SPECIAL ISSUE:

NON-PERSONAL DATA: POLICY AND REGULATORY CONSIDERATIONS

May 2020 | *Issue No. 202*

ESYA centre

OVERVIEW

The past year saw significant developments in India's data protection landscape. The Personal Data Protection Bill (PDP Bill) was introduced in Parliament in December 2019, and is under consideration by a Joint Parliamentary Committee. It aims to set out the governance framework for personal data and establish a Data Protection Authority for the purpose. Non-personal data, which does not relate to the data of identified individuals, remains outside the ambit of the PDP Bill[1]. The Union Ministry of Information and Technology (MeitY) also constituted a committee under the chairmanship of Kris Gopalakrishnan (Committee) to suggest pathways for the regulation of Non-Personal Data (NPD)[2].

In this introductory brief we introduce the concept of NPD, trace related developments, and suggest the contours of subsequent research.

[1] For a brief overview of the Bill and its provisions, please refer: prsindia.org/billtrack/personal-data-protection-bill-2019.

[2] Terms of Reference of the Committee are available here: https://meity.gov.in/writereaddata/files/constitution_of_committee_of_experts_to_deliberate_on_data_governance_framework.pdf.

1/ POLICY DEVELOPMENTS

While the Committee is a step towards the creation of a regulatory framework for NPD, various official documents have touched on the issue previously:

- The Telecom Regulatory Authority of India, in its Consultation Paper on 'Privacy, Security and Ownership' alluded to the need for using data to allow better service delivery. It also sought to identify the various potential stakeholders in such an ecosystem[3].
- The 2018-19 Economic Survey made an emphatic case for the use of NPD stored with the Government as a public good, by breaking silos and allowing easier data sharing among Ministries and Departments[4].
- The Draft National E-Commerce Policy 'acknowledges the importance of data as an asset and identifies the means to protect data generated in India, enhance data security, prevent violation of privacy and create domestic standards for devices which are used to store, process and access data'[5]. It also outlines data localisation measures, to ensure that data generated by citizens is used in the national interest.
- The Justice Srikrishna Committee, which preceded the PDP Bill, touched on the concept of 'community data', described as data sourced from multiple individuals. That committee advocated a legal framework to protect the privacy rights of individuals who contributed to such 'community data'[6].

These developments shed light on the reasons for regulating NPD, which could include:

- Harnessing and controlling data as an economic resource, especially given the success of data-based business models around the world.
- Using NPD to make evidence-based policy and facilitate public service delivery.
- Confining data flow within national borders in order to focus the benefits derived from it to Indian companies, and thereby Indian consumers.
- Ensuring the privacy of individuals and groups who have contributed to community data is given adequate protection.
- Levelling the playing field between large multinational companies and small Indian businesses.

Many considerations need to be taken into account in regulating NPD, however, some of which are outlined next.

[3] TRAI Consultation Paper on Privacy, Security and Ownership of Data in the Telecom Sector; Chapter V, Questions 5,6,7 and 9. Available here: https://main.trai.gov.in/sites/default/files/Consultation_Paper%20_on_Privacy_Security_ownership_of_data_09082017.pdf.

[4] Economic Survey 2018-19, Chapter IV, p. 12, accessible at: https://www.indiabudget.gov.in/budget2019-20/economicsurvey/doc/vol1chapter/echap04_vol1.pdf.

[5] Draft National E-Commerce Policy, accessible at: https://dipp.gov.in/sites/default/files/DraftNational_e-commerce_Policy_23February2019.pdf.

[6] Justice Srikrishna Committee Report, p.45, accessible at: https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf.

2/ KEY ISSUES IN THE REGULATION OF NPD

A. DEFINING NON-PERSONAL DATA AND ENSURING PRIVACY

Taxonomy

NPD can mean in general any data that does not relate to, or cannot be used to identify an individual. However, an official definition will play a crucial role in future NPD regulation. The Committee's terms of reference stress the need to determine a suitable taxonomy of data, listing 'aggregated data, derived data, anonymous data, e-commerce data etc.' as possible classifications[7].

The European Union (**EU**), which is one of the few jurisdictions that regulate NPD, defines it as any data that is not personal data[8]. Another approach is to explicitly define what constitutes NPD, and create sub-classifications for clarity. For instance, anonymised health data could receive different safeguards and protections from commuter traffic data. NPD could be classified as commercially sensitive, such as data submitted by companies to the Ministry of Corporate Affairs. Another possible classification is of data pertaining to or generated by a human, or by non-human sensors or machines[9].

Mixed Datasets

Although data may be split into personal and non-personal, in practice most datasets comprise both. In datasets where both forms of data can be identified and separated, the regulations governing each form may be applied with ease. However, where the two are inextricably linked, such separation becomes difficult: for instance the datasets used by banks, especially those with privileged client information, often consist of data concerning individuals (personal) as well as companies (non-personal).

The EU deals with the issue by specifying that if the two forms of data are *inextricably linked*, then the entire dataset is to be treated as personal data[10]. It is unclear what standard will be used to determine whether data are inextricably linked. It may be possible to separate the two forms of data in a mixed dataset, but prohibitively expensive. Are the data in such an instance inextricably linked?

Privacy

Although it does not contain any identifiable personal data, there are various privacy concerns associated with NPD. The Srikrishna Committee Report asks the Government to consider introducing a legislation that would facilitate privacy protection for both community data and corporate data[11]. But as seen above, certain NPD can contain sensitive information unrelated to an individual. Such NPD would then require safeguards to ensure privacy.

Scenarios are also possible wherein legislations dealing with personal data and with NPD are applied in parallel, with each undermining the other[12]. For example, companies may seek to classify certain data as personal to evade sharing obligations laid down by law[13]. In this context it is important to define the interaction between privacy regulations and NPD.

[7] *Terms of Reference of the Committee are available here: https://meity.gov.in/writereaddata/files/constitution_of_committee_of_experts_to_deliberate_on_data_governance_framework.pdf.*

[8] *Article 3(1), Regulation (Eu) 2018/1807 Of The European Parliament And Of The Council, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1807&from=EN.*

[9] *Anubhutie Singh et al., The Contours of Public Policy for Non-Personal Data Flows in India, Dvara Research. September 24, 2019. Accessible at: https://www.dvara.com/blog/2019/09/24/the-contours-of-public-policy-for-non-personal-data-flows-in-india/.*

[10] *Article 2(1), Regulation (Eu) 2018/1807 of the European Parliament and of the Council.*

[11] *Justice Srikrishna Committee Report, p.46, accessible at: https://meity.gov.in/writereaddata/files/Data_Protection_Committee_Report.pdf.*

[12] *Inge Graef at al., Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Innovation, TILEC Discussion Paper, September 2018, accessible at: http://ssrn.com/abstract=3256189.*

[13] *Martin Husovec et al., Feedback to the Commission's Proposal on a framework for the free flow of non-personal data, accessible at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3106791.*

Concerns over standards of anonymisation —the removing of all information that could be used to identify or trace an individual from a given dataset[14]—are also likely to arise. The process of anonymisation must be irreversible, as the PDP Bill envisions, but ensuring irreversibility has proven impossible thus far. This is because even anonymised datasets contain certain attributes or characteristics[15] which can be traced back to individuals by linking them with other datasets in the public domain[16]. Research suggests that even sampled anonymised datasets are unlikely to meet the requirements of the EU's General Data Protection Regulation[17] [18], which adopts a standard of anonymisation similar to the PDP Bill.

B. OWNERSHIP AND ACCESS

Rights of ownership of and access to NPD are important to consider in future regulation. Certain datasets enjoy protection under the Indian Copyright Act of 1957[19]. Copyright protection is justified on grounds of the creativity or originality that goes into selecting or arranging the contents of a database[20]. Indian Courts have usually used the 'sweat of the brow' test, or the devotion of time, capital, energy and skill, to determine whether a database is creative or original[21].

The extension of copyright to sets of machine-generated data is another uncharted area[22]. Jurisdictions including Germany, Singapore and Japan extend copyright protection only to human-authored works. However, the increasing creation of original works by artificial intelligence (AI) and associated technologies have led to growing demands for protection of these works by copyright[23].

Furthermore, data-driven companies may wish to restrict the use of data for commercial reasons[24]. Any regulation of NPD must seek to address questions on ownership and access, including who owns the data, who has the right to benefit from it, who has the right to access such data and for what purposes, and how to incentivise data sharing[25].

Various models have been conceptualized to resolve the above questions. Some of there are:

- regulating the sharing and use of NPD through existing contract law and intellectual property rights (**IPRs**);
- creating a new right in data, either in a form similar to IPRs[26] or a sui generis data producer's right[27];
- creating a system of self-regulation for the sharing of information, wherein interested stakeholders agree on the kind of data to be shared, and the process for such data sharing[28].

Any decision on ownership, access and sharing of benefits in this context must account for the need to encourage competition and innovation, which is examined below.

[14] S. 3(2), Personal Data Protection Bill 2019.

[15] Thomas Brewster, 120 Million American Households Exposed in 'Massive' ConsumerView Database leak, Forbes December 19, 2017, accessible at: https://www.forbes.com/sites/thomasbrewster/2017/12/19/120m-american-households-exposed-in-massive-consumerview-database-leak/#412e451f7961.

[16] Gina Kolata, Your Data Were 'Anonymized'? These Scientists Can Still Identify You, The New York Times, July 23, 2019, accessible at: https://www.nytimes.com/2019/07/23/health/data-privacy-protection.html?smid=nytcore-ios-share.

[17] Luc Rocher et al., Estimating the success of re-identifications in incomplete datasets using generative models, Nat Commun 10 3069 (2019), accessible at: https://www.nature.com/articles/s41467-019-10933-3.

[18] Recital 26, GDPR, accessible at: https://gdpr.eu/recital-26-not-applicable-to-anonymous-data/.

[19] s.2(o) of the Indian Copyright Act, 1957 includes 'computer database' under the definition of literary work.

[20] Apar Gupta, Protection of Databases in India: Copyright Termination Sui Generis Conception, Journal of Intellectual Property and Practice 8:2 553, 2007, accessible at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1682505.

[21] Shyam Lal Paharia v. Gaya Prasad Gupta Rasal, 1971 A.I.R. 58 (All) 192, 195, 199; Gangavishnu Shrikisondas v. Moreshvar Bapuji Hegishte, I.L.R. 13 (Bom.) 358.

[22] Wolfgang Kerber, Governance of Data: Exclusive Property vs. Access, IIC - International Review of Intellectual Property and Competition Law 47, 759–762 (2016).

[23] Andres Guadamuz, Artifical Intelligence and Copyright, WIPO Magazine, October 2017, accessible at: https://www.wipo.int/wipo_magazine/en/2017/05/article_0003.html.

[24] European Commission Communication on European Data Economy (n 24), 9.

[25] Ivan Stepanov, Introducing a property right over data in the EU: the data producer's right – an evaluation, International Review of Law, Computers & Technology, 34:1, 65-86 (2020), accessible at: https://www.tandfonline.com/doi/pdf/10.1080/13600869.2019.1631621?needAccess=true.

[26] Wolfgang Kerber, A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis, Joint Discussion Paper Series in Economics No. 37-2016, accessible at: http://www.uni-marburg.de/fb02/makro/forschung/magkspapers.

[27] Samson Yoseph Esayas & Angela Daly, The Proposed Australian Consumer Right to Access and Use Data: A European Comparison, Eur. Competition & Reg. L. Rev.2, 187 (2018).

[28] Heiko Richter and Peter Slowinski, The Data Sharing Economy: On the Emergence of New Intermediaries, IIC - International Review of Intellectual Property and Competition Law 50, 4-29 (2019).

C. COMPETITION AND INNOVATION

The digital economy is replete with competition concerns, because of various barriers to entry and also as a result of advantages gained due to pooling or mergers of datasets.

Barriers to Entry

Markets reliant on access to large volumes and variety of data can have technical, legal or economic barriers to entry[29], leaving new entrants unable to collect or access relevant data for the following reasons:

- Restrictions on data access as a result of technology standards. These include different standards of encryption and lack of interoperability of structured data[30].
- Legal barriers in the form of contracts restricting data transfer. IPRs such as copyrights and trade secrets can be used to protect large datasets which restricts access.
- Network effects and economies of scale can entrench dominant players[31]. Network effects refer to the fact that a larger number of users is likely to improve the value of a product or service[32]. Companies such as Google and Facebook enjoy large user bases and are able to gather insights from them to further improve these services. This can affect the ability of other players to compete.
- The scale (volume) and scope (variety) of data collection and processing can lead to improvement in algorithms and faster experimentation. This increases efficiency in the process of production and service delivery[33]. At the same time, economies of scale and scope can act as barriers of entry as they favour incumbents, who are likely to possess larger and more varied datasets.

Pooling of Datasets

The sharing of commercially sensitive information between competitors can hinder competition and lead to unfair trade practices[34]. The acquisition of datasets as a result of mergers and acquisitions between companies may also result in the acquirer gaining a dominant position in a particular market[35].

Yet the pooling of datasets can prove beneficial by enabling better insights and innovation, a greater quality and quantity of products, and more competitive prices[36]. Data moreover is the key input for numerous forms of artificial intelligence, such as machine learning and deep neural networks. These technologies use vast amounts of data to 'learn' how to identify patterns, recognise objects and make predictions. Given the potential of AI related technologies to revolutionise production processes and supply chains, the competitiveness of firms will increasingly be determined by timely access to relevant data[37].

[29] Jay Modrall, A Closer Look at Competition Law and Data, Competition L. Int'l 31 (2017).

[30] Thomas Tombal, Limits and Enablers of Data Sharing, An Analytical Framework for EU Competition, Data Protection and Consumer Law, TILEC Discussion Paper DP2019-024, accessible at: http://ssrn.com/abstract=3494212.

[31] Prüfer and Schottmuller, Competing with Big Data, CentER Discussion Paper 2017-007.

[32] OECD, The Digital Economy, new business models and key features, Chapter IV in Addressing the Tax Challenges of the Digital Economy 2014, accessible at: https://www.oecd-ilibrary.org/docserver/9789264218789-7-en.pdf?expires=1587827093&id=id&accname=guest&checksum=FF728501CEF29D5CB02539C1ACB3EFDF.

[33] Commission Decision of 11 March 2008, Case M.4731 Google/ DoubleClick, para. 273.

[34] B. Lundqvist, 'Competition and Data Pools', Journal of European Consumer and Market Law 146, 2017.

[35] Inge Graef , 'When data evolves into market power – data concentration and data abuse under competition law', in M. Moore & D. Tambini (Eds.), Digital Dominance: Implications and Risks, Oxford University Press 2018.

[36] Press release European Commission, 'Antitrust: Commission opens investigation into Insurance Ireland data pooling system', 14 May 2019, available at https://europa.eu/rapid/press-release_IP-19-2509_en.htm.

[37] European Commission, Competition Policy for the Digital Era 2019, accessible at: https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf.

This document has been prepared by Mohit Chawdhry, a Research Assistant at the Esya Centre.

For any further contact, please get in touch with us at:

www.esyacentre.org

ESYA centre