Feedback to the Ministry of Electronics and Information Technology

On the Draft India Data Accessibility and Use Policy

Introduction

NASSCOM welcomes the opportunity to submit our feedback to the Ministry of Electronics and Information Technology **(MEITY)** on the Draft India Data Accessibility and Use Policy (**IDAUP**) and on the supporting Background Note (**Note**).[i]

With the right design and implementation strategy, the IDAUP can create significant value for governance, research, transparency, and innovation in India. We believe that its success will depend upon how the balance is struck between improving data access, enhancing data quality, facilitating data reuse, and mitigating risks to privacy or security. With the aim of striking this balance better, we offer below some suggestions to make the IDAUP more clear, workable, and scalable. Our submission covers the following:

1 Objectives and principles

Several objectives are outlined for this policy, including increasing access to open data[ii], streamlining data sharing within government,[iii] leveraging public sector data for governance and research,[iv] whilst ensuring privacy and legal certainty.[v] The IDAUP also outlines several data sharing and governance principles[vi].

Comments & Concerns

Selected objectives

By themselves, these objectives appear to be valuable pursuits. However, they are broadly worded and are not articulated as quantifiable outcomes against which future performance can be measured. We are also left guessing about the rationale behind their articulation, how they are to be prioritised *inter se* each other, and how they are met by the proposals that follow.

For example, the IDAUP outlines an objective of protecting the privacy and security of all citizens, but, in terms of concrete proposals, states that anonymisation standards set by MEITY or by the proposed India Data Office (IDO) or under any other law or policy must be complied with and that "*all data sharing shall happen within the legal framework of India … and recognised international guidelines*".[vii] India does not yet have a comprehensive personal data protection law and existing data privacy obligations in law are only applicable to body corporates.[viii] Given these legal lacunae, the lack of analysis as to why these proposals in the IDAUP will be sufficient to protect privacy and security is wanting.

Despite the clear emphasis on open data, the IDAUP does not actually enumerate objectives most traditionally associated with open data policies, such as the promotion of transparency and accountability in democratic performance and governance or the development of new private products and services using open government data.[ix] Though they have been hinted at, these are important objectives meriting more emphasis. After all, prior efforts to release open government data proactively in India have been driven from the starting point of promoting the public's right to information[x] as well as to promote innovation and economic growth.[xi] As an example of the innovation potential, in the United States, high-quality open government data about the real-estate market has been credited with powering the growth of property technology companies.[xii]

Data sharing and governance principles

The various principles in the IDAUP are difficult to interpret or relate to the objectives and proposals. For example, "*privacy & security by design", "risk management over risk avoidance",* or "*user-centred practices & systems"* could serve as valuable design principles. However, currently, there is no discussion on their meaning and how they are being met in the proposals under the IDAUP.

It is not clear why these principles are the ones to proceed with. For example, the Note does state that several international policies and papers were referenced to evaluate India's data ecosystem. Some of these contain useful principles, such as the idea of "*open by design and default"*[xiii] and the *FAIR* principles (from the Open Data Directive in the European Union[xiv]) or the principles of "*project, people, setting, data and outputs"* (from the Data Availability and

Transparency Bill in Australia[xv])[xvi]. However, these principles have not been considered in IDAUP.

For instance, the IDAUP should look to secure compatibility with the *FAIR* principles (Findability, Accessibility, Interoperability and Reuse)[xvii] in relation to research and scientific data, given their emphasis on enhancing machine readability and usability.[xviii]

Suggestions

- **It is likely that the various objectives and principles articulated have been selected after much deliberation and they have been factored into the thinking behind the proposals in the IDAUP. However, stakeholders would be benefitted if provided with some insight into that thinking. This would lend more credibility to this policy exercise.**
- **Alternatively, it may be the case that these are for the proposed institutional framework to factor into their functioning; if that is the case, then the IDAUP would be benefitted by a more meaningful discussion of their intent and elements, so that subsequent standards, policies, protocols, and licenses developed under this policy can be evaluated against more measurable concepts.**
- **Additional objectives of high relevance, such as supporting transparency and public oversight of government functioning or innovation, should be incorporated.**
- **A comparison against international policies may be included in the Note to demonstrate how the data sharing principles in the IDAUP stack up against global standards.**

2 Material scope

The IDAUP states that it will apply to all "*non-personal data and information*" that is "*created/generated/collected/archived by the Government of India directly or through authorized agencies by various Ministries/Departments/Organisations/Agencies and Autonomous bodies*". In this submission, we refer to these various ministries, departments, organisations, agencies and bodies as "*covered public entities*".

Comments & Concerns

The term "non-personal data" (**NPD**) is not defined in the IDAUP or the Background Note or in any other law or policy that is currently in force. It is only defined in government reports[xix] or draft legislation.[xx] Notably, that definition is itself predicated upon a well-defined conception of "personal data", which is also not defined in the IDAUP.

Similarly, the term "information" has not been defined. If we go according to existing definitions of information found elsewhere – such as in the Information Technology Act of 2000[xxi] or in the National Data Sharing and Accessibility Policy of 2012 (**NDSAP**)[xxii] – then the scope of this policy practically extends to all data, since the term "information" is an all-inclusive term that captures all data and does not differentiate between NPD and personal data.

The IDAUP does not differentiate between different subsets of NPD – such as data that were never personal[xxiii] and anonymous data[xxiv] - or between different data collection mechanisms that covered public entities[xxv] may employ to collect NPD.[xxvi] It merely states that it will apply to all NPD that is "*created/generated/collected/archived*", which captures a wide range of interactions between covered public entities and external stakeholders involving NPD. Currently, due to this, it is difficult to determine the precise scope of the IDAUP and to what datasets it may apply to. This makes it difficult to determine how potential challenges that may arise with existing laws will be addressed.

For example, there is limited discussion on whether the IDAUP will also lead to covered public entities sharing NPD collected from private entities that contains commercially confidential or business sensitive information or in which such private entities hold intellectual property rights.

There is a risk, therefore, that information provided to public entities by private entities during joint collaborations could get categorised as NPD to which the IDAUP applies. The IDAUP does acknowledge this risk in stating that "*all data being shared must ensure compliance to guidelines for legal, security, IPR, copyrights and privacy requirements*".[xxvii] The intention, therefore, does seem to be to ensure that existing intellectual property rights are respected. However, we suggest that this be categorically excluded in the scope and objective of the IDAUP.

Since the material scope of the IDAUP – that is, the data to which it applies – is not precisely defined, it is difficult to determine what these various currently applicable requirements are, and, consequently, how this obligation on covered public entities and recipients of data being shared (the "*acquiring organisation/individual*") will be met.

Suggestions

- **The IDAUP should primarily apply to data that is created, generated, or commissioned by covered public entities using public funds.**
- **The IDAUP should only apply to anonymised data after due care has been taken to ensure identifying information has been removed and that other safeguards to preserve privacy have been implemented. We discuss this in more detail below.**
- **Within the overall scope, the IDAUP should apply to other non-personal data collected from private entities or individuals after adequate technical and organisational measures are implemented to ensure that any data containing information that is commercially confidential, business sensitive, or protected by intellectual property law is not shared. In this regard, we also suggest that these exclusions be specifically stated in the IDAUP.**
- **Adequate safeguards for data must be provided for in the IDAUP. These safeguards should be developed after consultation with relevant stakeholders. The consultation process should have inbuilt transparency and accountability safeguards.**

3 Jurisdictional overlaps

It is likely that the IDAUP will impact the operation of existing laws on government records, such as the Public Records Act of 1993. It will also overlap with several existing open data policies currently in force at the Central and State level, mostly notably the NDSAP, which is similar in both objectives and scope.[xxviii]

The scope of the NDSAP is currently larger – covering "all data or information" held within government bodies, not just non-personal data. The NDSAP also sets up certain institutions referenced in the IDAUP, such as Chief Data Officers.

Beyond this, there are also existing projects at the Central Government level that may overlap with those set up under the IDAUP, such as the India Urban Data Exchange **(IUDX)** under the Smart Cities Mission.

Comments & Concerns

Interplay with NDSAP

While the NDSAP is mentioned in the Background Note, there is no clarity on whether the IDAUP will replace the NDSAP or complement it as a parallel effort. If the IDAUP will replace the NDSAP, then the IDAUP should discuss how it addresses the gaps identified with the latter and the future goals set out as a replacement policy.

However, if the IDAUP will coexist with the NDSAP, then definitional or procedural overlaps should be avoided. In a connected vein, there is also no clarity on whether the licenses proposed under this Draft Policy will subsume or replace the Open Data License under the NDSAP.[xxix]

The NDSAP was owned by two ministries – while MEITY was the nodal department for implementation (through the National Informatics Centre), the Department of Science and Technology was the nodal department on policy matters.[xxx]

In this regard, we note that the IDAUP should operate in consonance with such projects and that the draft should explicitly clarify if such projects fall within its ambit or not.

Alignment with Puttaswamy I

It is also worth noting that implementation of IDAUP must be aligned with principles of data protection laid down in Puttaswamy I ("**Privacy**" judgment) to the extent where it is relevant and existing regulations like the Information Technology Act, 2000 and the regulations therein.

Suggestions

- **The interplay between the IDAUP and the NDSAP should be clarified in terms of scope and implementation. Any overlaps and duplicity of efforts should be minimised.**
- **The final decision taken on the interplay between the IDAUP, the NDSAP and other open data policies should also aim to simplify the overall regime for data governance going forward.**

- **A general objective that should be included in the IDAUP should be to ensure harmonisation and uniformity across such efforts by the Government. The aim should for the IDAUP to set the minimum baseline standards and processes.**

4 Institutional framework

The IDAUP envisages MEITY establishing two new institutions: an India Data Office (**IDO**) and an India Data Council (**IDC**) accompanied by a support unit. It also envisages that every covered public entity will set up a Data Management Unit (**DMU**) and appoint a Chief Data Officer.

The IDO is expected to monitor the implementation and enforcement of the policy. The IDC will be responsible for finalizing data standards. Members of the IDC will include the India Data Officer, the Chief Data Officers, as well as the entities who are primary owners of relevant datasets shall be associate members.

Comments & Concerns

Enforceability

The experience with the NDSAP has demonstrated the difficulties with not having clarity on how to incentivise compliance or penalise non-compliance. Researchers have noted that the lack of any clear method to enforce the NDSAP has meant that it has not been implemented consistently or to its fullest potential.[xxxi]

The problem with the NDSAP may be repeated with the IDAUP if the question of enforceability is not sufficiently discussed. There is a need for clarity on how the IDO will deal with practices that do not adhere to the policy. This may require bestowing the IDO with the power to take corrective measures, including punitive action and directions to covered public entities, and to explore novel strategies to incentivise compliance.

Agency design

The IDAUP does not clarify the legal status of the IDO or the IDC. It is not clear whether these will be attached offices, autonomous bodies, or independent statutory authorities. The IDAUP should discuss the design of these agencies at more depth.

One possible reference point to guide the design of the IDO could be the Financial Data Management Centre (**FDMC**) contemplated by the Financial Sector Legislative Reforms Commission.[xxxii] Though the FDMC was intended to be a data centre for financial sector data, it was also expected to manage requests from external stakeholders for access to such data for research purposes, to develop mechanisms for standardising data collection and reducing duplication of data, protecting the confidentiality and privacy of data.

The hierarchy between the IDO, the IDC, the DMU, and the support unit for the IDC is not clearly laid out. The IDAUP also does not clearly allocate responsibilities between such entities. Though there are some general directions, it is unclear how disagreements will be resolved or how accountability and responsibility for policy monitoring and enforcement will be distributed between them.

It is difficult to imagine this without there being a backing legislative framework. At present, the IDAUP or the Note does not clearly indicate whether such a legislative framework will be developed. It is worth noting that the NDSAP did not lead to satisfactory outcomes even with some legislative backing in place. It can be traced back to a voluntary obligation on public authorities to provide information to the public *suo motu* at regular intervals imposed on such authorities under the Right to Information Act of 2005.[xxxiii] Though this is not touched upon, it may be worth considering a dedicated "open data" legislation, such as the Open Data Directive in the European Union.

Composition of the IDO and the IDC

The current composition of the IDC and the IDO is limited to government stakeholders. We submit that there is both an opportunity and a need to leverage talent and perspectives available outside the government. There should be more multi-stakeholder participation in the IDC with representation from industry, researchers, civil society, technical experts, academia, and other stakeholders. This would strengthen the legitimacy of the decision-making processes adopted by IDC.

To ensure that the significant set of duties and functions allocated to the IDO and the Data Management Units can be effectively carried out, the appointments process for the India Data Officer and the Chief Data Officers should aim to fill these roles with dedicated full-time personnel with the requisite skills and available time, instead of requiring existing officers with other obligations already on their plate to simultaneously fulfil these roles. This was an important drawback highlighted with the NDSAP that we now can address.[xxxiv]

Feedback mechanisms

The institutional framework does not, at present, contemplate a feedback loop being established between end-users, re-users and the IDO or the IDC. There should be mechanisms developed for grievances and suggestions to be provided to the IDO and the IDC in relation to the data sets being shared and being provided for public release.

Suggestions

- **The legal character of the IDO and the IDC should be expressly clarified in the IDAUP.**
- **The IDAUP should clarify the division of responsibilities and functions and the hierarchies across the envisaged institutional framework.**
- **The IDAUP should envisage more multistakeholder participation in the constitution and composition of the IDC.**
- **Dedicated full-time personnel should be appointed to the roles of the India Data Officer and the Chief Data Officers.**
- **The IDAUP should examine new policy and legislative solutions that can incentivise compliance by covered public entities with the IDAUP's proposals and obligations.**
- **The IDAUP should envisage the creation of grievance redressal mechanisms for the IDO and IDC to receive and hear concerns regarding NPD being shared or released for reuse.**

5 Identification and classification of data

The IDAUP envisages covered public entities identifying and classifying non-personal datasets available them on their own. Three categories have been outlined: open, restricted, or non-shareable. The Note also mentions the need for standard classification criteria to assist CDOs in identifying "high-value data sets" (**HVDs**). This concept is not reflected in the text of the IDAUP itself. The IDAUP does state that the envisaged data sharing toolkit will help in identification and classification exercises.[xxxv] There is limited information on what this will look like under the toolkit.

Comments & Concerns

Discretion to covered public sectors

The NDSAP had afforded covered public entities much discretion in identifying and selecting datasets and did not create any incentive measures or penalties to compel performance. It has been suggested that, due to this, Chief Data Officers appointed under the NDSAP were not incentivised to provide access to many high-value datasets, release useful datasets of high quality when they did, or to meet deadlines.[xxxvi]

The IDAUP does not provide any criteria or procedure that can ensure standardisation in classification of datasets across different covered public entities. Though some definitions of "non-shareable" and "restricted" datasets are provided,[xxxvii] these do not go into detail, leaving much room to covered public entities to interpret them as they see fit. There is a risk of inconsistent decision-making and arbitrary or inconsistent classifications being followed.

Due to this lack of classification criteria or procedure, there is a possibility that the classification of datasets will follow existing frameworks that are designed for different ends – such as those to ensure security, such as the classification frameworks contained in the National Information Security Policy and Guidelines which requires Ministries and Departments to classify documents into "secret", "top secret" and "confidential" documents.[xxxviii] More specificity would be necessity to ensure that the ends of the IDAUP are met – for instance, the Open Data Directive in the European Union lays down criteria on the detailed applicability and on the datasets that are outside its scope.[xxxix]

Identification of high-value datasets

It is also unclear whether a goal of the IDAUP is to identify and enable access and reuse of HVDs, since this is not actually discussed in the policy and only mentioned in the Note as a challenge. This is in contrast with international reference points, such as the Open Data Directive, which provides:

- a clear definition of high-value datasets,[xl]
- a list of thematic categories basis which they may be identified – such as geospatial, earth observation and environment, meteorological, statistics, mobility, companies and company ownership,[xli]
- a set of factors to identify them,[xlii]

- a set of principles set out how they will be made available for publication and reuse – such as making sure they are available free of charge, machine readable, provided via APIs or as bulk downloads, where relevant.

Given the above context, we suggest that the framework pertaining to HVDs should either evolve from IDAUP or from the proposed NPD framework instead of two overlapping and conflicting governance frameworks. We would like to bring attention to the Committee on Non-Personal Data (NPD) Governance Framework which had noted that high value datasets would be datasets beneficial to the community and shared as a public good. This includes data sets pertaining to financial inclusion, diversity and inclusion, energy, urban planning amongst others. The Committee had noted that a data trustee such as a government organisation or a non-profit private organisation would be responsible for creation, maintenance, and data sharing of HVD data sets.[xliii]

Recourse for classification decisions

The IDAUP also does not discuss the possibility of any recourse mechanism for citizens and stakeholders to challenge decisions to classify NPD into restricted or non-shareable categories and to deny access to such categories of NPD. At the very least, a requirement may be incorporated to obligate covered public entities to provide adequate grounds for refusal that are in line with the rest of the guidelines laid down by the IDC and the IDO.

Suggestions

- **The objective of ensuring standardisation should be operationalised by setting out criteria and procedures to guide identification and classification exercises.**
- **Identification criteria and procedures are particularly relevant in the context of HVDs held within the public sector. The IDAUP should include a formal definition of HVDs or at least lay out the principles based on which HVDs may be identified.**
- **HVDs should be made available as per well-laid principles, in machine-readable formats, provided by APIs or, where relevant, in formats that enable bulk downloads. For instance, with respect to pricing of such HVDs, it may be considered if the service would merit any charges, how would it relate to the costs, and if charges will create cost barrier for users amongst others.**
- **An appeals process may be introduced to address disputes over denials of requests to release certain datasets on grounds that they are restricted or non-shareable.**
- **The framework pertaining to HVDs should either evolve from IDAUP or from the NPD framework instead of two-overlapping and conflicting governance frameworks.**

6 Data sharing, portals and retention

Covered public entities will be expected to create searchable data inventories that shall then be federated into a government-wide searchable database.

All data portals and dashboards currently maintained by covered public entities would also be integrated with the open government data portal. The India Data Office**(IDO)** shall provide technical & implementation support to achieve this integration.

The IDO will notify protocols for sharing of NPD and most datasets would be made available at no cost. However, "restricted datasets" may be subject to additional (undefined) protocols and processes. A data sharing toolkit will assist covered public entities to assess the risks of data sharing and release.

Each covered public entity will adopt and publish their own domain-specific metadata and data standards that should be aligned with existing guidelines. The IDC will finalise data standards that may cut across domains.

Each covered public entity will also be permitted to set their own data retention policies. A broad set of guidelines will be standardised and provided to guide such exercises.

Comments & Concerns

Risk of fragmentation in data management

Permitting covered public entities to create their own domain-specific metadata and data standards may create siloes and lead to fragmentation. Interoperability of datasets across covered public entities may also be frustrated by a lack of clarity and standardisation on the formats to be adopted for sharing different types of structured and unstructured datasets. Some state-level policies do seek to address this by mandating the use of common open-source machine-readable file and data interchange formats (like JSON or XML).[xliv]

Role of government in ensuring data quality

Neither the IDAUP nor the Note provide much insight into the envisaged role for covered public entities in certain stages of the overall lifecycle of NPD from a data quality. For example, a data production and analysis pipeline may involve data being (1) collected or generated (2) cleaned and validated (3) analysed and finally (4) real-world decisions can be made based on that analysis.[xlv] The IDAUP does not discuss the role of covered public entities in the cleaning and validation stages, yet the protocols and strategies they employ will have a significant impact on the quality of the NPD that is finally shared and released. This requires framing a definition of data quality. A suitable international reference point in this regard may be the UK Government's Data Quality Framework.[xlvi]

Learnings from existing open government data portals

The Note and the IDAUP does not investigate the experiences of different stakeholders with the NDSAP or with connected efforts such as the Open Government Data Platform (**OGDP**). Though these are briefly mentioned, there is no specific discussion on what did and did not work with such initiatives.

Researchers in India comparing the OGDP with similar initiatives in other countries have found that there are a variety of good practices that the OGDP does take – such as offering data via application programming interfaces, providing data visualization tools, providing a data suggestion feature, and consistently increasing the number of available datasets on the platform each year.[xlvii]

They also found that there is scope for improvement, such as by introducing machine-readable formats, making author information available for all datasets, ensuring open licensing and copyrights, providing general discussion forums to engage with the user community, promoting API-based harvesting of metadata, simplifying the use of API access.[xlviii] Such suggestions should be examined in more length.

For example, the IDAUP also does not discuss whether data sharing protocols will obligate covered public entities to release documentation on datasets to help re-users make the most of the data they receive access to. Such documentation should cover the use of data being delivered, the source and methodology employed to create or collect such data and whether there are any limitations on their use. This would significantly improve the likelihood of re-users trusting the data being accessed. Such requirements should be incorporated as part of the data sharing toolkit.

The IDAUP can also consider incorporating mechanisms for ranking open government data portals according to measurable criteria linked to accessibility, user-friendliness, completeness, quality, etc. This could help in incentivising the quality and usefulness of such portals.

Release of real-time dynamic data

There is no mention on whether the IDAUP will also lead to the release of real-time ("*dynamic*") data. A recurring concern with the quality of public sector data available in India today is the fact that it is often outdated. Real-time data (such as sensor or machine generated data) can be particularly useful for research and innovation. For instance, sharing of real-time data of air quality index **(AQI)** over the years has led to Graded Response Action Plan **(GRAP)** to combat deteriorating air quality in Delhi-NCR.[xlix] There are cases where data could be released within stipulated time-periods as well. In this regard, we note that a process stipulating timeline for release of data needs to be formulated. These stipulated timelines may be reviewed on a periodic basis to improve upon. However, the release of real-time data should not come at the cost of removing confidential or personal information.

Further, we suggest that standards be laid down for quality of data shared as well. Such as, what is the expected error rate of such data, how can it be improved, what metadata fields is required, if audit mechanisms are required, transparency and accountability measures to operationalise quality data sharing amongst others.

Design of data sharing toolkit

In the interest of transparency, the data sharing toolkit should be made public and be finalised using contributions from external stakeholders, such as community, academic and industry contributions. Based on a preliminary review of international reference points,[l] the elements of such a toolkit could include:

- Frameworks and tools to aid identification and classification
- Mechanisms for data-cleaning, formatting, standardisation, and quality control
- Mechanisms for data sharing and release via multiple channels
- Decision-making support for choice of licensing
- Data anonymisation requirements, guides, and tutorials.

In the past, accessibility and visualisation of data has been a challenge. For instance, RBI publishes Annual Handbook of Statistics on Indian Economy which has various indicators and data. One of these data is on Payment System Indicators which is while presented in two separate sheets for a span of 5 years, thereby, bringing down the user experience.[li] Since India Data Office (IDO) has been envisaged to work in tandem with Chief Data Officers of each ministry/department, it may address the concerns associated with absence of data in a user-centric manner. In this regard, we note that it must be the responsibility of India Data Office to ensure accessibility and visualisation of data being made available is in a user-centric manner such as in open, machine-readable formats, easy UI/UX amongst others.

Suggestions

- **The IDC should be tasked with creating a country-wide common minimum standards for data and metadata associated with commonly occurring elements.**
- **The IDAUP should articulate a conception of data quality and set out specific proposals and strategies to be followed by covered public entities towards ensuring data quality.**
- **The IDO should develop a template or reference portal that can serve as a gold standard for open data portals.**
- **The IDO should be tasked with determining how accessibility and visualisation of data being made available can be ensured in a user-centric manner.**
- **The use of open-source and machine-readable formats across the board should be explicitly clarified in the IDAUP to promote interoperability.**
- **MEITY should examine the scope for proposals on the release of dynamic datasets.**
- **The IDAUP should also explicitly clarify that covered public entities will need to release adequate supporting documentation with datasets released for public reuse.**
- **Mechanisms may be developed to index and rank data sharing portals made available by covered public entities.**

7 Licensing

The IDO will create "*innovative and just licensing frameworks"* to enable fair access and use which can be used by covered public sector entities.

The IDAUP does not discuss how the costs of compliance and of making NPD available for sharing and reuse will be borne by covered public entities and whether existing budgets will be reallocated to facilitate compliance.

Comments & Concerns

Overall purpose of licensing frameworks

There may be apprehensions that licensing frameworks will be used to generate revenue and profits from the sale of NPD. However, this initiative should be driven by the goal of maximising benefits for citizens and not aimed at revenue maximisation.

This flows from a general underlying principle that data held with the public sector, ultimately, intrinsically belongs to Indian people, and is created, collected, or generated using public funds collected from Indian people. More practically, setting high license fees may create an entry barrier for start-ups and SMEs. For pricing, it may be considered if the service would merit any charges, how would it relate to the costs, and if charges will create cost barrier for users amongst others.

Recovery of marginal costs

That said, there can be cases where providing access to re-usable datasets may be a costly exercise to undertake for covered public entities. Licensing frameworks could step in here – if used, at a principled level, primarily for the purpose of recovering marginal costs relating to the collation, reproduction, preparation, and dissemination of such datasets for sharing and release and to fund the use of anonymisation tools and measures to protect commercially confidential information.

Considerations for designing licenses

Licenses should be afforded in a transparent, non-discriminatory, objective, proportionate and fair manner. Any conditions imposed should be justifiable on public interest grounds and should not be used to restrict competition or possibilities for re-use. This would be in line with key international reference points cited in the Note.[lii]

Licensing frameworks should not lead to the execution of exclusive data sharing arrangements in favor of specific recipients and re-users. This is in the interests of ensuring fair competition and equal access to open government data to all. This should be explicitly clarified.

Further, drawing from analysis on the difficulties with the Open Data License and the NDSAP[liii], the aim with such licenses under the IDAUP should be to facilitate the use of truly "open data" that is free from copyright and places the onus on data providers – the covered public entities – to determine the legality of data sets that are released and validity of compliance with the IDAUP.

This would not mean that data re-users would not be required to ensure proper attribution. A suitable reference point, from an intellectual property rights perspective, is the Open Government License formulated and relied upon by the Government or the United Kingdom that is designed to work in parallel with current open-source licenses (such as the Creative Commons Attribution license).[liv]

Suggestions

- **Licensing frameworks should be primarily used for the recovery of marginal costs of making data or information accessible and available for reuse and should not become a tool for engaging in exclusive arrangements.**
- **Licenses should place the onus on data providers to ensure compliance with the IDAUP and be aligned with open-source copyright licensing frameworks.**

8 Anonymisation and privacy safeguards

The IDAUP emphasises that "*any data sharing shall happen within the legal framework of India, its national policies and legislation as well as the recognized international guidelines. This will prevent misuse of data and assure security, integrity, and confidentiality of data*".

The IDAUP envisages anonymisation as the principal tool to ensure the preservation of privacy. It notes that covered public entities "*must comply with anonymisation standards defined by IDO/MEITY or by any statute/act/policy issued by the government of India*". It envisages the provision of reference anonymisation tools and decision-making frameworks to covered public entities to assist data officers in managing data sharing requests.

Comments & Concerns

The creation of federated and integrated government-to-government data sharing infrastructures can create apprehensions of significant profiling of citizens, leading to a loss of their privacy, *if* sufficient care is not taken to ensure that such infrastructures do not involve the sharing of identifying information and measures are put in place to reduce the risks of identification of persons through linking and triangulation.

Role of anonymisation

The scope of the IDAUP is limited to *NPD* and *information*. However, the use of anonymisation tools on personal data may lead to the resulting datasets to be regarded as NPD. It is the use of these tools that will, in practice, determine the scope of data to which the IDAUP will apply. Anonymisation is also being presented as the primary solution to preserve individual privacy and to prevent identifying information from being shared. This twin role of boundary setting, and of privacy preservation makes anonymisation a crucial component of the IDAUP.

Need for complementary measures and safeguards

There is very limited discussion in the Note or in the IDAUP on how different privacy and digital security risks will be addressed during data sharing beyond the use of anonymisation. This may not be sufficient. While we do not discuss the various safeguards and frameworks that may be needed from a digital security perspective, it is worth noting that, from a privacy perspective, anonymisation as a means of protection may not be sufficient as a standalone solution.

Researchers[lv] and experts[lvi] have noted that anonymisation should be complemented with other technical and organisational measures, such as contractual agreements that bind data recipients to data security and disclosure practices, the use of reidentification risk assessments, or the use of newer privacy preserving technologies such as distributed machine learning, differential privacy, and homomorphic encryption (that is, encryption that allows processing of encrypted data without revealing its embedded information). Further, as a general principle, "release-and-forget" models of data sharing should be avoided. This should be reflected in policy monitoring and enforcement efforts through subsequent audits and review processes.

Setting of anonymisation standards, tools and frameworks

The IDAUP intends for MEITY or the proposed India Data Office (**IDO**) to prescribe anonymisation standards – a task that is entrusted to the proposed Data Protection Authority of India to be set up by the DPB 2021. Given the importance of anonymisation to both frameworks and the need for harmonisation and certainty, there should be a singular authority setting out such standards. We submit that this function should be entrusted to the Data Protection Authority as the primary authority intended to ensure the protection of personal data across sectors and data categories.

Till such time that the Authority is set up, MEITY should constitute an external working group with suitable multi-stakeholder participation to evolve the reference anonymisation tools and decision-making frameworks in a transparent manner. Such tools and frameworks should be subject to an open public consultation process to ensure they are scrutinised adequately and are sufficiently robust before being deployed.

Suggestions

- **The IDAUP would benefit from a more detailed discussion on additional privacy preserving technologies, as well as other technical and organisational measures to avoid "release-and-forget" models of data sharing may be employed to further reassure citizens of privacy and digital security risks being minimised.**
- **An independent working group with multi-stakeholder participation should be established by MEITY to frame reference anonymisation standards, tools, and decision-making frameworks.**

9 Need for phased implementation

The IDAUP states that detailed implementation guidelines will be brought out by MEITY in the form of an implementation manual.

Comments & Concerns

The IDAUP will apply to a wide range of covered public entities. At present, neither the Note nor the IDAUP discuss strategies on how its proposals will be scaled. There may be unintended

consequences to expecting each covered public entity to shift to adopting the IDAUP's proposals.

For example, with respect to the proposed applicability to "autonomous bodies", the IDAUP does not consider the cost implications of its proposals on entities that provide key citizen-centric services. For example, given that the current intention is to apply the IDAUP to autonomous bodies that are providers of healthcare services (like the All-India Institute of Medical Services), it could lead to an increase in the cost of providing such services which may then be passed onto citizens.

Even though the proposed scope of the IDAUP is to apply to non-personal data, there may be scenarios of deficient anonymisation or reidentification by data providers or data re-users. Under current law, it would be difficult to determine how to penalise such scenarios and deter non-compliance with privacy preserving safeguards without such a personal data protection law in place.

Suggestions

- **The IDAUP would benefit from a structured implementation plan that scales its proposals in stages – perhaps by first identifying priority HVDs, then identifying the relevant data providers, and then scaling slowly to other datasets.**

10 Conclusion

In this submission, we have sought to identify several areas where the IDAUP could be improved. We consider our suggestions as starting points towards developing a mature and robust open data policy for India. We would be happy to engage on specific areas at more depth as required by MEITY, including, for example:

- Analysis of existing open government data initiatives and international reference points
- Design and development of suitable agency design and legal frameworks
- Preparation of implementation manuals, data sharing toolkits and licensing frameworks
- Mapping of existing laws to determine how legal, security, privacy and IPR risks may be accounted for in the sharing and release of NPD

We look forward to continuing to engage with MEITY on this important topic. For any queries related to this submission, please contact S Jayakumar (jayakumar@nasscom.in); Varun Sen Bahl (varun@nasscom.in); or Apurva Singh (apurva@nasscom.in).

For any queries or information related to public policy in NASSCOM, kindly write to Ashish Aggarwal (asaggarwal@nasscom.in)

Endnotes

[i] *See* clause 2, Draft India Data Accessibility and Use Policy, (2022), available at: https://www.meity.gov.in/writereaddata/files/Draft%20India%20Data%20Accessibility%20and%20Use%20Policy_0.pdf (last accessed on 16th March, 2022) ("**IDAUP**").
[ii] This is indicated by these objectives: *maximising access to and use of quality non personal data available with public sector* (paragraph 2.1); *increasing the availability of datasets of national importance* (paragraph 2.12).
[iii] This is indicated by these objectives: *streamlining inter-government data sharing while maintaining privacy* (paragraph 2.6); *enabling secure and privacy compliant pathways to share detailed datasets for research & development* (paragraph 2.11); *building digital & data capacity, knowledge & competency of government officials* (paragraph 2.8); *promoting data interoperability and integration to enhance data quality and usability* (paragraph 2.9).
[iv] This is indicated by these objectives: *improving policymaking, evaluation, and monitoring* (paragraph 2.2); *enhancing the efficiency of service delivery* (paragraph 2.3); *facilitating the creation of public digital platforms* (paragraph 2.4); *ensuring greater citizen awareness, participation, and engagement with open data* (paragraph 2.10).
[v] This is indicated by these objectives: *protecting the privacy and security of all citizens* (paragraph 2.5); *promoting transparency, accountability, and ownership in data sharing & release* (paragraph 2.7); improving overall compliance to data sharing and privacy policies and standards (paragraph 2.13).
[vi] *See* clause 5, IDAUP, Page 3.
[vii] *See* clause 12, *IDAUP,* Page 7.
[viii] *See* Section 43A, the Information Technology Act of 2000.
[ix] For a discussion of the public and economic importance of "open government data", *see* Open Knowledge Foundation, *The Open Data Handbook,* (2015) available at www.opendatahandbook.org. For a discussion on the specific value that open government data can hold for India, *see* S. Asher *et. al., Big, Open Data for Development: A Vision for India,* National Council of Applied Economic Research, (2021), available at https://www.ncaer.org/Events/IPF-2021/Paper/Paper-I-Sam%20Aditi%20Alison%20Johns-Conference-Version-IPF-2021.pdf
[x] *See* Section 4, Right to Information Act, 2005.
[xi] *See* Ministry of Housing and Urban Affairs, *the India Urban Data Exchange (IUDX) by the Smart Cities Mission*, available at: https://iudx.org.in/
[xii] *See* S. Asher *et. al., Big, Open Data for Development: A Vision for India,* National Council of Applied Economic Research, (2021), available at https://www.ncaer.org/Events/IPF-2021/Paper/Paper-I-Sam%20Aditi%20Alison%20Johns-Conference-Version-IPF-2021.pdf (**NCAER Paper)**
[xiii] This requires that, where processes or systems are being updated or new processes are being designed, open access to data must be built in at the start. This comes from the Open Data Directive.
[xiv] *See* European Parliament, *Directive 2019/1024 on open data and the reuse of public sector information,* (2019), available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L1024 (the **Open Data Directive**).
[xv] *See* Parliament of Australia, *Data Availability and Transparency Bill 2020,* (2020), available at: https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r6649 (the **DAT Bill**)
[xvi] Namely, the Project principle – data is shared for an appropriate project or program of work; People principle – data is made available only to appropriate persons; Setting principle – data is shared in an appropriately controlled environment; Data principle – appropriate protections are applied to data; and Outputs principle – outputs are as agreed. *See* the DAT Bill.
[xvii] *See* GO FAIR, *Fair Principles*, (2022), *available at https://www.go-fair.org/fair-principles/*
[xviii] *See* M. Wilkinson, *The FAIR Guiding Principles for scientific data management and stewardship,* Nature, (2016), available at https://www.nature.com/articles/sdata201618
[xix] *See* Ministry of Electronics & Information Technology, Report by the Committee of Experts on Non-Personal Data Governance Framework, (2020), available at https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf (**NPD Report**).
[xx] A definition of non-personal data is contained in Clause 3(28), (Draft) Data Protection Act of 2021 (DPB 2021). *See* Committee under chairmanship of Shri P.P. Chaudhary, Seventeenth Lok Sabha, Report of the Joint Committee on the Personal Data Protection Bill, 2019, (2021), available at http://loksabhaph.nic.in/Committee/CommitteeInformation.aspx?comm_code=73&tab=1
[xxi] *See* Section 2(1)(w), Information Technology Act of 2000.

[xxii] *See* paragraph 2.6, National Data Sharing and Accessibility Policy, (2012), available at: https://geoportal.mp.gov.in/geoportal/Content/Policies/NDSAP_2012.pdf
[xxiii] In more specific terms, data that never related to an identified or identifiable natural person, such as data on weather conditions, data from sensors installed on industrial machines, data from public infrastructures, and so on. *See NPD Report,* 7, (2020).
[xxiv] In more specific terms, data which were initially personal data, but were later made anonymous. *See NPD Report,* 7, (2020).
[xxv] By covered public entities, we mean the "*the Government of India directly or through authorized agencies by various Ministries/Departments/Organisations/Agencies and Autonomous bodies" as well as the State Governments* to which the IDAUP may apply. This is as per clause 4, *IDAUP,* Page 3.
[xxvi] For a brief overview of different data collection mechanisms in relation to NPD held with the public sector, *see NPD Report,* 8, (2020).
[xxvii] *See* clause 15.4, *IDAUP,* Page 8.
[xxviii] *See* Department of Science & Technology, National Data Sharing and Accessibility Policy, (2012), available at https://dst.gov.in/sites/default/files/gazetteNotificationNDSAP.pdf (**NDSAP**)
[xxix] *See* Government Open Data License – India, National Data Sharing and Accessibility Policy – 2012, (2017), available at https://data.gov.in/government-open-data-license-india (**Government Open Data License**)
[xxx] *See Government Open Data License.*
[xxxi] *See* R. Bailey, R. Sane, *A missed opportunity,* The Hindu, (2020), available at https://www.thehindu.com/opinion/op-ed/a-missed-opportunity/article32507522.ece
[xxxii] *See* Ministry of Finance, *Report of the Financial Sector Legislative Reforms Commission,* (2013), available at https://dea.gov.in/sites/default/files/fslrc_report_vol1_1.pdf
[xxxiii] *See* Section 4(2), Right to Information Act, 2005.
[xxxiv] *See* S. Carmeli, *To Unlock the Potential of Open Government Data, India Needs New, Comprehensive Legislation,* (2021), available at https://thebastion.co.in/politics-and/tech/to-unlock-the-potential-of-open-government-data-india-needs-new-comprehensive-legislation/
[xxxv] *See* clause 15.1, *IDAUP,* Page 8.
[xxxvi] *See* S. Carmeli, *To Unlock the Potential of Open Government Data, India Needs New, Comprehensive Legislation,* (2021), available at https://thebastion.co.in/politics-and/tech/to-unlock-the-potential-of-open-government-data-india-needs-new-comprehensive-legislation/
[xxxvii] *See* definitions of "negative list" and "restricted access data sharing" in Annex I, *IDAUP,* Page 9.
[xxxviii] *See* Ministry of Home Affairs, *National Information Security Policy and Guidelines,* (2014), available at http://faridkotpolice.in/guidlines.pdf
[xxxix] *See* Article 1, *the Open Data Directive.*
[xl] These are defined as "*documents the re-use of which is associated with important benefits for society, the environment and the economy, in particular because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs, and of the number of potential beneficiaries of the value added services and applications based on those datasets". See* Article 2(10), *the Open Data Directive.*
[xli] *See* Article 13(1) read with Annex 1, *the Open Data Directive.*
[xlii] Specifically, an HVD will be identified based on whether it has the potential to *"(a) generate significant socioeconomic or environmental benefits and innovative services" (b) "benefit a high number of users, in particular SMEs" (c) "assist in generating revenues"* and *(d) "be combined with other datasets"* See Article 14, Open Data Directive.
[xliii] *See,* MeitY, Para 7.6, *Committee on Non-Personal Data Governance Framework, (2020)*, available at https://ourgovdotin.files.wordpress.com/2020/12/revised-report-kris-gopalakrishnan-committee-report-on-non-personal-data-governance-framework.pdf.
[xliv] *See* Tamil Nadu Data Policy, 2022.
[xlv] *See NCAER Paper.*
[xlvi] *See* Government of UK, *Government Data Quality Framework,* (2020), available at https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework
[xlvii] *See* S. Ganapathy *et. al., An investigation of National Open Government Data Platforms: How can India improve?,* Data Governance Network, Working Paper No. 26, (2021), available at: https://datagovernance.org/files/research/1640243994.pdf (**IDFC Paper**)
[xlviii] *See IDFC Paper.*

[xlix] See Indian Express, Explained: *What is GRAP, Delhi-NCR's action plan as air pollution increases?* , (2020), available at https://indianexpress.com/article/explained/explained-what-is-grap-delhi-ncrs-action-plan-as-air-pollution-increases-6719746/.

[l] *See* World Bank, *Open Government Data Toolkit,* (2019), available at http://opendatatoolkit.worldbank.org/en/open-data-in-60-seconds.html ; Government of Canada, *Open Data Toolkit,* (2019), available at https://open.canada.ca/en/toolkit/diy ; Government of New Zealand, *Open Data Toolkit,* (2019) available at https://codeforaotearoa.github.io/ ; Government of Australia, *Open Data Toolkit,* (2018), available at https://toolkit.data.gov.au/

[li] *See* RBI, *Payment System Indicators,* Annual Handbook of Statistics on Indian Economy, available at https://rbi.org.in/Scripts/Publications.aspx?Publication=Annual.

[lii] *See,* for example, Article 6, *Open Data Directive.*

[liii] *See* D. Joshi, *Open/Secret – Assessing India's Commitment to Open Data,* (2017), available at https://spicyip.com/2017/06/opensecret-assessing-indias-commitment-to-open-data.html ; Thejesh G.N, *Open Data in India: In a Restrictive Copyright Regime, Voluntary Organisations Pitch in to Make Data Accessible,* Economic and Political Weekly, (2020), available at: https://www.epw.in/node/157030/pdf

[liv] *See* National Archives, *Open Government License for public sector information,* Government of UK, (2014) available at: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

[lv] *See* A. Kak *et. al., Open Data and digital identity: Lessons for Aadhaar,* NIPFP Working Paper, (2017), available at https://macrofinance.nipfp.org.in/PDF/kakParsheeraKotwal2017_open_data_aadhaar.pdf

[lvi] *See* E. Ronchi *et.al., Enabling Access to and Sharing of Daa: Reconciling Risks and Benefits of Data Re-use across Societies,* OECD, (2018), available at https://doi.org/10.1787/276aaca8-en;