented to show the amount of research output of India in different subject areas.

International collaboration

The portal presents analysis of international collaboration patterns in Indian research output. India's international collaboration share has grown from 18.92% of total output in 2010 to 22.98% of total output in 2019. US, UK, Germany, China and South Korea are the top five collaborating countries.

Gender distribution

The portal includes an analysis of gender distribution of researchers producing research output in India. Overall, 29.30% of Indian research output is female first authored. However, subject area-wise variations exist in the proportion.

Open access

The availability of scientific research outcomes varies greatly across the world with journals providing free of charge open access to published articles and other journals placing access charge for each article. Across the world, the governments have started to promote open access of research outcomes and are funding research organisation and publication agencies for developing appropriate platforms. The platform provides number of open access publications by researchers at an overall as well as at the level of the institutions, and even presents the breakdown into the various open access sub-types. Subject area-wise variations in open access and also the open availability of public funded research output is analysed.

Altmetric attention

The role of social media has rapidly become important with platforms such as Twitter, LinkedIn, Research Gate etc. assuming significant role in communication and information exchange between the researchers. This has resulted in development of Altmetric (short for Alternative Metrics), a field which specifically focuses on the sharing and visibility of published research on social media platforms. The Indian Science Reports portal provides information about the Altmetrics for whole of the Indian research output as well as the altmetric attention obtained by individual institutions.

Research grants

The total volume of research grant received by Indian institutions from national and international agencies is analysed. Major funding agencies are identified and subject area-wise distribution of grants is presented. Finally, publications out of grants are identified and listed.

SDG related research

Indian research publications on sustainable development goals are presented. The major focus SDGs are identified and the subject area composition of research output on different SDGs is analysed.

Institutional reports

A comprehensive analytical report is generated for each of the 1000 institutions initially included in the portal. These are very elaborate and systematically organized reports on various parameters of research output and performance of Indian institutions. Figure 2 presents one such representative research report.

Concept-based search

The portal allows the user to search for institutions having research expertise in a given concept theme. A bubble chart for the top 15 major research concepts of each of the institutions is provided in the Institutional Reports section.

Conclusion

The main objective of the portal is providing an overview of the scientometric indicators for research output from India as well as individual Indian Institutions. In addition, it also provides useful insights into some aspects such as international collaboration for publication, comparative output for India versus selected countries. The platform can be a useful first step towards the development of a comprehensive listing of Indian institutions and their research activity as no such listing is yet available. The framework developed for identification of core competencies and expertise indices for institutions is the next step towards improving the coverage of this platform.

The developed portal is useful for policy makers, scientometric researchers, researchers looking for potential collaborators, data enthusiasts and students among others. It provides analytical information on STI data both at broad national level as well as on the granular level of institutions, universities, companies and hospitals. Hence, it can be used by a very wide audience. As it follows a framework which includes conventional indicators for assessing STI performance within a field, as well as reflect the thematic strengths of an institution it can be used to: (i) select institutions for funding in a specific thematic area, so as to eventually develop these as centres of excellence, and (ii) identify top performers in a given thematic area, which can help in several science policy related decisions. The framework can be used to determine the core competency of an institution in a given subject, determine its thematic research strength in different themes of that subject, its focus on sustainable development goals (SDGs), productivity in selected time period and international connections for STI activities. Combined together this level of information can be very useful for *ex-ante* performance-based funding in different thrust areas. We are further improving this part by implementing clustering of concepts into higher-level themes for better navigation of institutions by thematic areas.

The data used for analysis is obtained from multiple sources containing metadata about scientific research. Since the primary source of publication metadata is Dimensions, the analysis is limited to institutions with publications listed on the database. As the portal uses only selected set of indicators, in its present state it is an academic exercise which is limited to selected institutions. It may be used with other well established data sources to make more informed and valid assessments of research performances.

Acknowledgements The authors acknowledge the enabling support provided by the DST-NSTMIS funded project 'Design of a Computational Framework for Discipline-wise and Thematic Mapping of Research Performance of Indian Higher Education Institutions (HEIs)', bearing Grant No. DST/NSTMIS/05/04/2019-20, for this work. The authors would also like to acknowledge the access to Dimensions and Altmetric.com databases provided by Digital Science.

References

Bornmann, L., Stefaner, M., Anegón, F. D. M., & Mutz, R. (2015). Ranking and mapping of universities and research-focused institutions worldwide: The third release of excellencemapping.net. *COLLNET Journal of Scientometrics and Information Management, 9*(1), 65–72. https://doi.org/10.1080/09737766.2015.1027090

Khokhlov, A. N. (2020). How scientometrics became the most important science for researchers of all specialties. *Moscow University Biological Sciences Bulletin, 75*, 159–163. https://doi.org/10.3103/S0096392520040057

Lathabai, H. H., Nandy, A., & Singh, V. K. (2021a). Expertise-based institutional collaboration recommendation in different thematic areas. *In proceedings of the 11th International Workshop on Bibliometric-enhanced Information Retrieval co-located with 43rd. European Conference on Information Retrieval* (ECIR 2021a), Lucca, Italy.

Lathabai, H. H., Nandy, A., & Singh, V. K. (2021b). X-index: Identifying core competency and thematic research strengths of institutions using an NLP and network based ranking framework. *Scientometrics, 126*(12), 9557–9583. https://doi.org/10.1007/s11192-021-04188-3

Mukherjee, B. (2019). Ranking Indian universities through research and professional practices of National Institutional Ranking Framework (NIRF): a case study of selected central universities in India. *Journal of Indian library Association*, 52(4).

Prathap, G. (2017). Danger of a single score: NIRF ranking of colleges. *Current Science, 113*(4), 550–553.

Prathap, G. (2018). Profiling India's research impact in mathematics. *Current Science, 114*(4), 711–712.

Sooryamoorthy, R. (2020). Scientometrics for the study of sociology. *International Sociology, 35*(5), 461–479. https://doi.org/10.1177/0268580920957911

Vinkler, P. (2010). *The evaluation of research by scientometric indicators*. Chandos Publishing.