Non-Personal Data

Policy, Economics and Technology

Srijoni Sen
NLSIU Bangalore
srijonisen@nls.ac.in

Inder Gopal
IISc Bangalore
indergopal@iisc.ac.in

D Manjunath
IIT Bombay
dmanju@ee.iitb.ac.in

Abstract

In response to the Expert Committee Report proposing a Non-Personal Data Governance Framework, the authors of this paper organised a series of discussions involving leading experts and practitioners across different fields thinking about NPD. This paper captures and expands on the main issues brought to light, offers our own analysis and critique, and details some suggestions on the way forward. We particularly examine the claimed connections between the proposed framework and its potential to accelerate innovation for the public good.

1. Background

Governments around the world are placing increasingly strict limits on what can be done with the personal data of individuals. This follows from the nearly universal recognition that the privacy rights of the individual must be protected in the data driven economy. Thus, in many countries, the law is evolving to ensure that personal data is protected from abuse in a manner that can cause harm to the individual or to the group. In India, the Personal Data Protection Bill, 2019 is expected to form the basis for such laws.[1] Non-Personal Data (NPD), loosely defined as data that does not identify specific individuals, is a more complex matter. While there are several privacy, ownership and security issues in the handling of NPD, there is also increasing recognition that there is a potential public good that can be achieved by making such non-personal data available to the public and to the private sector more freely. There is also a belief that the availability of such data can generate significant economic activity.

However, as is to be expected, a wide range of issues, relating to public policy, economics, business models, technology, law, regulation, governance, etc., emerge in the conceptualisation and implementation of any system enabling the exchange and usage of NPD. There have been several efforts around the world to explore these issues and address some of them.[2] Clearly, these issues, and the solutions to address them, depend on the history and culture, and also state of technological development of each society, and hence are unique to each country. This is more so in a large and diverse country such as India, and requires country specific solutions.

The Government of India has recognized the economic and social potential of NPD and in September 2019, the Ministry of Electronics and Information Technology (MeitY) set up a committee of experts with Kris Gopalakrishnan as the chair. The committee was charged to study the various issues relating to NPD and to make specific suggestions on its regulation. The first version of the report,[3] titled "Non Personal Data Governance Framework, submitted in July 2020 and the public were invited to submit their comments. A revised version[4] was published in January 2021 and public comments were invited.

The publication of the first version of this report, henceforth referred to as the MeitY NPD Governance Framework and abbreviated as NGF (NGF-1 for the first version and NGF-2 for the revised second version), has generated significant interest among economists, lawyers, public policy experts, in addition to the industry. The main aspects of NGF are discussed later in this paper. Several organisations and individuals have analyzed the report.[5] The focus in many of

[1] See 'A Free and Fair Digital Economy: Protecting Privacy, Empowering Indians', A report submitted to MeitY by a Committee of Experts under the Chairmanship of Justice B. N. Srikrishna. Available at https://www.meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf

[2] For example, see https://digital-strategy.ec.europa.eu/en/policies/non-personal-data and https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en for the European perspective.

[3] https://static.mygov.in/rest/s3fs-public/mygov_159453381955063671.pdf

[4] https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf

[5] See, for example, https://www.pwc.in/assets/pdfs/news-alert-tax/2020/pwc_news_alert_28_july_2020_report_on_ndp_governance_framework_open_to_comments.pdf

these analyses is on the legal aspects of the proposed framework. To capture a wider perspective that included the economics and the technology aspects, along with the legal issues, we organized a series of panel discussions in which leading experts and practitioners from across different fields presented their views and engaged in debates on the possibilities and pitfalls of an innovative framework governing NPD in India in general and NGF in particular.

The introductory session of the Panel Series served to provide a wide-overview of the issues involved. The session was moderated by Inder Gopal and had the following panelists:

- Parminder Jeet Singh, Executive Director, IT for Change,
- Wendy Hall, Regius Professor of Computer Science, University of Southampton, UK,
- Shantanu Bhattacharya, Chief Data Officer, Airtel,
- Vijay Chandru, Pioneer at AI Foundry and Professor, BSSE, IISc, and
- D. Manjunath, Professor, IIT Bombay.

Sessions 2 and 3 focussed on issues of law, governance, policy and rights. Session 2 was moderated by Srijoni Sen, and the following were the panelists.

- Rahul Matthan, Partner at Trilegal.
- Neha Munjral. General Counsel, GE Healthcare.
- Rahul Singh, Associate Professor, NLSIU.
- Desh Gaurav Sekhri, Head, Access to Justice Initiative, NITI Aayog, and
- Kailas Kartikeyan, Founder, Gestalt Strategy Consulting.

While Session 2 of the series focussed on governance frameworks and regulatory issues, Session 3 saw a deep dive into the many questions of individual and community rights, and privacy and data protection concerns that the approach on non-personal data had raised. This session was moderated by Udbhav Tiwari of Mozilla and the following were the panelists.

- Amba Kak, Director, Global Strategy and Programs, AI Now Institute.
- Astha Kapoor, Co-Founder, Aapti Institute.
- Estelle Masse, Senior Policy Analyst & Global Data Protection Lead, Access Now.
- Anush Kapadia, Assistant Professor, IIT Bombay, and
- Sarah Mary Stanley, Consultant, Dvara Research.

The business models and regulatory frameworks associated with generating economic value from non-personal data (NPD) the subject of panel #4. The discussions were centred around the critical topics of data economics and data monetization. This panel was moderated by Prof V. Sridhar (VS) from the Centre for IT and Public Policy at IIIT Bangalore. The panelists were

- Inder Gopal, Research Professor, IISc, Bangalore,
- Wolfgang Kerber, Professor Economics, University of Marburg, Germany, and

https://www.ikigailaw.com/summary-of-the-report-of-the-committee-of-experts-on-non-personal-data/#acceptLicense
https://www.medianama.com/2020/08/223-non-personal-data-ispirt-comments/

- Amol Kulkarni, Director, Consumer Unity and Trust Society.

This panel happened in the backdrop of the release of the second version of the NGF, i.e., the release of NGF-2 and it provided the framing construct of the discussions.

The final session of this series explored healthcare data as a specific facet of NPD. The session was moderated by Prof. Vijay Chandru of the Indian Institute of Science, and had the following panelists:

- Ajay Mahal, Deputy Director, Melbourne School of Population and Global Health
- Ajay Nair, CEO, Swasth Digital Health Foundation
- T.V. Sekher, Professor, International Institute for Population Sciences (IIPS)
- Rahul Matthan, Partner, Trilegal, and
- Gautam Menon, Professor, Ashoka University

The discussions were held during November 2020-February 2021. In this paper, we capture the main issues brought to light in these discussions, provide our own critique of elements of the current discourse, and make several suggestions on the way forward. We also expand on many of the issues that were raised during the panel discussion.

We emphasize that the comments in this paper are heavily influenced by the above-mentioned panels but ultimately are entirely the views of the three authors of this paper. While the panelists generously gave their time and expertise, they have not explicitly or implicitly endorsed (or not endorsed) the contents of this paper.

Interestingly, the India Urban Data Exchange (IUDX) project at IISc, sponsored by the Ministry of Home and Urban Affairs, is building infrastructure to collect data from the Smart City project, an excellent example of a class of NPD, and developing systems to enable the sharing of the collected data and enable subscribers to the Exchange to extract useful information from the data and build services. We believe that the experience of building this infrastructure provides one of the authors (IG) a unique perspective that can be of immense value in realising the vision of NGF-1 and NGF-2. We will have more to say about this project later in this paper.

2. Overview of Issues

In this section we provide an overview of NGF, relevant international developments shared, and the perspectives of large industries. We also describe ways in which NPD can unlock value in different fields.

Data Ownership and Data Exchanges

We begin with an overview of the position of NGF-1.[6] It is emphasised that the proposed framework stresses on the need to rethink models of ownership for data that is generated from the community, or from public spaces, and also on the importance of laying down the conceptual foundations of data sharing. The NGF has consciously chosen to mirror the approach adopted for personal data by defining roles of the data principal, trustee and custodian. It also breaks new ground in describing arrangements on how NPD should be handled and shared. It is strongly emphasised that the individual's rights are protected in the NGF.

The arguments that make up the basis for sharing NPD can form the bedrock of arguments for mandated sharing of data, especially data collected from individuals and community spaces. There are many possible bases for data sharing; it could be in the provision of public goods and welfare services, voluntary or request-based private-to-private sharing of data, and also sharing in furtherance of a legal mandate.

To create value from data that serves communities meaningfully, and to convert it into a public good, a key step is the 'wrangling and ingesting of data'. Public agencies already generate large volumes of data using public funds, but inadequate implementation of the National Data Sharing and Access Policy (2012)[7] means that much of the data remains inaccessible. It must be noted that even where it can be accessed, processing of that data for critical use-cases is also a challenge. It is also important to be aware of this experience when the government moves to implement the vision and the mechanisms set out in NGF-1.

It is strongly suggested that infrastructure that enables data sharing, possibly in the form of a data exchange platforms[8], needs to be developed post haste. Such a platform (or a suitable alternative) can then be leveraged by companies to process large quantities of non private data and then present to the public for use, thus essentially converting them into a public good. Active data sharing and exchange mechanisms through an NPD governance framework can address much of this. These exchanges can also serve as data markets and also to provide other services. Like

[6] The revised version, NGF-2 had not been published. We remark here that we believe that there have been substantial changes in the revised version.

[7] https://en.wikipedia.org/wiki/National_Data_Sharing_and_Accessibility_Policy_–_Government_of_India provides a good introduction while the document is available at https://data.gov.in/sites/default/files/NDSAP%20Implementation%20Guidelines%202.4.pdf . In other fora it has been remarked that India among the pioneers of the Open Data movement. That it has not taken off and an NGF like exercise is necessitated is perhaps an indication that this movement has lived up to its expectations.

[8] An example is the IUDX platform that provides a framework to share urban data from the Smart City projects in India. This w

Internet exchanges, provisioning of such an exchange infrastructure could itself have a business model.

It can be argued that India, in particular, is well placed to contribute to the development of an NPD eco-system: we generate voluminous quantities of data, there are many open source data models that can be readily deployed, and there a large number of use-cases that can help address the challenges of our developing economy. An example pointed out during the panel discussions was with respect to increased precision of address data in India that could enable e-commerce sites to significantly reduce their delivery time. It is estimated this would eventually lead to a $10-12 billion increase in the country's GDP (0.5% increase). Another example is in traffic signal management. Traffic police have been known to control the traffic signals by estimating the amount of traffic backup at different choke-points using Google Maps. The feedback could surely be made more efficient and better algorithms could be used to control the traffic signals and perhaps even reroute traffic to reduce congestion. To summarise, the idea would not just be about selling data, but of selling data for the purpose of adding value.

When we talk about data exchanges and processing for value addition, there are natural questions about ownership, incentive structures and pricing of data. It is important to sound a cautionary note where poorly thought incentive structures can cause more harm than good, especially when there are conflicting data principles from 'overlapping communities.' There is a general agreement that new legal frameworks are needed to conceptualise data as a property right, and possibly thinking beyond markets alone setting the terms of price discovery. It is important to note that legal fictions [9] have operated in different spheres in response to public policy considerations; for example, community rights over water or other commons. Along with the core concerns that one might expect (see many of the public discussions mentioned earlier), two possibly innovative suggestions were made.

1. Consider certain categories of data as community resources or data commons that should be governed by an overarching framework.
2. Ownership of data clearly lies with the community that created the data in question and it could be considered as an input to the production process of a firm that is using it to create value. Thus, it could be that the community owns the data and it is community property.

Both of these would require the recognition of a data custodian.

Data Custodians

Clearly, people are not entirely comfortable sharing data due to a variety of reasons. It is also true that there are also technical and legal issues as well that present roadblocks in the collection and dissemination of data. Thus, government intervention in the form of regulatory mechanisms is essential for the full potential of NPD to be realized.

It is crucial to recognize that the collector of data is different from a custodian of the data. Independent custodians of data are being imagined all over the world, and attention is drawn to

[9] Assertion accepted as true for legal purposes, even though the assertion may not be true or proven.

the creation of data trusts in the UK,[10] and the need for legal recognition of institutional custodians of data. There are evolving methodologies to determine the mechanics of construction of these trusts and the specifications of their mandates. We mention that the Trust Laws in the UK can be used to set up Data trusts, and additional legislation is not required but there is some opposition to this view. The work of the Open Data Institute[11], founded in 2012 by Tim Berners-Lee and Nigel Shadbolt, is an early pioneer in open data that has been helping organisations steward data on behalf of others. The Ada Lovelace Institute[12] is another organisation, albeit a more recent one, that has a similar objective. In the US, academic research into the formation of data trusts has been initiated.[13] Also, law firms have begun piloting data trusts and tested out the concept and understand how to apply it in a business scenario.[14] This has enabled them to explore the identification of a business case and form a successful consortium, the necessary legal and ethical governance frameworks to enable data sharing, and understand the technologies needed to promote transparency and trust in the consortium. An industry effort on this front is the Emergent Alliance, a data collaboration initiative promoted by IBM and Rolls Royce to facilitate economic recovery post COVID-19 by using data sciences.

Larger global developments in data and internet governance also affect the way in which NPD is handled. It may be noted that a Data Act is being planned by the European Parliament that will introduce measures to create a fair data economy by ensuring better control over and conditions for data sharing for citizens and businesses. This will be achieved by facilitating data access and use, and also ensuring suitable legal protection of databases. Furthermore, ensuring fairness in the allocation of data value among actors of the data economy, including in business-to-business and business-to-government situations is also a key goal of the Act. A counter-argument in the European context has been that GDPR suppresses innovation, and the same may be said about any attempt to mandate data sharing through legislation.

At the same time, with different laws emerging in different countries and regions, there is fear that the Internet could be fragmented into at least three components—US, Europe, and China. This can possibly lead to significant lowering of the value of the Internet and its potential.

India's core approach as outlined in the NGD is pioneering. In its implementation, we can expect that India will also adapt the best practices from these emerging ideas elsewhere. However, an important theme that developed during the discussions is an especially critical requirement in India—furthering trust in these governance bodies and structures, an essential element of forming data trusts. There is also the issue of the availability of suitable technology for the data trusts and this is recognized as a chicken-and-egg problem. Technologies may evolve if there is a legal mandate but a mandate can be confidently made if there is a technology available, at least on the horizon. A good comparative example for this is the evolution of checks and balances in the management of large financial corporations. If these mechanisms are not in place, the risk holding responsibility of this data will remain undefined, and hence under utilized. It should also be kept

[10] https://datatrusts.uk
[11] https://theodi.org
[12] https://www.adalovelaceinstitute.org
[13] https://pacscenter.stanford.edu/research/digital-civil-society-lab/a-framework-for-data-trusts/
[14] https://hbr.org/2020/11/data-trusts-could-be-the-key-to-better-ai

in mind that the legal interface of data sharing with all types of intellectual property laws remains unclear.

For India, a possible way forward is a combination of investment of public funds, a legal mandate, and adopting an empowerment based data sharing approach to build trust in custodial institutions for NPD. The institutions and initiatives from the UK and the US also provide an important direction for the next steps in making NPD an economic engine.

Industry Responses

The concerns raised by the industry, from those that willing to discuss them,[15] are very similar to that seen in many of the publicly available discussion papers. For example, they raised questions on the ambiguity of concepts used in NGF, including the category of non-personal data itself, the meaning of the term 'community', or of 'community trustees'. Another set of concerns that the industry lawyers have raised is on the difficulty of governing only 'raw' data where, in practice, it is not often possible to distinguish between raw and processed data. Data driven organisations can legitimately lay claim to the proprietary nature of processed data; however, even current practices of data collection and processing may not allow a clear separation of raw data which is then subject to the kind of data exchanges being proposed. The degree of government intervention proposed in a hitherto unregulated space also raises concern; a major cause for concern is with the proposed decision-making being centralised in a single authority.

We conclude by noting that the industry representatives all acknowledged the value for public good that NPD could unlock, and has been unlocking. The COVID-19 pandemic, in particular, has made such instances particularly clear. However, clearly, as more data sharing takes place to respond to such crises, the more there is the need for secure and trusted institutions to facilitate the exchange.

[15] Many of the major data dependent service companies would not engage with us, even anonymously.

3. NPD Governance and Individual Rights

The following are the three main issues that will be discussed in this section.

1. The definition of NPD as adopted in the NPD Framework.
2. The means for, and the consequences of, introducing the proposed regulatory framework into the existing data governance landscape.
3. Ways in which greater agency and respect for individual and community rights can be incorporated into the proposed framework.

Definition of Non-Personal Data

What should 'Non-Personal Data' encompass? This is the key concern in many of the responses to NGF, both those that were submitted by the public and in the response papers that have been posted online. Thus, a common thread that appears when discussing the aforementioned issues is on the approach to defining this data category. The NGF originally defined NPD broadly to include all data without any personally identifiable information. This is further classified into two different categories depending on the origin.

1. Data that was 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.
2. Data which were initially personal data, but were later anonymised.[16]

These distinctions have been made while outlining the different governance mechanisms for personal and non-personal data. However, it needs to be pointed out that both share some common economic impetus in enabling seamless exchange of data within a regulated framework.

Public, individual, and community data are a further extension of this matrix, which will have to be thought through differently from a rights-based perspective.

The all-encompassing definition of NPD, coupled with the many determinations that need to be made on its categorization in accordance with NGF, raises many natural concerns about the regulatory burden that this would pose as well as its implications for privacy and data protection. For example, given the often blurred distinctions between personal, non-personal, and anonymised data in large datasets, it will be challenging if they are meant to be identified and treated differently according to different legal frameworks.

The first version, NGF-1, received wide attention and extensive stakeholder feedback. Thus it would be reasonable to assume that the preceding concerns have perhaps also been submitted as part of this feedback. The framework in NGF-1 has taken cognizance of the same in the revision

[16] Anonymisation is a technique in which the collected data is transformed via suitable processing techniques to the extent that individuals and specific events are no longer identifiable. Anonymisation techniques are an active area of research in the Information Theory community. Importantly, so is de-anonymisation, the reverse technique of identifying individuals and specific events from aggregate and transformed data. Thus, anonymous data today could become non anonymous in the future.

in NGF-2. Thus NGF-2 focuses more clearly on the kinds of data that justify the regulatory intervention, i.e., data that can be used for the public good. Fueled by the feedback the revised version defines more clearly the delineation between "public" and "private" NPD. It also added a dimension of demarcation of data based on the collecting entity—whether government or a private entity. It also introduces a wholly new concept of a "Data Business", a type of business that collects and manages both personal data and NPD, and adds attention to the issue of anonymisation. It must be mentioned here that the report also reflects on the need for high-quality India-relevant data sets to be made available in priority sectors to promote intensive data-driven research and development, and in the creation of public goods and services.

NGF-2, the revised version has two key features:

1. A more detailed analysis of the categories of data along with the rights.
2. Detailed management mechanisms for each category of data.

However, many believe that there remain open questions about the *basis* for regulation of various forms of NPD. For example, those who are familiar with large infrastructure companies pointed out, there is valuable raw data with these companies that is not generated or related to any human being and is considered proprietary information. The immediate question then is the following: To address the intellectual property concerns, is NGF suggesting an approach of acquisition of such data for the public good, perhaps using the 'right of way' lens to formulate the necessary regulation? Considering that the objective, and hence the emphasis, of both versions of the NGF is on the economic repercussions of NPD, this is a major gap to be bridged.

It must also be pointed out that there are key determinations to be made in the proposed sharing framework along multiple dimensions; for example, whether the data is business sensitive or relates to confidential information; whether it is anonymised data that bears as risk of re-identification; or whether it relates to national security or strategic interests. Statutory frameworks are required in establishing standards for these kinds of assessments, which will inevitably be done by multiple players. It is strongly suggested that the NGF follow the concept of "High Value Data (HVD) sets" as provided in Section 7.6 of NGF-2 and clearly define the categories which would be considered HVD. Similarly, it would be helpful to determine some inclusive examples as guidance in case of conflicting interpretations at a later point.

The NGF also introduces the concept of *sensitivity of NPD*. The sensitivity of data is determined by factors such as implications for national security, collective or individual privacy, and business interests. We believe a more detailed, if not precise, description of this concept should emerge from a vigorous public debate

Governance Frameworks

NGF-1 proposes to create an entirely new ecosystem within which the regulation of NPD is to occur. NGF-2 continues in the same vein, attempting to cull out the concept of NPD and all related data into its own ecosystem. Right from collection, categorisation to regulation, the Framework's obvious attempt has been to ensure that it comprehensively deals with any issues that may arise in relation to NPD.

The proposed mechanism in the NGF to regulate NPD is in keeping with India's overall approach to data governance. India has claimed to adopt a framework based on individual empowerment, one that is yet to be tested in case law. This approach of India is to be contrasted with the more *laissez faire* approach in the United States, the emphasis on procedural compliance in the EU, and the command and control structure in China. The same conceptualisation of data principal and data custodian is adopted in the NPD framework, possibly in an attempt at future harmonisation. However, two seemingly contradictory ideas seem to be at play in the NGF—one that locates the agency and interest in the individual, and another that emphasizes the non-personal nature of the data in question. Whether the data empowerment framework works in the case of non-personal data will have to be carefully thought through in ensuing policymaking in this area.

A key feature of NGF that has triggered significant discussions is that it has almost inevitably resorted to proposing the introduction of a new regulatory authority and process for non-personal data. This appears to have been inevitable given the multiple determinations to be made in the sharing and use of NPD. NPD has long existed prior to the deliberations on NGF by the Gopalakrishnan Committee and numerous other regulatory bodies and means have sought to regulate various avatars of NPD thus far. It is strongly suggested that having a new post-1991 type regulator is a recipe for more friction; and two kinds of friction at that—from the industry's perspective and from the perspective of the regulator. Ranging from the Competition Commission to the Telecom Regulatory Authority of India, there have been several ad-hoc categorisations created, with each body claiming jurisdiction over various types of NPD. While MNF-2 clarifies the roles of various entities it is creating, it has not yet clearly placed itself in the already existing ecosystems, or provided direction as to the resolution of jurisdictional conflict.

Individual and community rights

Individual and community rights need to be outlined in the NGF more precisely. The NGF's approach to community data is pioneering in several respects. Specifically, community rights over data, has not been discussed as extensively in government proposals in other parts of the world, and India's proposed directions on this aspect is of interest to observers around the world.

Data stewardship is one way of broadly thinking about these issues. Many believed that such stewardship would help unlock the value of data for individuals and communities, while empowering them to exercise greater agency in this process. Panelists examined the thinking in the report as a means of assessing whether it had managed to maintain the right balance between using NPD for economic purposes and the rights and interests of the individual and the community. It also recognises and provides for the concerns over safety of such data, endeavouring to provide agency to the community to control such data, often at odds with privately collected NPD datasets that are currently governed by contract. The European perspectives on the interaction of the GDPR with new lines of thinking on the Digital Single Market are instructive in this regard.

NGF-2 seems to have appreciated the stakeholder feedback, revising the portions on community data to detail the rights of the community. It also brings in the concept of data unrelated to any

data principal, providing that the community may exercise rights over it. However, a concern raised by the several that still remains is the 'top-down' nature of this design, which may not reflect the ways in which the communities being thought of actually tend to organise themselves.

It is recognised that both versions of the NGF have devoted a certain energy to ensuring agency to the community and identity involved in the NPD. NGF-2 contains an extensive appendix providing the framework for community data rights. While this is appreciated as commendable first steps, the concern largely remains as the conflicting community identities and creation of yet more institutionalised mechanisms (an issue covered later in this paper).

4. Data Economics for NPD

The discussions in the Panel series on this theme centered around the scope outlined in the framework, incentives, ownership, creation of value, the role of regulation, and data markets.

Data sharing for public good

The key premise of NGF is that sharing of NPD is in the public interest and should be encouraged. To this end the NGF makes several proposals that encourage NPD sharing.

The discussions started off with concern that the scope of public interest considered is probably not wide enough. Specifically, many aspects of public interest (particularly those of start-ups and of SMEs) seem to have been given a short shrift in NGF-2. Furthermore, innovative approaches to the use of data that serve wide ranging public interests are not considered. This view that NGF-2 considered a traditional and somewhat constrained view of how NPD is used and valued appears to be widely held and is a recurring theme in many fora. There also appears to be a sense of a climbdown in the ambition and scope of the NPD governance in NGF-2 as compared to NGF-1.

The following cautionary note on the unintended consequences of mandatory NPD sharing is made mandatory. Data custodians who are collecting NPD and providing derived metadata are expending considerable cost and must be encouraged and compensated. Any mandatory sharing requirement must be well-thought through and unintended consequences gamed out before implementation. It may be noted here that even in previous discussions, the need to carefully, and clearly, design the incentive structures are emphasised. Furthermore, in framing any regulation and incentive structure, the principle should be that the objective of NPD sharing should be in the best interest of the data principals and not necessarily controlled by the data holder (custodian). It is widely hailed that NGF is one of the first global efforts to recognize this and should be used as a global model.

Incentives for Creation and Sharing

The truism that there must be incentives for creation of data, otherwise there will not be any data to share is reiterated. Thus, any draconian or inflexible pricing model that unduly limits the benefits from data should be avoided. For example, the pricing model should take into account the cost of data collection as part of reasonable charges, otherwise data collection will be disincentivized. Business to Government data is an important concern of European Union and there must be strong and comprehensive incentives for the collection and sharing of such data. With the emerging digital economy in India, this is expected to be an important issue in India too.

The valuation of data is closely tied to the issue of incentives and compensation for the various actors in the data supply chain. There is consensus that the incentives must be reflected in a fair manner across the entire chain. All actors must be eligible for reasonable and non-discriminatory remuneration. For example, how do data trustees get compensated? Care must be taken in constructing incentives for data trustees who represent community interest. The maximum incentive for data collection will be from the data requester and if data trustees are dependent on

the data requester for compensation, this may result in misaligned incentives and objectives. Several cases of potential conflict of interest or areas for special focus can be identified. It would perhaps be important to lay out clear rules for such an identification.

Another consideration is compensation for public services generated based on community data. Community NPD should be free for all but we must ensure that the data custodian is adequately compensated or data will not be stored and served. Would the data custodian operate as a data business and charge the community for the services for the public? If so, is that in public interest even if the data service is used for public purposes?

The EU perspective is that there are corporations with huge data sets that could be of great public value but that these are not being shared. The ideal scenario would be to have a large-scale, voluntary data sharing regime which can be facilitated by reducing legal and technical barriers to data sharing. However, at this time, it is not clear if there would be sufficient non incentives for voluntary sharing. Hence, the EU policy is aimed to create a regime to enforce (or at least strongly encourage) such sharing.

Data Ownership

While not strictly an economic issue, the issue of data ownership is discussed extensively as it underlies the basis of a data economy.

It may be noted that German IP lawyers have pushed for new non-exclusive property rights for NPD.[17] However, since NPD has the properties of non-rivalry in the use of information and non-excludability in innovation it may be argued that such a new exclusive IP right is not needed. Furthermore, we should keep in mind that it is good economics to ensure that non-rivalrous goods are used as much as possible. Firms can use the same data and generate separate value, and therefore exclusive control of data might not be the best use of data.[18]

Data owners can effectively get de-facto exclusive property rights on it if they do not share the data and keep it a secret. De-facto exclusive data can work economically with the owner retaining exclusive control of data sets. (This may, of course, be in conflict with the need for NPD sharing described previously).

An important question that is raised often is about the ownership of inferred data (or the output data from a class of data processing procedures). Defining policies for raw data is relatively more straightforward than for inferred data and it is suggested that policy frameworks should encompass at least some forms of inferred data. Inferred data is often not within the purview of the same data fiduciary. A consensus view on this is that the legal framework is at an infancy and

[17] Kerber, W. (2016). Governance of Data: Exclusive Property vs. Access. IIC - International Review of Intellectual Property and Competition Law, 47(7), 759–762. doi:10.1007/s40319-016-0517-2

[18] The non rivalrous nature of data, even raw data, has been challenged in many fora. Time differentials in availability of data can result in differential advantages. Similar arguments about data not being non excludable are also being made in the literature.

much work remains to be done and much real-world experience needed to be gained before the framework could be matured.

There is also a consensus that data ownership rights would not be completely covered by the copyright and trademark laws and the associated enforcement regimes. Furthermore, trade secret laws also will not completely apply. Another view that has been put forward argues that it is in the public interest to share data that is of societal value, and the concept of eminent domain should be explored to mandate data sharing.

Creating value from data

Creating value from data is an important motif. We begin the discussion around the (in)famous statement from the Economist—"Data is the new oil." It is to be immediately pointed out that unlike oil, there are no markets for trading data, there are no data refineries, and there is no equivalent to the petroleum industry to derive value from the basic commodity of raw data. A prominent viewpoint is that the statement is simply absurd. It may be concluded that "Data is not the new oil—it burns up while data is reusable".

The issues of urban data are examined in this context. Many cities are aware of the heavy expectations around the value generation of their important asset—data. They know that the data is valuable but have no idea how to generate value. As a result, they hoard data believing that it is better to sit on a valuable asset rather than be branded a fool for giving away something of value.

Cities are being asked to think about the services they generate from their data in three tiers.

1. Data-as-service: This is the lowest value service and involves thinking of data as a base commodity. Data is priced by volume or through a subscription model. Access to data is metered and billed according to the pricing structure. As an example, some cities are starting to sell real-time raw data from sensors (e.g. air quality, traffic) to application providers who need real time access.
2. Insight-from-data: In this model, cities offer derived insight from data, rather than selling raw data. This was referred to as inferred data previously. As an example, a city may create a livability index that uses data from a plethora of sources, to assist city dwellers or developers.
3. Data-ecosystem: In this model, the city does not just create a simple stream of data but provides an entire platform including development kits, sample applications, etc, to assist application developers to create value. The assumption is that innovative application developers are often more capable of generating value than city officials.

In general there is consensus that the first and second models are limited in scope and are not likely to create a booming data economy. It is also observed that most regulators or legal observers think of data in those terms. In fact, the data ecosystem approach is far likelier to result in huge value but it is hard to regulate or even conceptualize a priori.

An important lesson from success stories is to explore non-traditional models for data valuation. Traditional pricing models have a tendency to be unduly restrictive. An example that can be

discussed in some detail is that of a parking company operating in New York which used public and private real-time parking availability to make recommendations to drivers trying to find parking spots for their automobiles. The start-up company is on the verge of becoming a highly valued "unicorn" generating value for its shareholders (which include public sector investors) and for citizens of the city. Thus, the providers of data will benefit enormously from their decision to share data, but not through a fee-per-use model but through a market valuation model that requires a willingness to take a shared risk with the start-up company.

Corporate data comparisons

We also mention here that private companies have been the most successful in generating value from data and it is worth exploring the models that they use.

A specific example is that of global telecommunications companies such as Verizon, Deutsche Telekom, and Telefónica that have monetized the customer call and presence data that they collect as a product of their operation. Till recently, they have used this data internally to achieve significant benefit by optimizing operations, customer segmentation, pricing optimization, etc. More recently, they have also used the same data externally, anonymized and aggregated, across various use cases for their B2B clients and partners by offering a wide variety of new revenue services such as:

- Traffic flow and density planning for ad agencies
- Fraud detection for financial institutions and credit card companies.
- Smart targeting and click-stream insights for brands and digital advertisers.

Another example of companies using data "exhaust" to create value is the US agricultural equipment company, John Deere. The company has built a data hub where it aggregates comprehensive agricultural data that it routinely collects from a variety of sources, such as telemetry and sensor data from its farm machines. It has connected the aggregated (and suitably anonymized) data with analytical tools and other sources of publicly available agricultural data to create a valuable asset for farmers and a major new source of revenue for itself along with a higher customer retention. This is an example of a win-win from clever use of data that previously would have been used for limited purposes within John Deere departments or often simply ignored. There is some more discussion of this in the next chapter.

NPD and Personal Data

The similarities and differences between NPD and Personal Data should be discussed in detail. A key concern is that Data Trustees will be authorizing the use of personal data to create NPD, often without a clear understanding of risks. The risk of individual privacy exposure is real and the use of techniques such as differential privacy and other forms of data anonymization are to be explored. The controversy around the sale of automobile registration data and resulting privacy exposures is to be viewed as a cautionary example.

Nevertheless, a strong view is that NPD data is to be shared as a default. Unless there is strong evidence that data must not be shared, the default should be to share. The burden of proof should not be on those advocating sharing but rather on those advocating not sharing. The only clear case for not sharing is when privacy rights of individuals are compromised.

Role of Regulation

Clearly, a cost benefit analysis must drive regulation. The universal view is to limit the role of regulators and light touch approach to regulation. Here, there are questions on the role of the regulator and real concerns about harms that can happen with an overly broad regulatory approach. There is a fear that regulatory overheads will become excessive. The issue of multiple regulators with overlapping jurisdictions possibly making life difficult for the different segments of the data businesses is reiterated. This becomes particularly relevant in the case of mixed datasets which are comprised of both personal as well as non-personal data. The personal data protection bill and the NGF1&2 state that mixed data sets will come under the ambit of the former. Since mixed datasets are frequently used for analysis and insights, the proposed scope of the NPD regulation may reduce significantly. Finally, it is strongly emphasised that all regulators should work together, though it is not clear as to how this would come about. Perhaps through MOUs between regulatory agencies, or through an integrated decision making process.

It is emphasised that excessive regulation is premature, particularly around pricing or valuation Any new regulators should limit their role to encourage data sharing. Anti-hoarding should be the only target of any new regulation. An example of NPD data hoarding against public interest is that of AQM data being hoarded to avoid impacting property values and upsetting property owners. It is further emphasised that there is no need for a separate NPDA regulator, and that this role should be subsumed in existing regulators by providing them with additional data sharing guidance.

The role of grievance redress rather than regulation should be discussed to deal with privacy and pricing abuse. A viewpoint is that redress has not worked well in India. There must be a clear picture on how NPDA can resolve conflict and there is a need to have grievance redress as part of the process.

Is a market for NPD possible?

The final question posed is whether a market for NPD is possible and likely to emerge?

There is optimistic unanimity that such a market is both possible and likely. One perspective is that it would certainly emerge, but the inhibitors could be information asymmetries and unduly concentrated market power. Another perspective is that the way to accelerate the market emergence would be to focus on incentives and mandates for data sharing, and less on governance and unnecessary regulation. A related view is to incentivize data sharing by starting with public data sets that are already available. Create centers for NPD data sharing and perhaps adopt a specific approach to create some success stories. There is general agreement that this issue is one of the critical aspects of a data economy. Overly aggressive and misdirected regulation could in

itself be part of the reason why this effort fails to take off, and this must certainly be guarded against.

5. Use Case Examples and a Technology Segue

In this section, we discuss some use cases and show how non-personal data (NPD) is used to create public good. A few have been mentioned in other communication outside of the panels but are summarized here. We also discuss a key technical component, a Data Exchange, that will play an important role in creating a Data Economy. This is done by presenting a use-case of an actual Data Exchange in operation, the India Urban Data Exchange.

Smart city

A prime example of the use of NPD is in smart city applications. Many such applications were explored during the India Urban Data Exchange (IUDX program).[19] Three illustrative examples are discussed below:

Transit: The use of GPS devices coupled with apps to show real-time location of city transit busses and trains has been a resounding success in many cities across the world. It is likely the benefit in saved productivity alone has been about US$100 Million per year in New York, USA. In smaller cities like Helsinki, Finland, a comprehensive investment in public smart infrastructure, has generated a huge amount of data[20] which has been used by public and private agencies to reduce travel times in the past 3 years by an estimated 30%, a significant step in a city where winter temperatures are well below freezing. India has actually taken this one step further by using real-time fare collection data to estimate current bus occupancy. In cities like Surat, bus customers can see whether or not seating is available in an arriving bus before attempting to board. This simple feature is credited with an increase of about 5% in ridership.

Solid waste: Many cities around the world have begun to see dividends from NPD data related to the collection. In India, the holy city of Varanasi, with its myriad gullies and alleyways, has created an application that uses data from garbage bin sensors, garbage cart and garbage truck GPS sensors, operator smart phones using accelerometers to estimate load, crowd sourced data from a Citizen engagement, and other sources. The application allows the city to dynamically manage waste pickup for operational efficiency and for higher citizen satisfaction, making it responsive to festivals etc., saving at least 15% in overall cost.

Safe city: Another major use of NPD data has been to improve citizen safety. The Indian city of Pune has created a citizen safety app that uses real time data from street-lights (e.g., providing information about their functioning), crowd density and crowd gender diversity from analytics on video feeds (such conclusions are drawn by analyzing the crowds for predominance of families or groups of young males), property records to know the nature of building usage (residential complex or bar), etc. Using this app, citizens can automatically find “the safest part” to walk through in Pune based upon current conditions.

[19] iudx.org.in
[20] helsinkismart.fi

Healthcare

The unprecedented COVID-19 crisis has taught us many lessons on improving the management of global public health and has disrupted many existing solution approaches and societal, technological, and regulatory norms. The critical resource underlying many proposed new solutions and approaches is data. We have seen the practical benefits of sharing accurate and trusted public-health non-personal data, namely data that does not violate any individual's privacy. To maximize global benefit, such data should be freely available for all, without excessive national or international legal restrictions on access or usage. New data-driven applications from public agencies and private sources correlate data, apply sophisticated analytics and inferencing, and are able to proactively address public health problems. These new techniques, together with the rapid availability of data, enabled the global community to identify the spread of COVID-19 in a matter of two weeks. In comparison, it took months in the previous outbreaks of Severe Acute Respiratory Syndrome in 2002, Ebola in 2013, and Zika virus in 2014.

In the panel series, and particularly the last panel, we examined a series of examples (or use-cases) where data-driven applications have shown the way to better manage disease outbreaks, improve chances of not catching a disease, or prevent outbreaks from turning into pandemics. Some typical health-care use-case examples are presented below.

Epidemic surveillance & management: These solutions provide a platform for predicting and managing high-exposure epidemic areas. They collect structured and non-personal datasets and enable analysis of complex correlations across demographic, migration, health, and weather dynamics sourced from government and international agencies, researchers, companies, and media. The dynamicity and timeliness of the sourced data vary across datasets. Some of the datasets like anonymized geotagged locations of cell phones or airport ticket data might be dynamic with real-time or near-real-time streaming. They also feed in static datasets like census and demographic statistics or simulated data like environmental predictors facilitating disease transmissions. A prime example of such a solution is BlueDot which has been used to detect severe outbreaks as much as ten days in advance, as seen in the COVID-19 outbreak[21] .

Tools for tracking pathogen evolution: Tools such as Nextstrain use data to enable tracking of pathogen transmission and evolutionary patterns to retrieve epidemic history from genomic data[22]. These platforms enable researchers to perform phylodynamic exercises using bio-sequences sourced from public repositories like NCBI, GISAID, ViPR, or GitHub. Nextstrain accelerated the process of sharing, mapping, and visualising the genomic data amidst the Coronavirus pandemic in close to two days. It took a year to complete these operations in Ebola

[21] Allam, Z., et al., 2020) Allam, Z., Dey, G., & Jones, D. S. (2020). Artificial Intelligence (AI) Provided Early Detection of the Coronavirus (COVID-19) in China and Will Influence Future Urban Health Policy Internationally. AI, 1(2), 156–165. doi:10.3390/ai1020009 https://www.mdpi.com/2673-2688/1/2/9 (last accessed on 14 December 2020)

[22] Hadfield, J., et al. 2018) Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R. A. (2018). Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. doi:10.1093/bioinformatics/bty407 https://academic.oup.com/bioinformatics/article/34/23/4121/5001388 (last accessed on 14 December 2020

and other previous outbreaks. Nextstrain's phylogenetic charts and family trees helped epidemiologists, policymakers, and medical practitioners dramatically accelerate the process.

Virtual repositories of biospecimens: Virtual repositories such as UK Biobank provide genotypic and phenotypic data from large population-scale studies. UK BioBank data is sourced from the genetic and clinical analyses of the medical samples donated by ~500,000 participants recruited across the United Kingdom from 2006 to 2010[23] . Virtual repositories such as UK BioBank and EuroBioBank have played a critical role in sharing such data for various forms of biological and medical research[24].

Integrated Command and Control Centres (ICCC): Integrated Command and Control Centers (ICCCs) which are the heart of smart cities have collected non-personal data from a wide variety of sources. ICCCs established in India's Smart Cities have been repurposed into COVID-19 War Rooms to tackle the Coronavirus outbreak. The datasets include details of COVID-19 diagnostic tests, Coronavirus hotspots, disease heat maps, geotags of home-quarantined people, lane closures and traffic data, audio files of helplines, surveillance camera footage, mobile questionnaires, etc. Some of these dynamic datasets are real-time while others are updated daily. These war-rooms have been essential in the fight against the scourge. Functionalities of ICCC's are based on the heavy stakeholder coordination between several municipalities and services. Similar crisis command centers are found in other parts of the world including the one at the State of Utah, USA[25] .

Smart Public Infrastructure: An example of health related smart infrastructure is the smart bus shelters being piloted for Seoul's travellers. Such shelters are equipped with IoT sensors and interconnected smart devices capable of collecting, transmitting, and delivering data remotely. COVID-19 oriented data-driven technologies auto-regulate their functionalities at the smart bus shelters, such as automated thermal-imaging cameras allowing persons with temperature below normal body temperatures, air-conditioning and sterilizing systems, sanitizer dispensers, bus arrival-departure schedules, automatic detection of crimes and fires[26] (Park, M., 2020). These sensors-based systems use de-identified dynamic data to enhance convenience for travelers. These systems also interact with the users in real-time, e.g., aids voice assistance to those who require it, provides suggestions on wearing a mask in a disease outbreak.

---

[23] Bycroft, C. et. al., 2018) Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature, 562(7726), 203–209. doi:10.1038/s41586-018-0579-z https://www.nature.com/articles/s41586-018- 0579-z (last accessed on 14 December 2020

[24] Mora, M., et al., 2014) Mora, M., Angelini, C., Bignami, F. et al. (2014). The EuroBioBank Network: 10 years of hands-on experience of collaborative, transnational biobanking for rare diseases. European Journal of Human Genetics, 23(9), 1116–1123. doi:10.1038/ejhg.2014.272 https://www.nature.com/articles/ejhg2014272 (last accessed on 14 December 2020

[25] Baird, R. P., 2020) Baird, R. P. (2020). How Utah's Tech Industry Tried to Disrupt Coronavirus Testing. The Newyorker. https://www.newyorker.com/tech/annals-of-technology/how- utahs-tech-industry-tried-to-disrupt-coronavirus-testing (last accessed on 14 December 2020

[26] Park, M., 2020) Park, M. (2020). Keeps out rain and COVID-19, Seoul tries smart bus shelter to fight virus. Reuters. https://uk.reuters.com/article/amp/idINKCN25A1U4 (last accessed on 14 December 2020

Private Sector

More interesting are models where companies have monetized non-personal data both internally and externally. Usually these are for improving their business operation but in some cases, the data is used to generate public good (and also company profit). Three interesting use cases are presented below.

Telecom carriers: An example is telecommunications companies such as Verizon, Deutsche Telekom, and Telefónica that have monetized the customer call and presence data that they collect as a product of their operation[27]. They have used the original Personal data internally to achieve significant cost reduction by optimizing operations. They are now creating NPD from this data, suitably anonymized and aggregated, for sharing externally across various use cases:

- With local governments, allowing city planners to design more effective traffic management systems and officials to better establish "smart city" technology solutions.
- For fraud detection for financial institutions and credit card companies.
- For city planners to allow better location, layout, and staff planning for stores, banks and other public facilities

Agriculture: John Deere, a US-based manufacturer of agricultural machinery, has used the data collected in its business to generate value for its farmer customers[28]. John Deere has built a data hub where it aggregates comprehensive agricultural data it routinely collects from a variety of sources, such as telemetry and sensor data from its farm machines. Much of this data would have previously been used for limited purposes within John Deere departments or often simply ignored. The platform also integrates public data sources, including soil type and weather. John Deere has connected the aggregated (and suitably anonymized) data with analytical tools such as Cornell Ag-Analytics. Farmers can use these tools to provide estimators for crop insurance, forecasts for yield and risk management, estimates on insurance, satellite vegetation imagery, real-time feeds on field conditions, and guides on conservation practices. The data hub would also be important for boosting supply-chain transparency between major food companies as a way to quantify progress and verify the practices farmers implement on the ground.

Energy: Energy and process industry companies (think refineries, hydroelectric dams, and other power-generating facilities) are reengineering to increase efficiency. To support this, one company, General Electric (GE), provides additional value to customers through data-based services that increase the efficiency of its machines[29]. GE delivers data about its systems and machines that makes predictive and prescriptive analysis possible for its customers around energy

[27] sloanreview.mit.edu/article/demystifying-data-monetization/
[28] digital.hbs.edu/platform-digit/submission/farm-to-data-table-john-deere-and-data-in-precision-agriculture/
[29] www.ge.com/digital/iiot-platform

use, maintenance, and other outcomes, allowing cost-reduction decisions by simplifying energy processes, leading to automation and operational efficiencies.

Data Exchange: Essential technical enabler for NPD sharing

We briefly take a technical detour and discuss a component that is essential for sharing disparate NPD data created by different data producers, namely that of a Data Exchange. The Data Exchange has been identified as a key component of the IndEA2.0 architecture[30]. In a traditional enterprise environment where all the data is under the control of a single data owner or producer, the tried-and-tested approach to data sharing is a central data repository (a data warehouse/lake). In environments with distributed data ownership, each with its own controls and terms associated with data usage, such an approach is not feasible. Each set of data has its own data access policies, as well as commercial, monetary or subscription aspects which must be observed. Thus, instead of breaking data silos by moving data en masse into a central repository, the data exchange approach interconnects the disparate and distributed entities without forcing data to be moved or copied. This provides a way for accessing data in a unified, common format, allowing for sharing of data between different departments in a city, as well as opening up data for third party developers to create innovative new applications and citizen services. In addition, there is an opportunity for third party providers of data, or third-party providers of data analytics or data annotation, to participate in what becomes a data marketplace.

IUDX as an case-study of a Data Exchange for Smart Cities

In late 2018, the Smart City Mission[31] in the Government of India, came together with the Indian Institute of Science, Bangalore to jointly develop and deploy the India Urban Data Exchange (IUDX)[32]. IUDX is now deployed as a production cloud service in 10 cities (Surat, Varanasi, Pune, Bengaluru, Chennai, Vadodara, Bhubaneswar, Bhopal, Agartala and Faridabad) and will shortly be deployed in several more. The intent is to make IUDX broadly available across all Indian cities in the next couple of years, with an ecosystem of collaborating partners. The development of IUDX is collaborative with contributors from various organizations.

IUDX facilitates secure, authenticated and managed exchange of data amongst various data platforms, 3rd party authenticated & authorized applications and other data sources, data producers and consumers, both within a city to begin with and scaled up across cities eventually at a national level, in a uniform & seamless way. Some of this data consists of streams of IOT data from installed sensors (e.g., Air Quality, Traffic, etc), some of the data is demographic or

[30] https://www.meity.gov.in/india-enterpise-architecture-indea
[31] http://mohua.gov.in/
[32] iudx.org.in

geographical, some may be from municipal tax or property records, some from legal documents or registrations, and some may be historical data from archival sources. The platform provides full control to the data owners as to what data to expose and to whom. Built-in accounting mechanisms connect with payment gateways which will form the foundations for a data marketplace. The whole platform is developer friendly, via definitions of open APIs (application program interfaces) and data schema templates (formats for interpreting data), so that a whole new application ecosystem gets created.

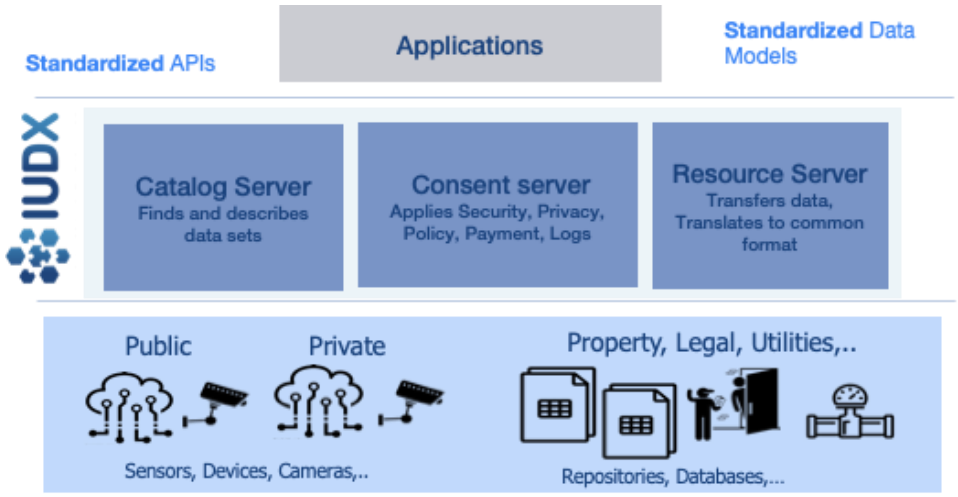

Figure: IUDX architecture overview

The IUDX platform (See Figure above) is based on interfaces and open APIs as described in the 'Unified Data Exchange Architecture' specifications'[33]. IUDX compliant applications will be able to use APIs to pull data from any of the underlying data platforms and using the publisher APIs to push data to any of the applications behind the individual platforms. Standardized APIs and data schema templates, will enable an IUDX compliant application to work in a city without needing any modification. Additionally, the standardized publisher APIs along with common data schemas, will enable vendor neutrality for IoT devices. In IUDX, there is a clear definition of data ownership and sharing mechanism, under the control of the data owner.

As shown in Figures above and below, IUDX consists of three main components:

- A Catalogue Server that allows applications to identify and locate pertinent data resources
- A Consent or Authorization server that validates data is accessed only by those authorized by data owner and consistent with the specified access policy

[33] https://bis.gov.in/other/USR_ICT_FSI_V_1_0.pdf

- One or more Resource Servers to provide data access and data ingestion through standardized API's and data models. Resource servers can either be part of the IUDX platform or reside outside the platform.

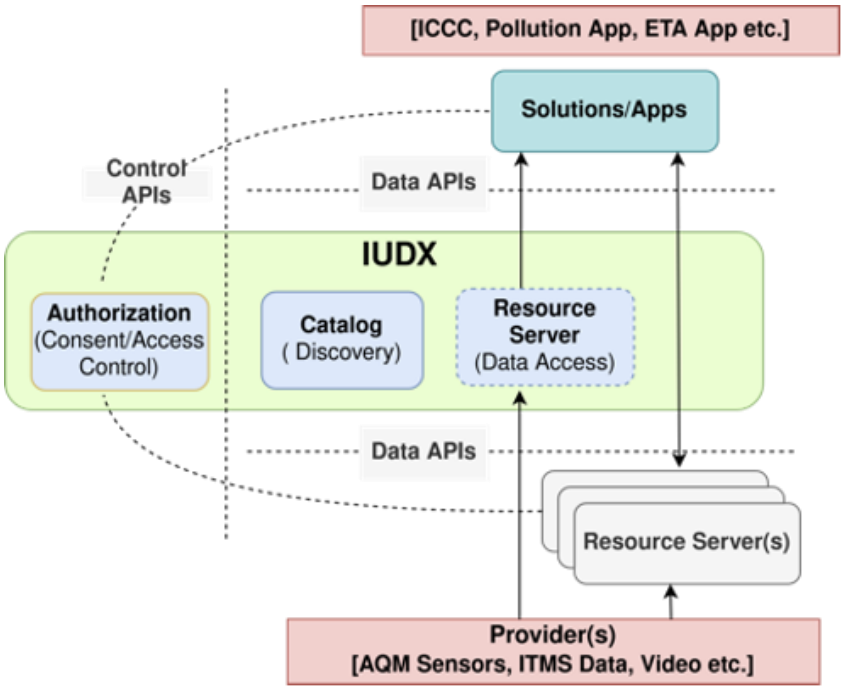

Figure: IUDX Components

Data resources managed by a Data Provider are hosted on the Resource Servers and a Data Consumer can access a data resource via open and standard data access APIs. Resource servers also provide publication services to enable data providers to ingest data from their respective data resources. A Data Consumer can discover the data resources relevant to its application using the search APIs provided by the catalogue service. The catalogue hosts information (e.g., data formats, units, type of the resource, etc.) for each data resource. This information is registered and managed completely by the provider of the given resource using open management APIs provided by the catalogue service. The meta-information, which is both human and machine understandable, enables the consumer to understand data and get additional context required for intelligent usage of this data. The Consent or Authorization server provides management APIs to register and modify access control policies associated with a protected data resource. The Consent Server also provides services to get access tokens in case the data consumer needs to access a protected resource. Using token validation services provided by authorization server, a resource server can ensure compliance to the data access policy set by the provider.

6. Concluding Remarks

We conclude with some remarks on the larger public discourse about NPD governance and on the way forward. These were not necessarily derived from the panel discussions but are the personal views of the three authors.

We begin by discussing the fallacy of innovation vs data sharing. We are concerned that a false dichotomy suggested in some discussions that increased mandatory data sharing will reduce innovation. This is reflected in the strong undercurrent against the framework presented in NGF on some limited sharing of community NPD. NGF-2 merely asks that the existence of certain categories of NPD be disclosed, but even this does not seem to be acceptable to many parties. They claim that this will inhibit innovation because the incentive to produce useful data and innovative applications will be reduced if the data is made available to all. Hence, they argue, voluntary sharing of NPD is adequate. They say, "Leave it up to companies and they will do the right thing. The government should not get involved in any coercive way."[34]

We argue that this is a fallacy for the following reasons:

- The data is collected for the mainline business and not for the ancillary public benefit. This is evident in the non-personal data being collected by social media, search, or ride-sharing companies. This data enables their core business and it does not impact the conduct of that business if the data is shared.
- Innovation often does not come from the data producer but from other parties. In most of our experience, the data producers are not the innovative companies that create new and innovative applications. For example, multi-modal transit applications are being created by innovative startups in cities like Bangalore and Surat. These companies could benefit greatly from some of the data being collected by ride-sharing companies and mobile operators, who would never have created such applications.
- Innovation requires data from different producers. Many applications require data from different sources to work effectively. For example, in Pune, an innovative start-up has created a citizen safety app that uses data from smart street-lights (which lights are on), video cameras (where are crowds gathered), property records (what is the nature of the buildings), crime records (where have crimes occurred), to compute the safest walking route for a citizen.
- Mandates apply to public as well as private sources. The mandates will apply to governmental and private sources of data. This will help ensure that data generated by central, state and local governmental and quasi-governmental agencies is shared. In smart cities, in particular, there is a plethora of innovative applications and services that can be created using such data.
- Publishing the existence of data can increase the monetization potential of the data. Far from decreasing the value of data, publishing its existence in a catalog will often increase the value by making more potential users aware of its existence. This is a bit like

[34] See, for example, https://thedataeconomylab.com/wp-content/uploads/2021/02/Comment-on-the-Revised-Report-on-NPD-Governance-Aapti-Institute.docx.pdf

advertising one's products to increase the customer demand. If the data is made available through a suitable data exchange, the provider can then enforce terms and conditions (monetary or otherwise) on the availability of the actual data itself.

Our experience has shown that sharing of high-quality data is the precursor to the creation of a data economy. Far from stifling innovation, it is the progenitor of new services and benefits for citizens and will empower an array of application developers and start-ups.

Valid concerns have been expressed in reactions to the NGF on the possibility of heavy handed regulation and conflicting regulatory structures. These should be taken into account when designing governance mechanisms, as should concerns of privacy and the right of the individual. However, opposition to a particular regulatory structure should not be conflated with opposition to the core concepts of NPD exchange.

The NGF has taken an innovative step in opening the door for significant economic activity and public good in the NPD space. It is essential, nay critical, that this be backed up by ramping up the research and development activity. Specifically, we suggest the following three pronged strategy to start off at the earliest.

- MeitY, DST, ICSSR and other funding agencies should initiate scholarly work on the technology, science, legal, and policy aspects and address the many concerns that are still open. The many entities that want to resist this bill are investing in making a case against NGF.
- Fund technology and marketplace development for exchange of raw and processed data. We specifically suggest that the technologies developed for IUDX is an excellent template and can be used for many other kinds of data. Of course, other technologies may also be independently developed.
- Make data available to researchers and for- profit organisations, subsidised or at cost, to develop the ecosystem. Many believe that startups and ideas have money but no data. A simpler way is to also ask the data analytics curricula that have sprung up to work on projects based on these data sets and make their outcome publicly available.

7. Acknowledgements

We would like to gratefully acknowledge the panelists who gave so generously of their time and expertise. We reiterate that the contents of this paper, while heavily influenced by the panels, are ultimately the sole viewpoints of the three authors. And we are particularly appreciative of the moderators of the panels, Udbhav Tiwari, Prof V. Sridhar, and Prof. Vijay Chandru, who so expertly guided the discussions.

We also acknowledge Rahul Patil and Asmita Verma of IISc Center for Society and Policy, and Ananya Das of NLS, who have provided valuable comments on this paper.

We also recognize the assistance of Namrata Aggarwal, Rohit Rai and Rajat Asthana in the organization of the panels that form the basis of this paper.