

4th Dvara Research Conference Regulating Data-Driven Finance April 5-6, 2019

Primer on Consumer Data Infrastructure

Malavika Raghavan, Beni Chugh & Anubhuti Singh

This primer presents a broad overview of the landscape, emerging concerns and some considerations for regulators to set the context for discussions around consumer data infrastructure in India. A reference list of background reading materials is available on page 7 & 8.

This primer has been prepared for participants at the 4th Dvara Research Conference.

Overview

The emergence of large-scale data infrastructures in India to support financial services delivery (such as the proposed Public Credit Registry, the GSTN system, and the Indiastack) raises foundational questions regarding the role of the State and of the regulation of these infrastructures. These data infrastructures include the databases that aggregate information and data utilities that allow for various databases to exchange information. They seek to re-imagine and improve the credit information markets in India with roles for private and public information-collecting entities. They also seek to bring the benefits of the aggregation of such data to improve financial inclusion and supervision of the financial sector.

The impact for individuals and firms whose information will be aggregated and shared needs to be understood. Consumer data forms the backbone of the information that will be passed through these infrastructures. This requires an understanding of the issues that arise from the interplay between the regulation of data infrastructures and the regulation of personal data. The impact of these new projects for existing financial infrastructures and databases in the country also need to be understood.

These developments raise new considerations for regulators. The approach to the regulation and supervision of these databases and infrastructures is by no means certain in the Indian context. The effect they will have on the financial sector and the wider market, and accordingly the optimal stance for regulation remains to be understood. Finally, there is a need for clarity on the liability framework for entities accessing these infrastructures and those in charge of running the infrastructures especially in cases where harms or losses are suffered due to their misuse, breach or failure.

1. Landscape

As new methods of generating, collecting and processing individuals' personal information proliferate, the robustness of the technology ecosystem through which this data flows becomes an important concern for all stakeholders. Data infrastructures can aggregate massive amounts of consumer data and enable data sharing across databases through a variety of mechanisms. The importance of strong data infrastructure to support the development of financial markets is well-recognised in India. These infrastructures hold out the promise of improving efficiency in the financial system, aiding supervision and early identification of systemic risks and instabilities. They also have the potential to create utilities and services that can harness the power of data to widen, deepen and improve the quality of access to finance for individuals and SMEs. Better data about individuals and firms can play a large role in correcting the information asymmetries that act as barriers to accessing finance. It can also play an important role to tailor financial services so that they are truly suited to consumers' unique circumstances.

The Indian financial sector already has several pieces of data infrastructure to assist in data sharing for various objectives. Four private credit bureaus (TransUnion CIBIL, Experian, Equifax, and CRIF High Mark) exist in India, collecting different kinds of information to help build out the credit information market over the last two decades.

Databases also exist to aide in the supervision of the financial sector. For e.g. the RBI's Central Repository of Information on Large Credits (CRILC) is a borrower level dataset of systemically important exposures (of INR 50 million and above) held by Indian banks. Separately, the RBI's Basic Statistical Return -1 database (BSR-1) contains information of all loans (without borrower identification) across the country, to give the regulator insight into the distribution of credit in the country.

In the recent past, we have witnessed the creation of data-driven infrastructures to act as utilities for the financial sector. For instance, [in 2015, the Government of India set up the Central Registry of Securitisation Asset Reconstruction and Security Interest \(CERSAI\)](#) to act as a Central KYC Records Registry for the entire financial sector. It aims to create a centralised repository of KYC records that can be shared by entities across the financial sector without the need for the consumer to undergo the KYC every time they seek to create a new relationship with another financial entity. Another prominent public infrastructure with relevance for widening financial access includes the Aadhaar database and related IndiaStack Application Programming Interfaces (APIs). These APIs aim to lower barriers to entry for private firms seeking to offer services to consumers, especially those who were harder-to-reach or previously excluded. For instance, the e-KYC layer of the IndiaStack¹ was launched in 2013 by the Unique Identification Authority of India (UIDAI) as

¹ The [e-KYC API of the Indiastack](#) enables the establishment of the identity of an individual, and the authentication of this identity against their records in the Aadhaar database electronically in a paperless manner by service providers either (i) remotely, using a One Time Passcode delivered to a linked mobile phone or (ii) in person, using biometric authentication.

information infrastructure to overcome barriers faced by financial sector entities when on-boarding consumers due to lack of adequate documentation or ability to authenticate identity. In 2016 the [Unified Payments Interface \(UPI\)² layer of the Indiastack](#) was created by the National Payments Corporation of India (NPCI), a not-for-profit set up to provide infrastructure for the entire banking system in India for payments and settlement systems. These kinds of infrastructures aimed at harnessing personal data, financial information and technology to streamline existing financial processes (such as customer verification and identification, and completing payments).

Financial sector regulators are also beginning to create specific regulatory and technical infrastructure purely for data aggregation and sharing. In September 2016, the RBI recognised a new class of Non-Banking Financial Companies, i.e. [Account Aggregators](#), exclusively dedicated to collecting, retrieving and sharing users' financial information with other financial entities with the users' consent. In June 2018 the RBI issued a blueprint for a [Public Credit Registry](#) which will consolidate various types of financial and non-financial data about individuals and entities currently held across different financial and non-financial sector databases.

Other sectors and regulators are also creating publicly backed data infrastructures. For instance, Information Utilities (IUs) have been envisioned under the [Insolvency and Bankruptcy Code](#), 2016 as specialised, public companies, created to provide verifiable financial information regarding the liabilities of an entity to assist in insolvency and resolution proceedings. The National e-Governance Services Limited ([NeSL](#)) was India's first IU that became operational in September 2017. Rich data about firms is held and updated by the Goods and Services Tax Network (GSTN) based on their GST returns including on GST paid, value of products, sale and purchase details etc.

These examples reveal the proliferation of data infrastructures in service delivery by public bodies in India. Private firms in the market, from SMEs to large multinationals and BigTech firms also have extensive infrastructure to support the collection, linking, storing and sharing of information. In terms of regulatory requirements, skeletal rules under the Information Technology Act, 2000 specify basic requirements to be followed by entities upon collection of personal data. There is also an uneven patchwork of requirements across various financial sector regulations on managing client data. There is limited enforcement and external oversight of these data infrastructures.

As the roles of these infrastructures become increasingly central to financial markets, several concerns and considerations emerge for regulation. The unprecedented scale of recent data infrastructures raise implications for the financial market, as well as the individuals and firms whose information will be aggregated and shared.

² The UPI API is built to enable all bank account holders in India to make payments to other accounts across banks instantly on their smartphone without the need to enter detailed bank transfer information.

2. Emerging Concerns

Implications for financial markets and competition: Globally, the impact of large-scale aggregation and data sharing on existing information markets and connected financial markets is still being understood. The design of data infrastructures is one of the factors that will impact these markets and the firms within them.

In India, these infrastructures are being erected in a changing financial services landscape where several new types of entities and firms are disintermediating traditional institutions and processes. These may drive healthy competition among entities who may not have previously competed. For instance, several products have been created to enable digital payments through the UPI infrastructure but three very different entities **currently have the largest market share**: PayTM (a Payments Bank, previously an e-wallet), GooglePay (a BigTech company) and PhonePe (an Indian fintech). On the flip side, there is potential for these infrastructures to erode existing infrastructures or markets. For instance, the impact of the proposed Public Credit Registry (PCR) on credit bureaus and the developing credit information markets in India is unclear, and evidence of the impact of their interaction from other countries is mixed (Policy and Economic Research Council, 2018).

Risks and harms for consumers from the aggregation and sharing of data: Individuals and firms can be exposed to a range of risks and harms if their personal data is misused or compromised in large data infrastructures. Individuals can be exposed to risks (i) immediately due to the compromise of their data (such as losing privacy, and being exposed to identity theft), and (ii) also on a continuing basis, if the data is used against them for unauthorised transactions or fraud, to discriminate against them etc. A strong data protection regulatory framework is seen as a solution to protect consumers, but enforcement frameworks to realistically protect users are still emerging. There is a need for the technical architecture of any data infrastructure to be built in a way that mitigates these risks, or to re-consider features where the risks outweigh the benefits.

Data security and data quality: The security of these large data infrastructures has been a major source of concern in recent years. Reports of data security issues in crucial public infrastructures (including **the Aadhaar system**) as well as breaches of large private infrastructures (such as **data breaches of Equifax**, the credit reporting agency) have revealed their vulnerabilities. Centralisation of data also raise the risks of a single point of failure, creating the need to consider alternative strategies for aggregation. Finally, compromises of data in one database can have implications for the data quality in related or linked databases. These concerns must be addressed as they strike at the heart of the benefits that large data-sharing systems have to offer.

3. Considerations for Regulators

Regulators must engage with these emerging concerns holistically. Doing so will require an understanding of the interplay between several fields including financial systems and regulation, competition, data protection and network security. Some emerging considerations for regulators are noted below to open deliberations in this session.

Aligning technical deployment and policy development: The experience of large data-infrastructures in India and globally has shown the need for regulatory thinking to be developed alongside technical architecture. Failure to do so can result in regulatory guarantees that may be technically unfeasible and consequently meaningless. It could also create systems without sufficient safeguards and governance, that could expose consumers and the entire system to risks. Early alignment on design principles of data infrastructures, such as access controls, purpose limitation and clear liability frameworks would be essential to create secure and trustworthy data infrastructures.

Public and private sector development of data infrastructure: Each new sphere within which data infrastructures are created must consider the context and effects of the balance of public and private sector involvement in building infrastructure. The benefits of public sector leadership in creating data infrastructure must be carefully weighed up in each situation against creating the space for a market to develop for private firms to create and deploy infrastructures.

Considering Competition, Data Privacy, Systemic Risk and other regulatory imperatives: Regulators will need to reconcile other regulatory imperatives for consumer data privacy protection, managing systemic risk, and competitiveness. A loss of consumer privacy could have adverse consequences for consumers' access to finance and their wider lives. The potential for data infrastructures to compromise privacy must therefore be offset by mechanisms to mitigate these risks. Likewise, issues of competition and market power need to be considered and aligned with the development of data infrastructure. Systemic risk concerns must also be addressed. For instance, the Basel Committee on Banking Supervision (2018) has raised the interdependencies between market players and market infrastructures, could cause IT risks to escalate into systemic crises especially if some firms do not have expertise or experience in managing such risks.

Against this context, this session on **Consumer Data Infrastructure** for data-driven finance could examine the following questions:

- What are the obstacles faced in the exchange of data within the financial system currently? Are public utilities an efficient way to resolve the obstacles of financial data exchange?
- What are the key opportunities and risks for finance from the creation of large-scale data infrastructures for the country?
- What are the design principles that should guide the regulation of data infrastructure? Should public and private data utilities be held to different design and regulatory principles?
- What are the lessons from the last decade of the financial and data infrastructures that should guide considerations for future large-scale projects?
- What are the likely effects on competition within the financial services landscape from the existence of these infrastructures? Will they create welfare gains for consumers?
- What is the interplay between the regulation of consumer data and the regulation of data infrastructures?

Resources

1. Bank for International Settlements. (2018). Sound Practices: Implications of fintech developments for banks and banking supervisors. Basel: BIS. Retrieved from: <https://www.bis.org/bcbs/publ/d431.pdf>
2. Consumer Financial Protection Bureau. (October, 18 2017). Consumer-authorized financial data sharing and aggregation. Retrieved from Consumer Financial Protection Bureau: https://files.consumerfinance.gov/f/documents/cfpb_consumer-protection-principles_data-aggregation_stakeholder-insights.pdf
3. Directorate for Financial and Enterprise Affairs Competition Committee. (2018, November 28). Quality Considerations in Digital Zero-Price Markets. Retrieved from Organisation for Economic Co-operation and Development: [https://one.oecd.org/document/DAF/COMP\(2018\)14/en/pdf](https://one.oecd.org/document/DAF/COMP(2018)14/en/pdf)
4. Hurley, M., & Adebayo, J. (2016). Credit Scoring in the Era of Big Data. Yale Journal of Law and Technology, 148-216. Retrieved from <https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=1122&context=yjolt>
5. Independent Expert Advisory Group on a Data Revolution for Sustainable Development. (2014, November). A Word That Counts. Retrieved from UN Data Revolution: <http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf>
6. Policy and Economic Research Council. (2018). The Case for a Public Credit Registry in India: Additional Evidence for Consideration. Durham: Policy and Economic Research Council (PERC). Retrieved from: http://www.perc.net/wp-content/uploads/2018/03/India_PCR.pdf
7. Parsheera, S., Shah, A., & Bose, A. (2017, April 3). Competition Issues in India's Online Economy. Retrieved from National Institute of Public finance and Policy: https://www.nipfp.org.in/media/medialibrary/2017/04/WP_2017_194.pdf
8. Reserve Bank of India. (2014, January 31). Report of the Committee to Recommend Data Format for Furnishing of Credit Information to Credit Information Companies. Retrieved from Reserve Bank of India: <https://rbidocs.rbi.org.in/rdocs/PublicationReport/Pdfs/APR220314FS.pdf>
9. Reserve Bank of India. (2018, April). Report of the High Level Task Force on Public Credit Registry. Retrieved from Reserve Bank of India: <https://rbidocs.rbi.org.in/rdocs/PublicationReport/Pdfs/PCRRR09CF7539AC3E48C9B69112AF2A498EFD.PDF>
10. Sakle, A., & Chand, A. (2018, May 18). Big Data: Emerging Concerns under Competition Law. Retrieved from Cyril Amarchand Mangaldas Blog: <https://competition.cyrilamarchandblogs.com/2018/05/big-data-emerging-concerns-competition-law/>

11. Schwartz, P. (2010). Data Protection Law and the Ethical Use of Analytics. Retrieved from The Centre for Information Policy Leadership:
https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/data_protection_law_and_the_ethical_use_of_analytics__paul_schwartzwhite_paper_2010_.pdf
12. Vora, H. (2019, February 21). Opinion | Why India needs to set up a public credit registry. Retrieved from Livemint:
<https://www.livemint.com/opinion/online-views/opinion-why-india-needs-to-set-up-a-public-credit-registry-1550773274423.html>