O.P. JINDAL GLOBAL UNIVERSITY
INSTITUTION OF EMINENCE DEEMED TO BE UNIVERSITY
A Private University Promoting Public Service

AI IMPACT SUMMIT भारत 2026 INDIA
सर्वजन हिताय | सर्वजन सुखाय
WELFARE FOR ALL | HAPPINESS OF ALL

CYRIL SHROFF CENTRE FOR AI, LAW & REGULATION
O.P. JINDAL GLOBAL UNIVERSITY

Policy Brief
Exploring a Regulatory Framework for Open Data in India

India AI Impact Summit 2026

Cyril Shroff Centre for AI, Law and Regulation
O.P. Jindal Global University

I Background

Algorithmic breakthroughs, advances in computing architectures, and access to large volumes of high-quality data are the three key determinants of progress for AI-enabled systems. Amongst these determinants, the data itself is a critical input as it shapes the other two determinants. The performance of existing AI models can also be augmented for broader applications using improved access to data. These applications include healthcare, urban governance, environmental management, and several other fields of public importance.

However, this growing dependence on data is often met with data scarcity in practice due to various reasons. These reasons include the data remaining inaccessible due to being segmented in silos, lacking standardisation, reflecting poor accuracy and quality, or being confronted with regulatory constraints. In India, public sector data represents one of the most significant untapped resources, given its scale. Yet, across different government departments and the centre-state jurisdictional segregation, this data is often fragmented across agencies and governed by inconsistent policies due to regulatory uncertainty and risk aversion. These challenges are impacting the trajectory of AI development and are preventing the full scale of benefits of AI from flowing to all stakeholders in India.

This problem is not unique to India. Governments worldwide are evaluating the role of open data. But India occupies a unique position in this evolving landscape. As one of the world's largest producers of public data, there is an opportunity for India to be the leading voice from the Global South in AI governance. Initiatives such as the Open Government Data (OGD) Platform, AIKosh and e-Sankhyiki already reflect a sustained policy commitment to data sharing and reuse. Yet what India still lacks is a systematic approach to open data.

Since India's broader regulatory strategy aims to strike a balance between innovation and responsibility, the challenge in this regard is also how open data should be governed. This may require navigating complex trade-offs. Balancing these considerations would require weighing openness in data against other public interests, including privacy, institutional accountability, data security, and equitable distribution of benefits arising from access to data. Lastly, private sector data is also relevant, as many high-value datasets for AI development and domain-specific applications are generated and held by private entities. However, enabling access to such data raises distinct difficulties, including privacy and consent constraints, protection of trade secrets and intellectual property rights, competition law issues, and cybersecurity risks.

In this policy brief, we first look at the policy framework in India that may impact open data sharing in conjunction with the specific challenges and opportunities in this context, substantiated by case studies. To make these more salient, we follow up with a comparative overview of other jurisdictions. Lastly, we conclude with general suggestions that emerge from the discussion in the brief.

II Indian Law and Open Data

India does not have any legislation specifically directed at ensuring access to or regulating open data. Therefore, in India, unlike jurisdictions such as the United States, the European Union, South Korea, and Japan, public authorities are under no legislative mandate to ensure availability, accessibility, and open use of their data. In 2019, the Ministry of Electronics and Information Technology (MeitY) constituted an eight-member committee to deliberate on the regulation of non-personal data in India.

The committee published a report in 2020, which recommended the creation of a national regulator for non-personal data and the enactment of legislation to implement the framework for free access to certain non-personal data. These recommendations have not been implemented. Similarly, unlike the European Union's efforts to enhance access to privately held non-personal data, there exist no obligations upon private players under Indian law to share access to their data with other enterprises, users, or the public at large.

Despite the lack of a dedicated legislative framework for open data in the public or private domain, India has made strides in open public data initiatives, starting with the National Data Sharing and Accessibility Policy (NDSAP) of 2012, which was aimed at enhancing access to non-sensitive data generated or collated using public funds. While the NDSAP was initially intended to enhance access to public data to enable better decision-making and transparency in government functioning, it has also proved useful in spurring innovation and training of AI. This has been made possible by certain steps that the government has taken under the Policy, such as the establishment of the OGD Platform. The OGD Platform has played a key role in making public data more accessible. It facilitates access to machine-readable data, which is crucial for AI training and innovation. This is further enhanced by the introduction of the government open data license which provides a "worldwide, royalty-free, non-exclusive license to use, adapt, publish (either in original, or in adapted and/or derivative forms), translate, display, add value, and create derivative works (including products and services), for all lawful commercial and non-commercial purposes".

III Challenges, Opportunities and Developments in India

Despite this framework being in place, there are significant challenges and opportunities for open data access in India. Here, we look at some of the specific challenges and opportunities, along with certain key developments, that highlight their salience.

A key challenge in India is that while platforms like OGD Platform and NDSAP have the data, they often suffer from issues of 'quality, completeness, interoperability and AI ready data'. The data on these platforms is uploaded in formats that are not suitable for Large Language Model training, as these models require datasets that are annotated, machine-

readable and high volume. Thus, there is a need for better data repositories and a regulatory framework governing the same.

There is also no dearth of opportunities. Since open data is increasingly viewed as a critical tool for addressing complex societal challenges (often termed 'wicked problems') through evidence-based policy-making, an appropriate open data regulatory framework could unlock innovative and out-of-the-box solutions to these problems.

Consider, for instance, the wicked problem of air pollution. The underlying data necessary to achieve this objective is present. If data from multiple sources, such as emissions inventories (e.g., from Central Pollution Control Board (CPCB) stations), meteorological data (from the India Meteorological Department (IMD)), industrial discharge information, and public health outcomes (from health ministries), were made openly available and linkable, it would enable integrated AI models aimed at reducing air pollution. These models may, in turn, link sources, predict hotspots, and evaluate interventions based on shared evidence. Currently, NCAP and CPCB/IMD data portals provide for real-time data, which is openly downloadable for monitoring. However, they still lack seamless standardised formats or APIs for AI models to analyse these vast datasets.

Work on AI-ready integration is accelerating, and several initiatives are pushing in that direction. An exemplary platform in this regard is AIKosh. It is India's first-of-its-kind centralised repository for AI-ready datasets, pre-trained models, and toolkits. It represents a significant step towards an open data sharing model and empowers several stakeholders, including startups, developers, researchers, businesses, and institutions, to build solutions tailored to Indian contexts and markets. It can facilitate innovation across several sectors, such as smart energy management and urban growth planning.

For instance, in smart energy, AI models trained on granular, anonymised data about energy consumption patterns (from smart meters), grid load, and renewable energy generation (from power distribution companies or DISCOMs) could predict demand fluctuations and optimise energy distribution, thereby enhancing grid stability. Many other Indian startups and tech companies are actively working in AI development, and access to qualitative standardised datasets is crucial for their growth. IIT Bombay has launched 16 culturally rooted datasets on AIKosh related to Indian languages, documents, audio, and drone imagery in order to facilitate innovation. This move is a bold step towards asserting India's AI sovereignty in a world dominated by western-centric datasets.

Further, Indian startups like Sarvam AI and Ola Krutrim, and institutions like IIT Bombay (BharatGen Consortium) and IIT Madras (AI4Bharat), are actively leveraging AIKosh. They are bypassing the issues of data silos, which is crucial for growth amid limitations of synthetic data. Further, government institutions such as the Ministry of Agriculture, with its Kisan Call Centre data, and Open Data Telangana are also providing real-world data that startups can use to train AI models.

IV Comparative Jurisprudence on Open Data

Globally, open government data frameworks sit on a spectrum. At one end are statute-based regimes, notably in the European Union, the United States, and Australia. At the other end are approaches driven mainly by policy instruments and standard licences, such as in Singapore and the United Kingdom. Across jurisdictions, an 'open by default' principle is common, but it is almost always qualified by exceptions for privacy, security, and confidentiality.

The European Union has one of the most detailed legal frameworks for the reuse of public-sector information set out in the Open Data Directive (EU) 2019/1024. The Directive treats public-sector data as a resource intended for reuse and introduces the concept of 'High Value Datasets' (HVDs). These are datasets identified for their significant socio-economic value and they span categories such as geospatial, earth observation, environment, meteorological, statistical, corporate, and mobility data. The Commission Implementing Regulation (EU) 2023/138 specifies that HVDs must be made available in machine-readable formats via APIs (and where indicated, as bulk downloads), while the Open Data Directive (EU) 2019/1024 establishes the general rule that HVDs are to be available free of charge, subject to limited statutory derogations. In practical terms, the EU model is not only about openness in principle, but it is also about enforceable access and usable data infrastructure.

In the United States, open data related provisions are found in federal legislation. The OPEN Government Data Act establishes an 'open by default' mandate and requires federal agencies to publish public data assets in machine-readable formats. In addition, they are expected to use standardised metadata.

Singapore, on the other hand, relies on a licence-based approach. The government offers the Singapore Open Data Licence, which grants rights to use government-owned datasets worldwide and in perpetuity. While these rights come with conditions of attribution and non-endorsement, they are free of royalty. The government also manages a centralised portal and provides extensive API support.

Lastly, the Australian model focuses on data sharing but in a controlled manner. The Data Availability and Transparency Act 2022 establishes the Data Availability and Transparency (DATA) Scheme. Under this, government data may be shared only with accredited users for defined purposes. Further, individual sharing decisions are based on the satisfaction of safety and governance principles established by the law. This approach is highly relevant to the development of AI models in those contexts where datasets are sensitive.

V Conclusion and Way Forward

Taken together, these approaches underline a clear lesson for India. Mature open data regimes increasingly treat data access as a form of public infrastructure, where usability, standardisation, and accountability matter as much as formal openness. On this basis, we make the following suggestions that may be considered and adopted in India:

- Adopting an 'open by default' stance with clearly defined exceptions;
- Prioritisation of a curated set of High Value Datasets for early compliance, thereby enabling phased implementation; and
- Pairing openness with enforceable requirements on machine readability, metadata, and API-based access to ensure practical usability of open data.

Concurrently, the Australian model demonstrates the efficacy of establishing a parallel pathway for controlled access to sensitive datasets. This suggests that a balanced Indian framework would benefit from a dual-track approach: open publication for low-risk public data on the one hand, and accredited, purpose-limited access mechanisms on the other, particularly where considerations of confidentiality, security, or data linkage risks render full openness inappropriate. Such a framework would position India to harness the transformative potential of open data for AI development while maintaining appropriate safeguards for sensitive information.

Building upon existing initiatives such as AIKosh and the OGD Platform, India has a strong foundation from which to develop a comprehensive open data regime that addresses current gaps in data quality, standardisation, and interoperability. By learning from comparative models while tailoring solutions to its unique federal structure and scale, India can emerge as a leading voice from the Global South in shaping the future of open data governance for AI.

O.P. JINDAL GLOBAL
INSTITUTION OF EMINENCE DEEMED TO BE
UNIVERSITY
A Private University Promoting Public Service

AI IMPACT SUMMIT
भारत 2026 INDIA
सर्वजन हिताय | सर्वजन सुखाय
WELFARE FOR ALL | HAPPINESS OF ALL

CYRIL SHROFF
CENTRE FOR AI,
LAW & REGULATION
O.P. JINDAL GLOBAL UNIVERSITY

O.P. Jindal Global University, Sonipat Narela Road,
Sonipat-131001 (Delhi NCR), Haryana, India

jindalglobaluni jindalglobaluni jindalglobaluni jindalglobaluni jguvideo

JGU - An Initiative of Jindal Foundation

WWW.JGU.EDU.IN