

India - Open science country note

Open science and the national context

The National Data Sharing and Accessibility Policy of the Indian Government places emphasis on the sharing of government data as well as data from scientific and research and development institutions (www.dst.gov.in/nsdi.html [1]). This policy initiative has led to the formation of a government portal (<http://data.gov.in> [2]) that facilitates open access data sharing.

The National Innovation Council (NInC) (www.innovationcouncil.gov.in/ [3]) was set up to discuss, analyse and help implement strategies for inclusive innovation in India, and prepare a roadmap for innovation 2010-20. NInC was intended as a first step toward creating a cross-cutting system to provide mutually reinforcing policies, recommendations and methodologies to implement and accelerate innovation performance in the country.

Open science research and innovation actors

There are several stakeholders that contribute substantially to the initiative.

1) Research councils and funding agencies

The Indian Council of Agricultural Research (ICAR, established 1929) (www.icar.org.in/ [4]) is an autonomous organisation under the Department of Agricultural Research and Education (DARE) of the Ministry of Agriculture. With [99 ICAR institutes](#) [5] and 53 agricultural universities throughout the country, this is one of the largest national agricultural systems in the world. The Council supports a host of databases – including state-specific databases focused on agricultural needs – and has elaborated State-specific Technological Interventions for Higher Agricultural Growth (www.icar.org.in/files/state-specific/contents.htm [6]). Each ICAR institute is mandated to set up an open access institutional repository (<http://icar.org.in/en/node/6609> [7]). ICAR is to set up a central harvester for the metadata and full text of all records from all OA repositories of ICAR institutes, for one-stop access to all agricultural knowledge generated in ICAR. Metadata and resources are licensed for use, reuse and sharing for academic and research purposes. Commercial and other reuse requires written permission.

The Council of Scientific and Industrial Research (CSIR, est. 1942) (<http://csir.res.in> [8]) now comprises a chain of about 40 laboratories and 50 field stations across the country in different areas of science and technology. Each laboratory generates a huge amount of data and has exhaustive data repositories, partially or fully in the public domain.

The Indian Council of Medical Research (ICMR, est. 1949) (<http://icmr.nic.in/> [9]) is a body in India for the formulation, co-ordination and promotion of biomedical research. It is also one of the oldest medical research bodies in the world, with [21 mission-oriented national institutes](#) [10], 6 [regional medical research centres](#) [11], and 5 units. The Council promotes biomedical research in the country through intramural as well as extramural research, generating mammoth amounts of research data.

The Bio-Informatics Centre (BIC) of ICMR has developed a Management Information System that captures information about extramural research activities of the Indian Council of Medical Research (<http://bmi.icmr.org.in/bic/index.php/services-provided-by-icmr/management-information-system-mis> [12]).

The Department of Biotechnology (DBT, est. 1986) (<http://dbtindia.nic.in/> [13]), under the Ministry of Science and Technology, provided fresh impetus to the development of modern biology and biotechnology in India. Several data centres marking different content streams have evolved with

DBT support.

The Microbial Culture Collection (MCC, est. 2009) (www.nccs.res.in/ [14]) has a broad charter to preserve, characterise and authenticate microbial resources. The MCC is affiliated with the World Federation for Culture Collections, and registered with the World Data Centre for Microorganisms. An international depository, the MCC supplies authentic microbial cultures and provides related services to the scientific community.

The Silico Gene Bank for Adaptation to Abiotic Stresses (<http://igbaas-ifgtb.icfre.gov.in/> [15]) is a database under DBT that consists of over 2 500 gene sequences. The Gene Bank is a one-point resource for bioinformatics analysis, used by plant researchers working on candidate gene-based molecular markers and functional genomics.

The Department of Science and Technology (DST, est. 1971) (<http://dst.gov.in/> [16]) is a nodal department for organising, co-ordinating and promoting S&T activities in the country and undertaking or financially sponsoring scientific and technological surveys, as well as research design and development where necessary. The DST has funded numerous projects in mission mode, activity that resulted in organised scientific resources and databases for supporting scholarly activities.

The National Spatial Data Infrastructure (NSDI) was launched by the DST to provide the national infrastructure for the availability of and access to organised spatial data, using a standard structured database.

2) Ministries for higher education and research

The Ministry of Human Resource Development (MHRD) (<http://mhrd.gov.in/> [17]) supports and [strengthens science-based higher education and research in the](#) [18] country. The ministry, through the institutions and organisations under it, has funded open access repositories and open data projects. Examples include the ICAR repository, CSIR repositories, and the Shodhganga repository of the Information and Library Network (INFLIBNET).

3) Higher education institutions

Almost all the major institutions of higher learning have begun collecting and curating their research data and output. Many universities have begun digitising their theses and dissertations. Shodhganga (see next paragraph) is now collecting all such digitised theses from many Indian universities and centres of higher learning.

Shodhganga (<http://shodhganga.inflibnet.ac.in/> [19]), funded by the MHRD, is a reservoir of Indian theses available online through centrally maintained digital repositories. Shodhganga not only ensures easy access to and archiving of Indian doctoral theses, but also will help raise the standard and quality of research.

The Open Index Initiative (<http://oii.igidr.ac.in/> [20]) of the Indira Gandhi Institute of Development Research (www.igidr.ac.in/ [21]) is an endeavour to develop an online, searchable bibliographic database for most of the Indian social science literature and resources available in Indian libraries.

4) Public research organisations and other government labs

The primary actors among public research organisations are the mission-oriented science agencies. The eleven major public research organisations of India are the Council of Scientific and Industrial Research; Department of Atomic Energy; Department of Science and Technology; Department of Biotechnology; Department of Electronics; Department of Space; Defence Research and Development Organisation; Indian Council of Agricultural Research; Indian Council of Medical Research; Department of Non-Conventional Sources of Energy; and Department of Ocean Development.

Some 300 national laboratories are housed under these leading public research organisations. There

are in-house R&D laboratories in large public sector enterprises in, for example, steel, fertilisers, railways, power, transport and aviation, chemicals, petroleum and energy. A massive network of public research organisations and government labs is mandated to carry out research in various S&T fields; these generate voluminous data, partially or fully in the public domain.

ENVIS (Environmental Information System) (envis.nic.in/), implemented by the Ministry of Environment and Forests, is a programme for environmental information dissemination among decision makers, scientists and other stakeholders. The ENVIS scheme extends its reach to involve institutions/organisations in state governments, academia, the corporate sector and NGOs, with 85 network partners.

The [Foundation for Revitalisation of Local Health Traditions \(FRLHT\)](http://envis.frlht.org/indian-medicinal-plants-database.php) [22] (<http://envis.frlht.org/indian-medicinal-plants-database.php> [23]) (est. 1991) is the “ENVIS Centre on Medicinal Plants” and its database, which stores 7 637 botanical names (6 198 medicinal plant species) with 119 183 vernacular names from 12 languages across India. There are nearly more than 2 000 plant images of traded medicinal plants, and the database houses 960 traded Indian medicinal plant species that are the source of 1 289 botanical raw drugs in trade in the country. The Digital Atlas section (<http://envis.frlht.org/digital-atlas-main.php> [24]) has 1 877 geo-distribution maps, corresponding to that number of species that are downloadable.

The Zoological Survey of India (ZSI, est. 1916) (www.zsi.gov.in/ [25]) collects and documents baseline data on the taxonomy, distribution, bio-ecology, etc. of various animal groups. ZSI’s National Zoological Collections (www.zsi.gov.in/achievements/national_holdings.html [26]) presently holds nearly 1 million specimens of about 70 000 species. The fauna of the Indian region represents nearly 8% of the world’s fauna, and approximately 90 000 recorded species along with 15 000 type specimens.

The Botanical Survey of India (BSI, est. 1890) (http://bsienvis.nic.in/Database/Type_Specimens_BSI_16547.aspx [27]) hosts the Central National Herbarium, comprising 2 500 000 items. The data are available free and can be downloaded after free registration.

The Geological Survey of India (GSI, est. 1851) (www.portal.gsi.gov.in/ [28]), under the Ministry of Mines, conducts geological surveys and studies. The geo-information furnished by the Survey takes the form of maps, geo-reports, and meta-databases that include three repositories of fossil, meteorites, rock and minerals. An active participant in international geo-scientific forums, the GSI has data sets covering over 150 years.

The Water Resources Information System of India (WRIS, est. 2008) (www.india-wris.nrsc.gov.in/ [29]): India-WRIS WebGIS, the joint venture of the Ministry of Water Resources and the Indian Space Research Organization (ISRO) within the Department of Space, is dedicated to managing water resources and providing the foundation for modelling and the Spatial Decision Support System (SDSS). India-WRIS WebGIS aims to be a “single window” solution for comprehensive, authoritative and consistent data/information concerning India’s water resources and allied natural resources in a standardised national GIS (Geographic Information Systems Resource) framework.

The National Data Centre (NDC) (www.imdpune.gov.in/research/ndc/ndc_index.html/ [30]) is the sole custodian of all meteorological data collected from various parts of India. The available data cover over 125 years. The mandate is to preserve quality-controlled data for weather prediction, aviation, agriculture, environmental studies, oceanography and shipping, as well as for researchers at various institutions and universities.

The National Remote Sensing Centre (NRSC) (www.nrsc.gov.in/index.html [31]), under the Department of Space, is the focal point for the distribution of remote sensing satellite data products in India and its neighbouring countries. The NRSC disseminates over 50 000 satellite data products annually. The NRSC/ISRO open data and product archive allows the user to select, browse and download 219 data products (www.nrsc.gov.in/Data_Products_Services_Satellite.html [32]).

Bhuvan (<http://bhuvan.nrsc.gov.in/data/download/index.php#> [33]), the Geo-portal of the Indian Space Research Organisation, comprises: visualisation of multi-resolution satellite images; application-enabled platform for government data; crowdsourcing and enrichment of maps with people's inputs; disaster support services; and free downloading of domestic and international data, through National Open Electronic Data Archive (NOEDA)

The Atomic Minerals Directorate of Exploration and Research (AMD) (www.amd.gov.in/ [34]) is under the Department of Atomic Energy (<http://dae.nic.in> [35]), and produces mostly airborne geophysical data on the areas flown by AMD over the years. Data are available on request (www.amd.gov.in/work/amdc.htm [36]).

The World Data Centre for Geomagnetism (<http://wdciig.res.in/WebUI/Home.aspx> [37].) is the part of the [Indian Institute of Geomagnetism](#) [38], an autonomous research body under the Department of Science and Technology. It provides the online data services of a magnetic data repository comprising over 150 years' data.

The Ocean Data and Information System (INCOIS) (www.incois.gov.in/Incois/incois1024/index/index.jsp?res=1024 [39]) provides ocean information and advisory services. The INCOIS National Oceanographic Data Centre is the central repository for marine data in real time in the country, as well as data received from other countries.

The Institution of the Ministry of Health and Family Welfare was created to set and update standards for common drugs in the country for treating disease.

The Indian Pharmacopoeia Commission (IPC) (<http://ipc.nic.in/> [40]) is an autonomous institution of the Ministry of Health and Family Welfare, created to set and update standards for common drugs in the country for treating disease. Its database Indian Pharmacopoeia (IP) has a total of 2 548 monographs.

5) Other

Digital India (<http://india.gov.in/> [41]) - is a programme to transform India into a digitally empowered society and knowledge economy.

The Tata Energy Research Institute www.teriin.org/ [42] (est. 1974) is a non-profit, scientific and policy research organisation, specialising in the fields of energy, the environment and a range of sustainable development options.

National Data Bank - The Ministry of Statistics and Programme Implementation (MoSPI) (<http://mospi.nic.in/> [43]), through its highly interactive and informative National Data Bank of socio-economic and religious categories (http://mospi.gov.in/national_data_bank/index.htm [44]), provides demographic databases in a single window.

Policy design - Open data

- The National Data Sharing and Accessibility Policy (NDSAP) sets forth the government's objective of facilitating access to government-owned shareable data. The data are to be presented in machine-readable format and made available across the country in a proactive manner and updated periodically, within the framework of various related policies, acts and rules of the government of India.
- To facilitate the contributions from various government departments, the National Informatics Centre (NIC) (www.nic.in [45]) has formulated discrete guidelines for implementing the open data initiative for each department. According to the guidelines, each department is to nominate a senior-level person to be data controller/data manager. Each

department is also advised to set up an open data/NDSAP cell with skilled persons from different disciplines, such as statisticians, economists, data analysts, IT professionals & subject specialists. NDSAP policy also advocates having departments spend a small percentage of their budget on the open data initiative.

- Also in compliance with the NDSAP, made effective by the government of India in March 2012, the National Informatics Centre has set up a data portal at <http://data.gov.in> [2]. A number of government departments are participating by contributing their datasets: today there are more than 10 000 datasets published in open format, free to use, reuse and redistribute. These datasets are distributed across 3 220 catalogues; they have been viewed 1.58 million times and downloaded 623 457 times by people around the world with diverse demographic and professional profiles. Implementation of the policy is monitored by a multidisciplinary taskforce set up by National Informatics Centre, which has members from government, academia, industry and civil society.
- The National Repository of Open Education Resources (NROER) (<http://india.gov.in/national-repository-open-educational-resources-ministry-human-resource-development> [46]), launched in 2013, aims to offer resources for all school subjects and grades in multiple languages. The resources are available in the form of concept maps, videos, audio clips, talking books, multimedia, learning objects, photographs, diagrams, charts, articles, and textbooks.

Policy design- Open/increasing access to scientific publications

One example of a repository is Shodhganga: A Reservoir of Indian Theses (<http://shodhganga.inflibnet.ac.in/> [19]). This is a federal, government-funded project; its budget is generally calculated on the basis of the number of institutions likely to be covered per annum, coupled with the number of theses from the respective institutions. On an average, INR 700 000 is spent per university.

The institution involved with the project is the Information and Library Network (INFLIBNET) (www.inflibnet.ac.in/ [47]) an Autonomous Inter-University Centre (IUC) of the University Grants Commission (UGC). The sectors targeted are research and academia; the expected outcome is online availability of electronic theses through centrally maintained digital repositories. Shodhganga should not only ensure easy access to and archiving of Indian doctoral theses, but also help raise the standard and quality of research.

As to the legal framework for open access with regard to copyright, the INFLIBNET Centre obtains the non-exclusive right to content by signing a memorandum of understanding (MoU) with the institution concerned. At the national level (<http://copyright.gov.in/> [48]), in 2012 the government amended the Copyright Act (under s.21), which deals with the right of authors to relinquish copyright. This places creative commons, the GNU General Public License, and other open licensing models on a much sound footing in India.

Data/evaluation results of open access initiatives:

1) The empirical studies have shown that such initiatives have made authors aware of the value of visibility, and also have motivated institutions in different ways to promote OA culture at the institutional level.

2) Several initiatives were taken by various scientific and technological departments and institutions in the government, particularly with regard to open access to scientific literature published in journals.

3) The Indian National Science Academy (INSA)'s signing of the Berlin Declaration on OA placed all academy journals as well as all CSIR journals in the public domain for free and unrestricted/unlimited access. It also led to the creation of OA repositories of multi-variate data content and streams. Journals and databases that are now open access include, for example: INSA journals online

(<http://insa.nic.in/INSAuth/Default.aspx> [49]); journals from the Indian Academy of Sciences (www.ias.ac.in/pubs/journals/ [50]); CSIR's NISCAIR (National Institute of Science Communication and Information Resources) Online Periodical Repository (NOPR) (<http://nopr.niscair.res.in/> [51]); IndMED, the bibliographic database of peer-reviewed Indian biomedical literature (medind.nic.in/imvw/), as well as the national database of medical journals, covering the full text of Indian biomedical periodicals (<http://medind.nic.in/medindcf/medinda.shtml> [52]), both hosted by National Informatics Centre.

Shodhganga is subject to the University Grants Commission (UGC)'s Notification (Minimum Standards & Procedure for Award of M.Phil./Ph.D. Degrees, Regulation, 2009) dated 1 June 2009, which mandates submission of the electronic versions of theses and dissertations by the researchers in universities. The Regulation aims to facilitate the worldwide academic community's open access to Indian theses and dissertations.

The project does not differentiate between free and libre for the models discussed above; however, the print equivalents of the online journals mentioned are subject to a subscription fee.

Skills for open science and open data

Several workshops dealing with open science, skill development, and tools and technology training and awareness have been and are being held in India. To mention a few:

The National Informatics Centre (NIC) has hosted several workshops on open data in the government sector. For example:

- 1) Workshop on Inclusive Web Programming – Programming on the Web with Open Data for Societal Applications, 31 May-4 June 2014 in Hyderabad.
- 2) Workshop on Open Government Data & Civic Engagement, 4 April 2013 in New Delhi.

NIC also held an “Opendataapps” Challenge, inviting challenging applications that use the data sets of open government data (OGD).

Other programmes:

- Workshop on “Open Data in Research”, conducted by INSA as a precursor to the Committee on Data for Science and Technology (CODATA) and World Data System (WDS) Conference, on 19 April 2014. The proceedings of this workshop are published in *Current Science*.
- OSOD – The International Workshop on Open Science and Open Data (<http://drtc.isibang.ac.in/osod/> [53]), hosted by the Documentation Research and Training Centre of the Indian Statistical Institute, Bangalore, 7 October 2014.

Datameet (www.datameet.org [54]) aims to provide a platform for different communities to work on standards and tools that will facilitate the visibility and reuse of data. It also conducts skill development workshops periodically.

Open science and international co-operation

The major international collaboration initiative is OSDD the Open Source Drug Discovery (OSDD) (www.osdd.net/ [55]), a CSIR-led consortium and global partnership. The initiative's vision is to provide affordable healthcare to the developing world by furnishing a global platform where the best minds can collaborate and collectively endeavour to solve the complex problems associated with discovering novel therapies for neglected tropical diseases such as tuberculosis, malaria and

leishmaniasis.

The Open Government Platform (OGPL) (<http://ogpl.gov.in> [56]) was developed in collaboration between

the Indian Government (National Informatics Centre, NIC) and the United States Government (General Services Administration, GSA). OGPL is an open source-based platform to enable any government agency, research institution or local government to establish an open data platform for their community.

Other information

The National Knowledge Commission (NKC), a high-level advisory body to the Prime Minister of India, was set up with the objective of transforming India into a knowledge society. Around 300 recommendations for 27 focus areas were submitted by the Commission, and the government of India has been committed to implementing those recommendations (www.knowledgecommission.gov.in/ [57]). Action has been taken on several fronts; two recommendations that are highly relevant here are the introduction of the Science and Engineering Research Board Bill 2008 in Parliament; and the launch of National Mission on Education through Information and Communication Technology (ICT), to leverage the potential of ICT in the teaching and learning processes. The aim of the latter is to enhance the gross enrolment ratio (GER) in higher education by 5 percentage points by the end of the XI Plan. Under this Mission, 20 000 institutions of higher education and nearly 10 000 university departments were earmarked for connectivity, beginning with a minimum of 5 Mbps for each. The central government was expected to bear 75% of the connectivity charges for five years.

The National Knowledge Network (NKN) (www.nkn.in/news.php/ [58]) is a state-of-the-art multi-gigabit pan-India network that provides a unified high-speed network backbone for all knowledge-related institutions in the country. NKN was set up with the purpose of building quality institutions with the requisite research facilities, and creating a pool of highly trained professionals. NKN endeavours to enable scientists, researchers and students from different backgrounds and diverse geographies to work closely to advance societal development in critical and emerging areas.

The issues of collecting and consolidating data are being deliberated at various forums; these focus on concerns about the reuse of data sets by multidisciplinary groups. Naturally, it is felt that more concerted efforts are needed to concentrate on metadata issues at the time of data collection itself, so as to address reuse of data by multivariate groups.

There are a growing number of Scientific Interest Groups (SIGs) that are involved in the creation and curation of scientific data. At national level however, where there are substantial data pools, serious attention must be paid to the management aspect. There is thus a need for a more comprehensive data policy that addresses issues holistically and not just the accessibility and sharing of data. Some activities in this direction are being planned (though at an elementary level).

The NDSAP was instrumental in setting up the Open Data Portal (<http://data.gov.in/> [2]) – a major initiative by the government of India. The government is also endeavouring to boost open data initiatives through promotional and motivational mechanisms, such as instituting the Open Data Champion Awards, christened the “Web Ratna Awards” (<http://webratna.india.gov.in/> [59]).

Links

- [1] <http://www.dst.gov.in/nsdi.html>
- [2] <http://data.gov.in/>
- [3] <http://www.innovationcouncil.gov.in/>
- [4] <http://www.icar.org.in/>
- [5] <http://www.icar.org.in/en/node/325>
- [6] <http://www.icar.org.in/files/state-specific/contents.htm>
- [7] <http://icar.org.in/en/node/6609>
- [8] <http://csir.res.in/>
- [9] <http://icmr.nic.in/>
- [10] <http://icmr.nic.in/institute.htm>
- [11] <http://icmr.nic.in/institute.htm#Regional%20Medical%20Research%20Centre>
- [12] <http://bmi.icmr.org.in/bic/index.php/services-provided-by-icmr/management-information-system-mis>
- [13] <http://dbtindia.nic.in/>
- [14] <http://www.nccs.res.in/>
- [15] <http://igbaas-ifgtb.icfre.gov.in/>
- [16] <http://dst.gov.in/>
- [17] <http://mhrd.gov.in/>
- [18] http://mhrd.gov.in/schemes_he_B
- [19] <http://shodhganga.inflibnet.ac.in/>
- [20] <http://oii.igidr.ac.in/>
- [21] <http://www.igidr.ac.in/>
- [22] <http://www.ihstuniversity.org/>
- [23] <http://envis.frlht.org/indian-medicinal-plants-database.php>
- [24] <http://envis.frlht.org/digital-atlas-main.php>
- [25] <http://www.zsi.gov.in/>
- [26] http://www.zsi.gov.in/achievements/national_holdings.html
- [27] http://bsienvis.nic.in/Database/Type_Specimens_BSI_16547.aspx
- [28] <http://www.portal.gsi.gov.in/>
- [29] <http://www.india-wris.nrsc.gov.in/>
- [30] http://www.imdpune.gov.in/research/ndc/ndc_index.html/
- [31] <http://www.nrsc.gov.in/index.html>
- [32] http://www.nrsc.gov.in/Data_Products_Services_Satellite.html
- [33] <http://bhuvan.nrsc.gov.in/data/download/index.php>
- [34] <http://www.amd.gov.in/>
- [35] <http://dae.nic.in/>
- [36] <http://www.amd.gov.in/work/amdc.htm>
- [37] <http://wdciig.res.in/WebUI/Home.aspx>
- [38] <http://iigm.res.in/>
- [39] <http://www.incois.gov.in/Incois/incois1024/index/index.jsp?res=1024>
- [40] <http://ipc.nic.in/>
- [41] <http://india.gov.in/>
- [42] <http://www.teriin.org/>
- [43] <http://mospi.nic.in/>
- [44] http://mospi.gov.in/national_data_bank/index.htm
- [45] <http://www.nic.in/>
- [46] <http://india.gov.in/national-repository-%20open-educational-resources-ministry-human-resource-development>
- [47] <http://www.inflibnet.ac.in/>
- [48] <http://copyright.gov.in/>
- [49] <http://insa.nic.in/INSAuth/Default.aspx>
- [50] <http://www.ias.ac.in/pubs/journals/>
- [51] <http://nopr.niscair.res.in/>
- [52] <http://medind.nic.in/medindcf/medinda.shtml>
- [53] <http://drtc.isibang.ac.in/osod/>
- [54] <http://www.datameet.org>
- [55] <http://www.osdd.net/>
- [56] <http://ogpl.gov.in/>

[57] <http://www.knowledgecommission.gov.in/>

[58] <http://www.nkn.in/news.php/>

[59] <http://webratna.india.gov.in/>