SFLC.IN
2nd Floor, K9
Birbal Road, K-Block
Jangpura Extension, Delhi – 110014
https://sflc.in | mail@sflc.in | +91-11-43587126

Comments on the Draft India Data Accessibility and Use Policy

Executive Summary

In India, the Government is the biggest repository of data. The Draft India Data Accessibility and Use Policy[1] ("**the Draft Policy**") introduces several provisions for the adoption of certain measures to allow for greater data-sharing amongst government bodies, businesses, and private parties. The stated objective of the Policy is to improve governance and service delivery in the data-driven economy in India.

This draft Policy was published by the Ministry of Electronics and Information Technology ("**MeitY**") keeping in view the absence of a consolidated policy on Open Government Data (or "**OGD**") in India.[2] There have been efforts in the past towards the promotion of OGD such as the National Data Sharing and Accessibility Policy ("**NDSAP**"), but the ambit of the NDSAP is narrower as compared to that of the Draft Policy which also proposes government-to-government (G2G) sharing of data.

The summary of our suggestions in the Draft Policy are as follows:

- ❑ The Draft Policy claims to promote OGD but in fact is in a conflict with such a framework as it proposes to price data on the basis of a 'data ownership' model
- ❑ Work created by the government like databases must be in the public domain as they are made by state employees using taxpayers' money, and must not be copyrighted
- ❑ Moreover, copyrighting databases is not a very strong way of protection, as modification in the original database which results in a new and better arrangement of the original database is not considered as copyright infringement
- ❑ The model of ownership of data should be done away with and be replaced with data stewardship in line with the Non-personal data framework. The citizenry should be the owner of the data.
- ❑ The policy, as it stands, lacks a grievance redressal mechanism and must provide for a nodal person that may be contacted for issues related to re-identification of data principals etc.
- ❑ The lack of a data protection law disincentivises the government departments/agencies from keeping high anonymization standards
- ❑ The India Data Office has overlapping functions with the Data Protection Authority ("**DPA**") under the proposed Data Protection Bill
- ❑ In the absence of the Data Protection Act, the Government has an absolute overreach and discretion on the sharing of data with private entities.
- ❑ The classification of data into 'open', 'restricted', and 'non-shareable' confers too much executive control to the government
- ❑ The MeitY itself must specify the data retention timelines within the Draft Policy itself, instead of delegating the task to the various Ministries/Departments as it would lead to a lack of standardization

1 https://www.meity.gov.in/writereaddata/files/Draft%20India%20Data%20Accessibility%20and%20Use%20Policy_0.pdf.

2 Page 6 of the Background Note of the Draft Policy available at https://www.meity.gov.in/writereaddata/files/Background%20Note%20for%20India%20Data%20Accessibility%20and%20Use%20Policy.pdf.

Introduction

OGD has become an important governmental practice in an attempt to champion accessibility and accountability with respect to data processed by governments globally. 'Open data' refers to such data which anyone is free to use, reuse, and redistribute. Such data must be machine readable and be easily accessible.[3] Therefore, Open government data would refer to such open data that the government collects, processes, and generates in its day-to-day functioning.[4]

A large quantum of government data remains inaccessible to citizens, civil society, although most of such data may be non-sensitive in nature and could be used by the public for social, economic and developmental purposes. This data needs to be made available in an open format to facilitate use, reuse and redistribute, and should be free from any license or any other mechanism of control.[5]

Opening up government data in open formats would enhance transparency and accountability while encouraging public engagement. The government data in open formats has a huge potential for building various types of innovative apps and services around the published datasets.[6]

More than 250 governments across national, subnational and city levels, as well as entities such as the World Bank and United Nations have launched Open Data initiatives, while more are in the pipeline.[7]

India has also partaken in OGD policies. Starting with the NDSAP in 2012,[8] then the Government Open Data Use License that was published in 2017 under the NDSAP,[9] India has been experimenting with such policies for over a decade now. However, these policies are fraught with inconsistencies.

For instance, on one hand the Government has mentioned the importance of OGD and on the other hand it has proposed pricing datasets in order to use, access, or share them in the Draft Policy now and the NDSAP before. According to this policy, government works[10] such as datasets/databases that are 'owned' by the government can be accordingly priced and may be sold to both government departments/agencies and private individuals/entities.[11] Pricing of open data goes against the principles of open data as followed by most Governments across the world.

Issues around ownership of data

The Policy has proposed ownership rights in the datasets put in place by the Government. While a government work in the form of a database may be eligible for a copyright, this copyright only lasts for a period of 60 years.[12] After that the government work is released into the public domain.

However, copyrighting databases is not a very strong way of protection, as modification in the original database which results in a new and better arrangement of the original database will not be considered as copyright infringement.[13] Thus, anyone can modify the original database and not be held liable for infringement.

The concept of "ownership" could be done away with in the policy document, and could be aligned with the Non-Personal Data framework, which suggests that the data principal is the "beneficial owner" and the government is the data custodian/trustee responsible for the management of the data[14], and has a fiduciary role towards the data principal(s). Scholars such as Scofield thus suggest

---

3 Annexure - I, Draft India Data Accessibility and Use Policy, https://www.meity.gov.in/writereaddata/files/Draft%20India%20Data%20Accessibility%20and%20Use%20Policy_0.pdf.

4 Implementation Guidelines for National Data Sharing and Accessibility Policy (NDSAP), page 6, https://data.gov.in/sites/default/files/NDSAP_Implementation_Guidelines_2.2.pdf.

5 Id.

6 Id.

7 http://opendatatoolkit.worldbank.org/en/open-data-in-60-seconds.html.

8 https://data.gov.in/sites/default/files/NDSAP.pdf.

9 https://data.gov.in/sites/default/files/Gazette_Notification_OGDL.pdf.

10 As defined under section 2(k) of the Copyright Act, 1957.

11 https://www.meity.gov.in/writereaddata/files/Background%20Note%20for%20India%20Data%20Accessibility%20and%20Use%20Policy.pdf; https://www.meity.gov.in/writereaddata/files/Draft%20India%20Data%20Accessibility%20and%20Use%20Policy_0.pdf.

12 Section 28 of Copyright Act, 1957.

13 https://www.meity.gov.in/content/copyright.

replacing the term “ownership” with “stewardship”, as it better captures the responsibility that organisations are actually looking to promote.[15]

Scholars also suggest that a lot of stakeholders can claim data ownership – ranging from the creator i.e. the party that creates or generates data; the consumer i.e. the party that uses the data, the compiler i.e. the party that selects and compiles information from different information sources; the funder i.e. the user that commissions the data creation and therefore claims ownership; purchaser/licenser as owner – the individual or organisation that buys or licenses data may stake a claim to ownership, etc.[16]

Ownership of data is often perceived as a means of assuring data quality and curation, as well as data protection and security along the complete data life cycle. In this case, ownership of the data, it is argued, should be assigned without any copyright being granted to the “data owner”. Thus the “owner”, if any, should be the data principal(s) whose data is being collected, which in this case would be the citizens of India.[17]

Lack of provisions for a Grievance Redressal Mechanism

The major drawback of the policy is that there is no provision for a grievance redressal mechanism present either in the Policy document or in the Background Note. Even the Government Open Data License, for that matter, appoints an arbitrator through the Union Law Secretary for the resolution of any difference of opinion and/or dispute arising out of the license.[18]

If the database, for the sake of argument, is considered to be copyrightable, and is copyrighted – then, in that case, disputes may be referred to the Appellate Board constituted under section 11 of the Copyrights Act, 1957.[19] But even those disputes may be limited to the copyright itself, and would not be in the nature of a wider ambit that could potentially include data protection issues, for instance.

Therefore, for instance, individuals might have their identity known by re-identification if the anonymization is not secure enough, and in such cases, the aggrieved party needs to have a nodal person to contact and lodge their grievance(s).

Lack of a data protection law

Interestingly, the Government has proposed this Policy without any statutory backing, and without the Data Protection Act being passed and implemented yet.

In such a climate, the selling of data seems like a dubious thing to do – since there are no security safeguards put in place for anonymization, and the task has been delegated to the respective ministries.[20] This would leave room for states to decide the standards for themselves. The chances are that the Government will exempt itself from the provisions of the Data Protection Act when passed. Thus, there will be no checks and balances on government-to-government (or G2G) sharing of data.

High security standards demand better infrastructure, which would disincentivize government departments/agencies from having proper security standards in place for the protection of data. Thus, MeitY should have secure anonymization standards in the Policy.

The framework of the India Data Office (IDO)

---

14 Page 48 of the Revised Report by the Committee of Experts on Non-Personal Data Governance Framework available at https://static.mygov.in/rest/s3fs-public/mygov_159453381955063671.pdf; also *see* https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf.

15 http://www.dmreview.com/editorial/dmreview/print_action_articleId_296.html.

16 Loshin D (2002) Knowledge Integrity: Data Ownership.

17 Page 23 of the Revised Report by the Committee of Experts on Non-Personal Data Governance Framework available at https://static.mygov.in/rest/s3fs-public/mygov_159453381955063671.pdf.

18 https://data.gov.in/government-open-data-license-india.

19 Section 11 of the Copyrights Act, 1957, available at https://copyright.gov.in/documents/copyrightrules1957.pdf.

20 Available at https://www.meity.gov.in/writereaddata/files/Draft%20India%20Data%20Accessibility%20and%20Use%20Policy_0.pdf.

For the purposes of consolidating and sharing of non-personal data, the MeitY has proposed to set up an India Data Office ("**IDO**"). This sharing has been constituted to be across government and 'other stakeholders'.

The establishment of the IDO by MeitY means that the members of the body shall be appointed solely by the Executive. The Government will have sole discretion to appoint members, without any oversight. The manner of appointment to the IDO itself would attract arbitrariness in the absence of transparent and reasoned guidelines and protocols.

The IDO is also responsible for coordinating with ministries, states, and other schematic programs to identify and facilitate data access and sharing of non-personal datasets which are stored with the government bodies. Such vast amounts of data-sharing necessitates smooth coordination between the various Governmental Ministries and Departments, as also the State and Central Governments, as per Clause 10.2 of the Draft Policy.

It is worrying to note that Clause 6.6 of the Draft Policy provides that the access to non-personal datasets would be provided by the IDO to interested parties such as researchers, start-ups, enterprises, individuals, and government departments. The same shall be facilitated via data licensing, sharing, and valuation methods. This effectively means that data can be sold to the highest bidder, with no say of the user. For instance, under the Bulk Data Sharing Policy and Procedure,[21] the Government effectively admitted to selling bulk data of vehicle owners to private parties from its Vahan and Sarathi databases, earning Rs. 65 Crores in the process. A similar exercise is bound to be carried out under the Draft Policy as well, involving exponentially greater amounts.[22]

The roles and responsibilities of the IDO have not been clearly defined under the Draft Policy. It still remains to be ascertained whether the IDO will perform any grievance redressal mechanisms, or act as an oversight body simpliciter according to Clause 16.1. The precise nature of the IDO must be clarified in order to have a streamlined process of grievance redressal, and it remains to be seen whether its functions would overlap with those of the DPA.

The complaints made at that time are still being echoed today and run thus; In the absence of the enactment of a Data Protection Bill which governs data-sharing and access protocols, the Government has an absolute overreach and discretion on the sharing of data with private entities.

Inherent risks in the labeling of datasets as 'open', 'restricted' or 'non-shareable'

The Draft Policy asserts that data assets which are considered to be 'confidential' and/or 'in the interest of the country's security' will not be shareable by the Ministries/Departments. The Government would be required to analyze non-personal datasets in order to label them as confidential.

As part of its role and functions, the Government must clearly identify which datasets would be termed as confidential and the rationale behind the same. Effective security measures do not find mention in the Draft Policy; so it remains to be seen whether pseudonymization techniques which involve subjecting data to technical and organizational measures to ensure it cannot be linked back

21 Available at https://parivahan.gov.in/parivahan/sites/default/files/NOTIFICATION%26ADVISORY/8March%202019.pdf.

22 Available at https://inc42.com/buzz/indian-govt-is-selling-vehicle-owner-data-to-companies-and-citizens-dont-have-a-clue/.

to identifiable natural persons.[23] Several issues which have not been addressed by the arise for consideration:

- How will such datasets be identified and by whom?
- How will security concerns be dealt with in the process of such identification?
- Are pseudonymization measures likely to be employed to further anonymize datasets?

In the process of analysis, if the non-personal dataset can be linked back to a person, i.e., re-identification occurs, then there arises risk of clear data leakages occurring for personal datasets. Research has shown the likelihood of non-personal datasets being linked back to personally identifiable persons as highly likely using advanced techniques.[24]

For instance, a perusal of Clause 7 of the Draft Policy which pertains to Identification of Datasets reveals that every Government Ministry/ Department/ Organization will identify non-personal datasets as Open, Restricted or Non-Shareable. This implies that each of such organizations shall be analyzing and identifying the non-personal datasets in order to characterize them into one of the three categories. This raises several concerns on security measures, risk-mitigation and risk-analysis measures to safeguard such non-personal datasets from data breaches.

No relevant timeline provided for the data retention period of datasets

Since the Government will view public data from the commercial viewpoint to harness the public sector data and provide better services, this entails greater time-periods of data retention by the various Government agencies and departments.

-

It is glaringly discouraging to note under Clause 13 of the Draft Policy, it shall be the responsibility of ***each*** Ministry/Department to define its data retention period for the particular dataset it has stored. No time-period or estimate is provided. Therefore, data retention periods can be 30 days or 3 years, or even 30 years; it shall be the sole prerogative of the Ministry to decide. While Clause 13.2 makes mention of the DQGI framework as a model for basing data-retention policies by the Ministries, this acts merely as a suggestion with no binding effect. afterMinistries and Departments shall have the sole discretion to decide the data-retention periods for each dataset they contain.

For instance, non-personal datasets collected by Municipal Corporations or Public Utility Corporations contain vast amounts of data which when processed, would constitute as community non-personal data, available for sale to the different stakeholders mentioned under Clause 6.6 of the Draft Policy. In absence of any framework or time-line for the data-retention period, such entities will in effect, be allowed to store and share such data for excessively long periods of time.

There must be well-established policy guidelines providing a time-scale for specific categories of non-personal data to be followed by Central and State Government Departments.

The Principles of the Draft Policy make indication of the timeline to be observed for data retention by each Ministry. Therefore, the prerogative of the Government on retaining data looms large on every non-personal dataset that it possesses.

23Article 4(5), Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, or General Data Protection Regulation.

24Michèle Finck and Frank Pallas, (2020). *They who must not be identified – distinguishing personal from non-personal data under the GDPR.* Oxford University Press.

Concluding Remarks

While the Draft Policy is a step in the right direction, it needs to be more robust and privacy-preserving, and must not be implemented until there is a Data Protection Act in place. This would ensure that minimum levels of protection are there for the citizens. Further the government needs to implement such policies in tandem with its body of law in totality, and must harmonize the OGD policies with the Draft Policy and the Copyright Act, 1957 as well.

Moreover, the government should do away with the data ownership model and replace it with the data stewardship model, so as to ensure that the government does not monetize citizens' data with low security standards in place, and without any checks and balances.

While the concept of data ownership is full of inherent inconsistencies, if there must be an 'owner' in the truest sense it must be the citizens and not the State.

About SFLC.in

SFLC.in is a New Delhi based not-for-profit organization that brings together lawyers, policy analysts, technologists and students to protect freedom in the digital world. We promote innovation and open access to knowledge by helping developers make great Free and Open Source Software (FOSS), protect digital civil liberties by providing pro-bono legal advice, and help policymakers make informed and just decisions with the use and adoption of technology.

We hope that our submission proves to be useful. Please feel free to contact us for any clarification or further information.