ISSN 2321 – 9378

IMI Konnect

Volume 10, Issue 4, 2021

An IMI Kolkata Publication

Special Issue

Managing in a Data Driven World

About *IMI Konnect*

IMI Konnect (ISSN 2321 9378) published quarterly from International Management Institute Kolkata is an open access scholarly publication in Management. It publishes original research articles (non-technical) by scholars in the field of management and first-hand perspectives from business thinkers and practitioners on contemporary issues. It also publishes interviews with eminent personalities in the field of business. The publication currently enjoys a pan India reach with a growing audience in the international markets as well.

The articles are non-technical as the objective is to create and disseminate knowledge not only among specialists or experts, but also to create an interest about the academic knowledge among the stakeholders like individuals in corporate bodies, government departments and policy makers. There are no processing fees or publication charges for publishing in IMI Konnect and publication considerations are entirely based on the quality of the contribution that justifies the theme of specific issue or the overall aim.

Editorial Advisory Board

Editorial Team

Contents

Manuscript Submission

IMI Konnect publishes non-technical articles, case studies, commentaries, opinion pieces and interviews

Articles – full-fledged research articles

Case Studies – case studies with citations, references and exhibits.

Commentaries – writings on current issues/ facts/ events

Opinion Pieces – short write-ups involving opinion of experts from academia and/ or industry

Interviews – firsthand perspectives of experts on pertinent issues

The article should be non-technical and should be within 2500 - 7000 words. There should be an abstract of 150-200 words. 3 to 4 keywords should be included. There should not be any mathematical expressions or equations or technicalities of methods in the main text. The article should not contain any notations or equations at least in the main text. If required, they may be put in Appendix. Appendix should be placed before References. It should be typed in MS Word, Font: Times New Roman 12, Line spacing 1.5. Kindly mention your complete name, surname, affiliation, email ID and contact number while submitting articles online.

Academicians, practitioners and researchers are welcome to submit their original contributions on contemporary issues and topics related to business, viz. economics, finance, operations, strategy, OB & HR, analytics and marketing.

Please register at https://imikonnect.imik.edu.in/ to submit your articles online.

Kindly adhere to the submission guidelines mentioned at
https://imikonnect.imik.edu.in/submission-guidelines.php.

For copies and queries, contact editorial team at
Email: imikonnect@imi-k.edu.in.
Telephone: +91 33 6652 9664
Address: Amrita Datta
IMI Kolkata, 2/4C, Judges Court Road, Alipore, Kolkata 700027, India

Data Privacy in the Era of Artificial Intelligence

Pooja Sarin* and Arpan Kumar Kar**

"If you put a key under the mat for the cops, a burglar can find it, too. Criminals are using every technology tool at their disposal to hack into people's accounts. If they know there's a key hidden somewhere, they won't stop until they find it." -Tim Cook, Apple's CEO.

Abstract

The importance of Artificial Intelligence (AI), considering the way information is increasingly becoming a crucial asset, has increased in the human society. There is always a debate surrounding, if the collectors are the owners or just the safe keepers of the data. The article discusses about data and its several facets of privacy in the current scenario of evolving Artificial Intelligence (AI) applications. AI is discussed from the perspective of digital traces. Since, we are shifting towards identifiable personal information even within big data, these models become important to build into AI based initiatives. Data privacy, trust and integrity are very contextual terms and hence are discussed from an individual's and organisation's perspective. The framework for AI Governance from privacy preservation angle is discussed. Further, privacy models are proposed from privacy engineering to law and regulation perspective. Potential client's faith in the organisation and the technology utilised is critical to the success of AI-based businesses. To fulfil the social obligation, AI businesses must be compliant and reflect this extrinsically. If regulatory changes are acknowledged and implemented while maintaining public trust, only then this can be accomplished.

Keywords: Data Privacy, Artificial Intelligence, Privacy Laws and Regulations, Big Data Management, Privacy Engineering

1. Introduction

In the current age of evolving data driven society, in every aspect whether technological, societal, political(Manheim & Kaplan, 2019), environmental, economic, health and sciences, financial (Lee & B, 2018), education, sports , (Mehra *et al.,* 2021), and entertainment etc. the stakeholder's data and its access to the various other parties whether commercial or government (Reis *et al.,* 2020) play a major

* Research Scholar, Data Science, Department of Management Studies, Indian Institute of Technology Delhi

** Associate Professor, Information Systems, Department of Management Studies, Indian Institute of Technology Delhi

role in determining the society and evolving culturally based norms into a more deterministic way demographically. All this is happening with the use and misuse of Artificial Intelligence. We can't deny the importance of diligent Artificial Intelligence (AI) which has eased human society and its ways, if we look ten years back and today. AI as a technology has grown big and evolved in real sense with the help of the granular, and smaller level datasets e.g., the evolution of customised user data through popular platforms like Amazon, Facebook, Instagram, Google Maps etc. and it has a long way to go. We are lucky enough to entice the transition phase because the world is shaping up to the newer dimensions with the help of collection and analysis of individual's dataset(s). It is not only critical to the individual whom it describes, but also critical and crucial to the organizations which process this information to provide services, products, goods, and experiences to that individual. Sometimes the data getting captured, gets stored in the cloud (Kar and Rakshit, 2015), which offers greater scalability and compute, but at the cost of concerns surrounding how the data is stored and accessed outside the organization which is collecting the data. However, there are different models of cloud computing, between on premise solutions and multi-tenancy that address majority of these challenges with the appropriate data governance framework. While the positive impacts of technologies like AI and cloud computing are recognised, so are the downside of usage of these technologies documented in existing literature (Sarin *et al.,* 2020; Malik *et al.,* 2021).

In current times, data and information are crucial for citizens and government to interact in a positive, productive, and future facing way. Generally, following four key questions are to be considered while sailing through data driven journey-

1. Data collection and storage policy of the organization collecting (and owning) the data. There is also a debate surrounding if the collectors are the owners or just the safe-keepers of the data.
2. Data usage strategy within the organization that owns the data and organizations that may be granted permission for operating on the data.
3. Extraction of value and protection of data by the organization collecting it or the organization that analysed it outside the firm.
4. Strategize shift towards innovative data centric society, taking care of the above stated key-points without hindering the cultural values and norms.

We are in the maturing phase of AI-based industries impacting various societies and unifying several parameters at a global scale; one needs to be careful not to infuse inherited bias into a bigger picture and face devastating consequences later. Even the best AI product or service will not be accepted if its users do not trust it. Only by recognizing and

implementing regulatory and ethical obligations can trust be established. Identification and tracking governmental regulations and public opinion regarding AI to translate relevant AI-regulation and ethics into actionable requirements are needed for parallel progression to get things done more efficiently over time. For example, Personally Identifiable Information (PII), IT security, IT operations, legal, privacy, marketing, HR, etc. glued together would make sense if digging deeper for AI-based practices.

1.1 Artificial Intelligence: From the Perspective of Digital Traces

"The digital traces we leave behind each day reveal more about us than we know. This could become a privacy nightmare—or it could be the foundation of a healthier, more prosperous world" -Alex "Sandy" Pentland, MIT Professor

The use of AI touches more and more parts of daily life. With various opportunities like revolutionary disease detection methods, unique design concepts in engineering, or intelligent assistants for our everyday lives, AI brings many challenges. As an example, nearly a decade ago, justice systems across the United States started implementing an AI program predicting the likeliness of future crimes for various people. Shortly after its introduction, first voices claimed that the algorithms used would be racist. People of Colour (POC) were predicted to be much more likely to recidivate and therefore served longer sentences. Even if the used AI was not programmed that way, the data it learned from must have given it the impression that PoC is more likely to commit crimes again. Information is collected over a time where PoC was more likely to be imprisoned – a self-fulfilling prophecy.

Digital Traces Narrate a Story: Data tells a (Pastor-Galindo *et al.,* 2021) about every one of us every day because our daily activities are attenuated by leaving behind a virtual bread crumb. These breadcrumbs disclose more about our lives than anything we want to reveal about ourselves. Our Facebook status updates and tweets deliver information we choose to say to people, edited according to the day's standards. Digital breadcrumbs, on the other hand, record our actions in real time.

Data as Intellectual Property: Individual data can be leveraged by someone, can harm someone, build, or show value, but data that is not just some sort of social exhaust where we keep collecting it, and it has no impact at all. It won't ever be a zero-sum game. There will always be the opportunity to protect it, guard it, treat it with ethics or ignore it, allow it to be stolen or mishandled, or give false information. But a neutral zone where nothing terrible or nothing good happens when we collect information exists. Why? Because data is a version of intellectual property.

1.2 Data Privacy: A General Outlook

Data Privacy and Security (Lu *et al.,* 2014) is one of those threads that tie together every industry and, really, every individual! Suppose one thinks about data in its most personal sense. In that case, the data about healthcare,

medical conditions, the number of children they have, their ages, family members, geography, all these things make a big deal of difference to each person. They are now moving out from that circle, how to find each other in space and time, who's allowed to meet, and know where one is at what time. How to have room to go and be quiet and creative and turn everything off?

All those things are aspects of *authorized sharing of information*. What does one decide to share, with whom, and when? These foundational principles that are incredibly personal and individual, extend out as one gets to bigger and bigger groups of people. Whether we are people in a classroom deciding what we need to learn and what the agenda is, whether we are students sitting in a classroom learning from content and attempting to judge the speaker's veracity or the quality of the information, or whether we are citizens grappling with complex issues deciding how many people should be saved by crossing our borders when they do, we fear for national security reasons, whether that's valid or not. All these decisions are based upon the data stories and the individual stories that form a basis for humanity. Hence there's a need why we all have a considerable interest in understanding, at the least, and getting in and innovating and thinking about data.

Personally Identifiable Information vs. Personal Data are two classifications of data that often confuse organizations that collect, store, and analyze such data. In the United States, PII is utilized, although no single legal document defines it. The American legal system is made up of a patchwork of federal and state laws as well as industry-specific rules. Under the PII umbrella, they all represent and classify different types of data. They all represent and classify other pieces of information under the PII umbrella.

On the other hand, personal data has one legal meaning, defined by the *General Data Protection Regulation (GDPR)*, accepted as law across the European Union (EU). Both terms cover familiar ground, classifying information that could reveal an individual's identity directly or indirectly. But why is all that so important? As a website administrator, app maker, or product owner, you must be aware that the traces left by visitors and users may be sensitive. These traces may be used to identify persons. Thus they must be handled with extreme caution. From a legal standpoint, it could be a matter of breaches and violations with severe consequences. Grasping the bigger picture is crucial for an organization's security and legal compliance.

Privacy vs. Secrecy: Secrecy is different from privacy, but it's often confused. The cultural definition of the word privacy depends on the region where we're born and raised, e.g., "When I put myself out there on social media, this is information that other people may read about me." Privacy is not the same as secrecy. Privacy is the authorized processing of Personally Identifiable Information (PII), e.g., the processing of credit cards by not its

authorized user who are given access to the transaction data. The violation of privacy is sharing that information for any other purpose than a payment.

2. Proposed Framework

Following are the four critical elements to the framework of a well-governed data system.

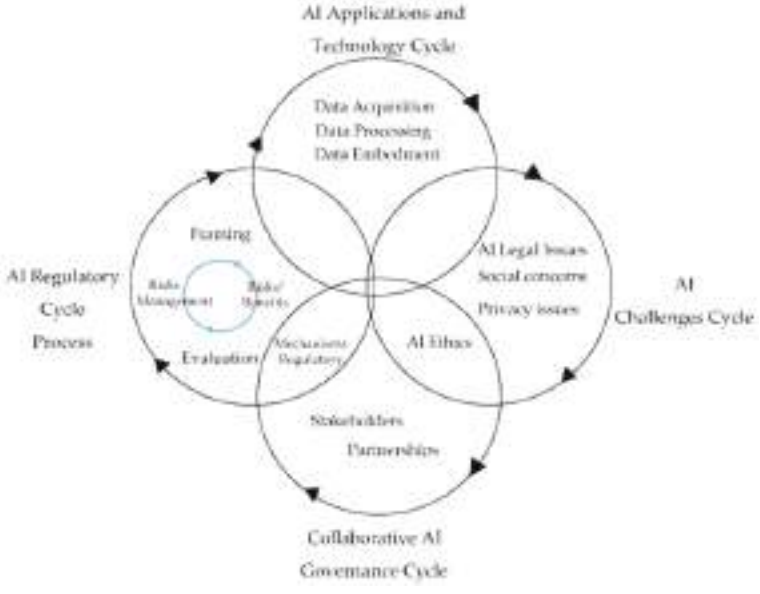
1. *Moral Processing*: A few incidents in the past were not executed, thankfully! To implant a chip under the skin in prisoners and keep a permanent record of every evil act they've ever done, whether they wanted it or not. It felt like a massive moral disruptor, which is an understatement. But there are specific codes of morality, which we can agree on, even cross-culturally. It must be moral.
2. *Ethical Processing*: From a brand perspective, above and beyond what is a moral anathema, if we look at ethical violations, it is the way of treating someone in that consistent way that an organization knows is right for their brand, including the degree of risk and profanity.
3. *Legal Processing*
4. *Fair from the start to end of the process.*

2.1 The Importance of Data Privacy:

There are several perspectives of Data Privacy as follows-

The Individual's Perspective: Data in aggregate describes stories about communities and

Figure 1: Implemetation of AI Governance Framework (Adapted from Wirtz *et al.*, 2020)



cultures. Data understands where the value lies, sometimes where bad people want to disrupt or make things worse. And one can discern what to do about that, all tied to data. It's fascinating to investigate how data is constantly organically moving, how it can be protected, how we think and develop rules about data. Data is a mirror of humanity! And, what is more important to protect than data that helps us live our daily lives, allows us to record the history of our lives, and helps us try to move forward as better societies for our future generations or people we don't even know! Data reflects value on a wide range of levels. Data is used to tell stories about people's lives.

The Organization's Perspective: Organizations care about data and privacy because it brings value to the organization. Data is that core asset that helps to hire the right people, also help to connect with respective customers. It helps to build customer satisfaction, understand how to develop better products and services pivot, and change to grow business over time that communicates with their investors and various stakeholders. Data is the new currency!

Privacy is Contextual: Depending on the situation, individuals may view confidentiality and private information differently. Now that may be a small example of contextualization. Still, if we can think about the different use cases over time, the information that we share with our doctor, is other than the information we share with all your co-workers. If we think about how data naturally flows in our life, we can understand how the scale-out and use case of personally identifiable information context comes sharply into view.

Trust and Integrity: Organizations can build trust (Kumar *et al.*, 2020) by owning, operating, and governing systems that show they are made, architected and built with the best interest in mind. That means the information you process is at work for a particular purpose over time. Organizations can organize in a trustworthy manner. Having individuals who are accountable for information throughout its lifecycle and always have an eye on that data asset is critical to build trust and integrity in the system. Organizations can audit their systems. Audit in an organization, that deals with personally identifiable information is glorious because it helps us look across our networks and understand those weaknesses? Where are we doing well? Where are our training programs hitting home so that people genuinely care?

Consumer Trust Cliff and Data Transparency: The Data Trust Cliff occurs when a person begins to create trust, then a relationship and that person continues to give more and more until something happens. And it generally happens quickly. The faith is on the verge of collapsing. Someone has had a severe security breach that could have been avoided. One may have shared the information with third parties that the person is not aware of it. By sharing

information, the organisation has not supplied services based on an individual's deal with them.

2.2 Privacy Solution:

Protecting the Story: When a story built up by data is lost, or that story is changed somehow, or the transparency to that story is lost, several different things can happen. We may not get the goods and services that we want. They may not meet our expectations anymore because the story that we've told in one way of communication isn't returned correctly in the other. Lost data or its report may result in a loss of integrity in the systems. There is a gap between the stories being told and recorded to survive digital economies in the era of gamification. The lost or hacked data stories which are not revealed may cause damage, both to an organization and an individual. The damage may include monetary penalties, integrity penalties, trust penalties, the inability for a consumer to trust that a system is telling the right story that will lead to the correct result over time (Chen, 2020). Maintaining the balance between information privacy is tricky, as customers often seek higher personalization based on unique needs that AI can enable, but at the cost of some elements of privacy (Kar, 2020; Kushwaha *et al.,* 2021).

Data in the Wrong Hands: Data breaches, data losses, data vulnerabilities, data mishaps, we hear the headlines every day in the news. What happens when there is an attack or data breach, or data loss? There can be several different things that indicate data is not well-governed and curated. Data curated in a certain way when it's attacked can be its vulnerability. Why do bank robbers rob banks? It's where the money is. Why did train robbers steal trains? It's where the money is. Why do hackers hack systems? It's where the money is. This is a digital economy. Guess what that economy is comprised of? Digital information. So, the bad guys that can take data out of systems, either from the inside or without, will traffic it on the black market. They will pose as an individual that they are not and take on their integrity as their own to steal things, to hide from their bad behaviour to harm for no good reason for sport. There are all sorts of different motivations, but the primary driver, today in most of these prominent hackers, falls in the camps of greed and power.

Privacy Engineering and Emerging Technologies: Privacy Engineering is the term having a broader scope conceived as possible beyond the definition of engineering. It is not just computer sciences, hardware, software, bridge building, or carpentry. But, if we think about engineering and how it's defined, engineering is taking all the tools and resources, and math available to you and solving a given problem in a structured way. So, privacy engineering then looks at how to design and create privacy, privacy by design, and how we use information in the information economy? How do we have authorized the processing of personally

identifiable information according to moral, legal, ethical, and fair propositions and policies?

Organizations that plan and execute statement-centric, data-centric, and privacy-centric systems best start with the approach of privacy by design. Privacy by design is a concept. It's a policy that says we build in privacy from the start. We don't believe in zero-sum thinking, or it must be private, or it must be security. It can be both. We believe that privacy, by default, is something that can happen. These are beliefs and policies, and pleasant to have, and outcomes—all important. Privacy engineering supports a company or an organization that already believes in and decides to do privacy by design. So, privacy engineering is the problem solving, the tool-using, the planning, the architecture, the testing, the transparency, the accountability structures that all support a goal and an aspiration of privacy by design. The GDPR (General Data Protection Regulation), which sets out the data protection principles, obligations, and data subjects' rights, applies to the European Economic Area (EEA). Apart from the general protective screen of the GDPR provisions, it is essential to pay attention to the critical elements of the AI and IoT-based context to focus on the main privacy issues regarding IoT. A story that is built up by data is lost or that story is changed in some way or the transparency to that story is lost, there're several different things that can happen. We may not get the goods and services that we want because of the disrupted story.

Anonymization: According to state-of-the-art, the security measure of anonymization consists of transitioning personal data into non-personal data. The GDPR's principles and requirements do not apply to anonymous data that does not identify or identify a natural person (Recital 26)[1]. As a result, it may prove to be an effective security strategy for extensive IoT data.

Purpose Limitation: Among the seven principles in Article 5 of the GDPR, this processing principle applies, lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity, confidentiality, and accountability. Further data processing that is incompatible with the intended collecting purpose must be limited. Therefore, focusing on the true meaning of each processing via IoT could limit unnecessary data processing with a different context from the primary processing.

Data Minimization: This processing principle of GDPR refers to prioritizing the necessity of each processing. This processing principle requires only the necessary personal data for each processing in the context of IoT. For example, the location data of a device is not considered an essential element of the processing, if the specific processing is not related to the location.

[1]Recital 26, EU GDPR. Retrieved from: https://www.privacy-regulation.eu/en/recital-26-GDPR.htm

DPIA (Data Protection Impact Assessment): The Data Protection Impact Assessment is an assessment of the "impact of the envisaged processing operations on the protection of personal data" (Article 35 para 1). This assessment is not required for every processing, but it is needed where "it is likely to result in danger to natural persons' rights and freedoms. IoT is a new technology (Article 35 refers to the contribution of new technologies in the DPIA conduction), and it would be necessary before launching any new IoT application.

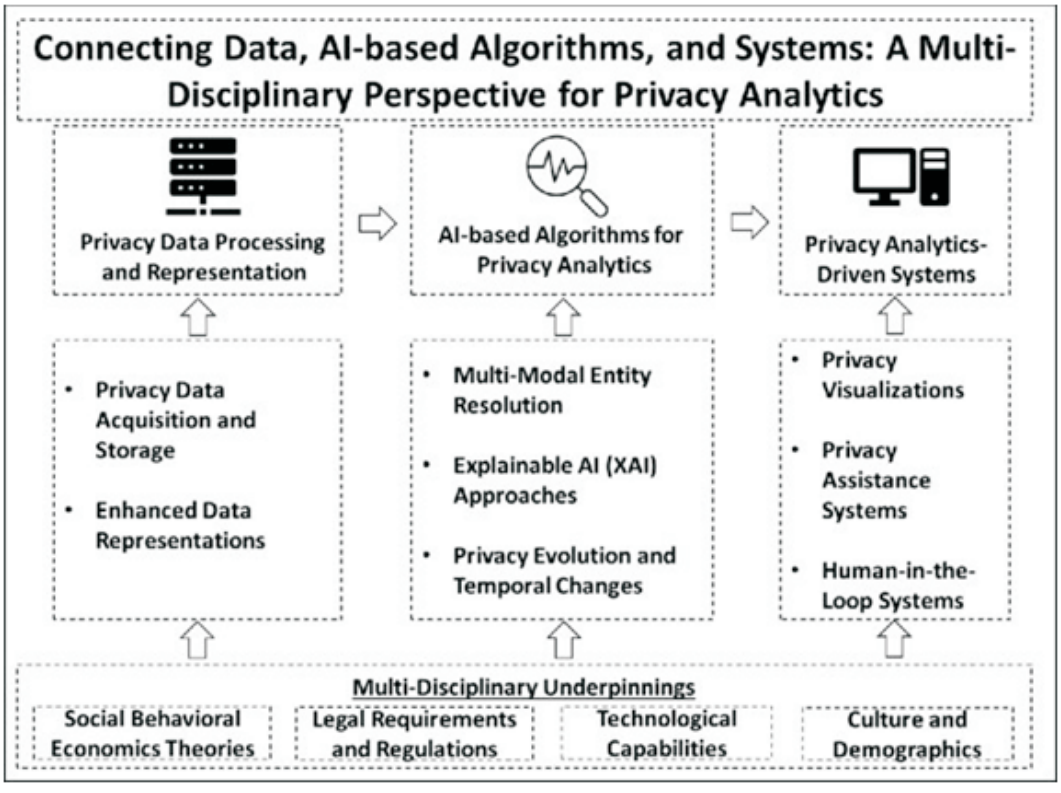
Lawful Consent: The data subject's consent is a fundamental and everyday legal basis of data processing. In the context of the consent via IoT environment, two significant elements should be mentioned. Firstly, according to Recital 32[2], if the consent request is provided with electronic means, this should not interrupt the use of the specific service. Furthermore, withdrawing consent should be as simple as giving consent. Regarding IoT applications, special consideration should be given to the procedures that include the departure of the data subject's consent. The potentials of new and ground-breaking technology and its socio-economical consequences cannot even be imagined when developed. But at the same time, societies have learned from past experiences that regulations for the usage of upcoming inventions are essential and handy to prevent unknown negative consequences. Companies developing applications with these new technologies are faced with such changing and adjusting regulations – legally and ethically. AI technology has just started to present its possible benefits for our society. But we already have multiple examples of its implementations supporting (or even creating) organizations opposed to liberal democracies. Corporations are stuck in the middle of this debate. They are experiencing an increasing speed of developments and growing complexity as digitization progresses. Legal issues contribute massively to this subject. For companies working with AI, the speed of technological development and the complexity of the subject matter are compounded because many legal issues are still completely unresolved or legal requirements are in constant motion. A specific example of this is the now planned EU regulation of AI. In addition, there are numerous ethical considerations (Fluitt *et al.,* 2019) about AI, such as social responsibility, which is feeding into the public debates. For example, the High-Level Expert Group on AI appointed by the European Commission also published Guidelines for Trustworthy Artificial Intelligence.

3. Discussion

The success of companies working with AI also depends to a large extent on potential customers trusting the company and the technology used. In addition, legal violations

[2]Recital 32, EU GDPR. Retrieved from: https://www.privacy-regulation.eu/en/recital-32-GDPR.htm

Figure 2: Connecting data, algorithms, and systems: A multi-disciplinary framework for AI-enabled privacy analytics. (Adapted from Samtani, Kantarcioglu & Chen, 2021)



can have consequences beyond the control of companies. It is, therefore, the task and the challenge of companies to adapt to regulatory and ethical requirements by both complying with and implementing legal requirements as well as communicating and ensuring integrity to external parties, especially potential customers. In essence, AI companies must be compliant and reflect this externally in order to live up to their social responsibility. This can only be achieved if regulatory changes are recognized and implemented while creating trust by the public. Excellent technology is not enough. Even the best autonomous car will not be driven by anyone if it is not trusted. Internal compliance cannot be contemplated at all without external integrity. Following are the guiding questions that should be addressed in current scenario-

1. What are the most relevant regulatory developments that corporates are currently confronted with?

2. How do you ensure trust on AI systems given challenges of transparency and replicability of the outcome given a specific data based context?
3. What ethical discussions surround the development of AI regulations from a multi-stakeholder perspective?
4. How might integrity and ethical corporate conduct be defined from multiple stakeholder perspective?
5. When corporate decisions are considered ethical and what determines the boundaries of fair usage of data within and across the enterprises?
6. What challenges do firms face with regards to understanding and implementing the implications of new AI regulation and ethics?
7. What role does firm communication play in terms of visibly displaying trustworthiness and regulatory as well as ethical compliance?
8. What other departments might play a role in ensuring corporate integrity alongside legal departments?
9. What functionalities would a solution need to provide for and for whom and who may it compromise in facilitating such an outcome?

The Facebook-Cambridge Analytica scandal, as well as technological company's data aggregation activities have demonstrated how vulnerable modern society is to privacy infractions. The National Science Foundation (NSF) has stated that AI-enabled models, artefacts and systems may swiftly and effectively sift through vast amounts of data from legal papers, social media, DarkWeb sites, and other sources to prevent privacy infringement(Samtani, Kantarcioglu, & Chen, 2021). However, analysing current data sources, developing AI-enabled privacy analytics to meet growing difficulties, and deploying systems to fulfil essential privacy demands all require significant work. Because of their exceptional accuracy, deep learning methods have become the foundation of new AI-based services on the Internet in the big data age. In the meantime, it presents obvious privacy concerns. The deep learning–assisted privacy attack can extract sensitive personal information from unstructured data such as photos and videos as well as text. Considering the adversarial example approach and ongoing active research in this field, distinct image privacy protection techniques based on the various metrics are evolving e.g., especially for photographs of complex structures, image privacy can be safeguarded by adding a modest amount of noise that has no discernible influence on image quality (Liu, Ding, Zhu, Xiang, & Zhou, 2019).

References

Chen, Z. (2020). Privacy Protection Technology in the Age of A.I. *IOP Conference Series: Materials Science and Engineering*, *750*(1). https://doi.org/10.1088/1757-899X/750/1/012103

Fluitt, A., Cohen, A., Altman, M., Nissim, K., Viljoen,

S., & Wood, A. (2019). Data Protection's Composition Problem. *European Data Protection Law Review*, *5*(3), 285–292. https://doi.org/10.21552/edpl/2019/3/4

Kar, A. K., & Rakshit, A. (2015). Flexible Pricing Models for Cloud Computing Based on Group Decision Making Under Consensus. *Global Journal of Flexible Systems Management*, 16(2), 191-204.

Kar, A. K. (2020). What Affects Usage Satisfaction in Mobile Payments? Modelling User Generated Content To Develop the "Digital Service Usage Satisfaction Model". *Information Systems Frontiers*, 1-21.https://doi.org/10.1007/s10796-020-10045-0

Kumar, A., Braud, T., Tarkoma, S., & Hui, P. (2020). Trustworthy AI in the Age of Pervasive Computing and Big Data. *2020 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2020*. https://doi.org/10.1109/PerComWorkshops48775.2020.9156127

Kushwaha, A. K., Kumar, P., & Kar, A. K. (2021). What Impacts Customer Experience for B2B Enterprises on Using AI-Enabled Chatbots? Insights From Big Data Analytics. *Industrial Marketing Management*, 98, 207-221.

Lee, S., & B, J. Y. (2018). Trustworthy AI in the Age of Pervasive Computing and Big Data. https://doi.org/10.1007/978-3-319-93563-8

Liu, B., Ding, M., Zhu, T., Xiang, Y., & Zhou, W. (2019). Adversaries or Allies? Privacy and Deep Learning in Big Data Era. *Concurrency and Computation: Practice and Experience*, *31*(19), 1–14. https://doi.org/10.1002/cpe.5102

Lu, R., Zhu, H., Liu, X., Liu, J., & Shao, J. (2014). Toward Efficient and Privacy-Preserving Computing in Big Data Era. *IEEE Network*, *28*(4), 46–50. https://doi.org/10.1109/MNET.2014.6863131

Malik, N., Tripathi, S. N., Kar, A. K., & Gupta, S. (2021). Impact of Artificial Intelligence on Employees Working in Industry 4.0 Led Organizations. *International Journal of Manpower*.https://doi.org/10.1108/IJM-03-2021-0173

Manheim, K., & Kaplan, L. (2019). Artificial Intelligence: Risks to Privacy and Democracy. *Yale Journal of Law & Technology VO - 21*, *106*(1), 106. Retrieved from https://www.foreignaffairs.com/arti-%0Ahttps://search.ebscohost.com/login.aspx?direct=true&db=edsgao&AN=edsgcl.587366495&amp%0Alang=pt-br&site=eds-live&scope=site

Mehra, V., Sarin, P., Singh, P., Sawhney, R. S., & Kar, A. K. (2021, September). Impact of COVID-19 Pandemic on E-participation of Fans in Sports Events. *In Conference on e-Business, e-Services and e-Society* (pp. 692-703). Springer, Cham. https://doi.org/10.1007/978-3-030-85447-8_57.

Pastor-Galindo, J., Marmol, F. G., & Perez, G. M. (2021). Nothing to Hide? On the Security and Privacy Threats Beyond Open Data. *IEEE Internet Computing*, *25*(4), 58–66. https://doi.org/10.1109/mic.2021.3088335

Reis, J., Santo, P., & Melão, N. (2020). Impact of Artificial Intelligence Research on Politics of the European Union Member States: The Case Study of Portugal. *Sustainability (Switzerland)*, *12*(17), 1–27. https://doi.org/ 10.3390/SU12176708

Samtani, S., Kantarcioglu, M., & Chen, H. (2021). A Multi-Disciplinary Perspective for Conducting Artificial Intelligence-enabled Privacy Analytics: Connecting Data, Algorithms, and Systems. *ACM Transactions on Management Information Systems*, *12*(1). https://doi.org/10.1145/3447507

Sarin, P., Kar, A. K., Kewat, K., & Ilavarasan, P. V. (2020). Factors Affecting Future of Work: Insights From Social Media Analytics. *Procedia Computer Science*, 167, 1880-1888.

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. International Journal of Public Administration, 43(9), 818-829.

Challenges and Opportunities of handling data in a Data Driven World : Experience from the Indian Census

Sayantika Basu

1. Introduction

Census is an exercise of collecting information on the entire population of a certain place on a regular periodicity. The study of "Demography" which is associated with the study of population has its origin in the Greek word "demos" (population). Hence the linkage of the Census process with that of the systematic study of the population under Demography can easily be established.

In terms of methodology and volume, Demography and Population Studies have always been guided by processing, deciphering and analysing an enormous amount of data. The challenge of interpreting and forecasting the human population characteristics by analysing large volume of data collected through population Census and surveys contributes significantly in understanding the pattern of economic growth. This also helps us in identifying the weaker areas seeking greater humane intervention and preparation of inclusive planning. Thus, Census acts as the base for conducting more in-depth studies and channelise the resources for the betterment of the society.

Conduct of Census in India, the second most populous country in the world, is itself a monumental exercise offering multiple challenges. Handling a huge volume of data is one important part of the exercise. In this vast expanse of the nation, Census is conducted in more than 6.4 lakh villages, more than 7000 towns using the services of more than 27 lakh Census (Census of India 2011) functionaries in many different layers and activities – the hierarchy goes from senior bureaucrats of the Union of India and States/UTs to that of the Enumerator and Supervisor involved in the field level data collection. For house-to-house data collection across the length and breadth of the country, the Census Schedule is translated into 16 languages and the instruction manuals explaining the modalities of the work, is prepared in 18 languages. The magnitude of the exercise thus requires no further elaboration.

Unlike another giant exercise of Indian

* Deputy Director, Directorate of Census Operations, West Bengal, Government of India, Ministry of Home Affairs

democracy i.e. the General Elections, which happens in every five years, directly involving the general populace in different ways, majority of our population do not know much about the Indian Census, one of the largest civil exercises of the world. As a matter of fact, the Census in India, indeed holds a leading position among the best Census exercises in the world. Periodicity of 10 years, simultaneity across this vast country and collection of all information for 100 per cent population are few of its most significant, but fascinating characteristics.

There are two distinct but correlated phases of modern-day Census. The first phase is called the House Listing and Housing Census Operations (HLO) where all places of probable human habitations are systematically documented collecting vital information on habitation, housing pattern, basic amenities and assets of the household concurrently. This phase also forms the base for individual data collection for the second and final phase of Population Enumeration (PE). In this final phase, also termed as the main phase of Census, the demographic particulars are collected for all residents in the country available for enumeration during the 3-week enumeration period (subject to certain conditions). Identical Enumeration Schedule is canvassed for all 3 types of households - normal, institutional or houseless - for culling information on name, age and date of birth, sex, literacy and education, religion and mother tongues, disability, marital status and working status, migration and fertility particulars. Handling information for all these parameters for 121,08,54,977 (as per Census 2011) population is genuinely a herculean task. Still, for every decadal exercise more or less than 300 Census data tables are published, through numerous stages of data processing, scrutiny and meticulous checking. The volume information, thus, collected and compiled in Census of India is certainly "BIG".

In this article, section 2 focuses on the genesis of Census followed by the methodology followed in India in section 3. Section 4 presents the data management in Census and the challenges faced; section 5 concludes the article by focusing on the prospects of census data in the era of big data.

2. Antecedents and Genesis

The earliest reference of efforts of ascertaining the population of a nation-state in India is found in "Arthashastra" of Kautilya (Chanakya). The basic purpose Census, as depicted in this 3rd. century BCE treaties on state policy, economy, warfare and welfare of the state was to find out the number of able-bodied young males for building a strong army for the safety and security of the nation-state. Reference of conducting Census as a part of state policy is also found during Medieval period in "Aieen-E-Akbari" by Abul Fazal, one of the 'Navaratna' in Mughal emperor Akbar's Court and his state historian. The Indian Census exercise, in today's "Modern" concept is essentially based on post-industrial revolution theories and practices. India was

introduced with these methods during the British "Raj".

Since the early nineteenth Century attempts have been made by the British efforts for ascertaining the extent of the diversity of the population of newly acquired regions in Indian sub-continent and conducting the population headcount. The earliest evidences of such endeavours can be found for Census in Bengal region, especially in present day Kolkata and its neighbourhood during CE 1800. In subsequent decades, as the British territory grew larger, the Census exercises became more and more structured and range and coverage of the exercise also diversified and expanded. Decades of practical experimentation and learning- resulted in an elaborate Census exercise in and around 1871 CE. Census 1871, as a matter of fact, is considered as the first modern Census of India. The process was initiated in 1865 in some portions of the country and could reach a stage of completion in 1872. Experiences of conducting Census 1871 provided thorough insights and helped in identifying the methodological deficiencies of conducting Census in a staggered manner as the population is a dynamic phenomenon owing to regular events of births, deaths and migration. This resulted in the synchronous census exercise all over the country in 1881. Since 1881, the Indian Census is being conducted every decade, the last one in 2011 being the 15th in sequence. The unprecedented havoc created by COVID-19 has, for the first time in the history of modern India, broken the decadal periodicity consequent to its postponement on account of the prevailing pandemic.

3. Census Methodology in India – Some Glimpses

Census in India follows 'Canvasser method' for statistical data collection, where the Census Enumerator visits every household with pre-designed Census Schedule and enquires notified set of questions. The time period for this visit is also notified in the Official Gazette. The data collection during most of the British period was that of a "One Night Affair" where on the particular Census night, the population of the entire country was enumerated on "As is where is" basis. This method is known as 'De Facto method' of conducting Census. Till date, many countries in the world use this method by declaring curfew on a particular night and complete enumeration before daybreak. However, in case of a diverse country like India this could not continue very long and "extended de facto" method was adopted since Census 1941. In this process, a particular period of time is notified for conducting Census, usually 3 weeks and all details are captured all over the country simultaneously during this period. That's why Indian Census is a Synchronous Census.

The responsibility for conducting Census in India lies on the Office of the Registrar General and Census Commissioner. India has

its Head Quarter in New Delhi and State Directorates across all States and Union Territories of the country. Census organisation is thus also the nodal authority of registration of births and deaths in India. Under their guidance and with the help of Governments of all States and Union Territories - Census in independent India is conducted every decade.

3.1 Enormity of Indian Nation and Volume of Census Data- Initiation of Census in Independent India

Assessing the enormity of Census exercise in today's digital era where the whole world is mapped in our palm, calls for glancing through its evolution process.

In the words of Sardar Vallabhbhai Patel referred in Census (Census of India, 1951, Volume I India, Part 1A, Report) "….Census as an institution goes back to the remote past, but it (Indian Census) is no longer a mere counting of heads; it involves extraction of the information which plays a vital role in determination of many of our administrative policies."

The Census (1951), the first exercise after the adoption of Census Act, 1948 in independent India, was a pioneering effort in many aspects. Data on all villages and towns in the country were collected, compiled and published systematically. 5,93,518 enumerators and 80,006 supervisors were engaged for the task of enumeration and supervision of fieldwork. The population of India was reported to be 36.10 crores for statistical purposes, the country was demarcated into six broad zones – North, East, South, West, Central and North West India. State names like Madras and Coorg; Bombay, Saurashtra and Kutch; Vindhya Pradesh, Mysore, Travancore and Cochin, Bilaspur and Ajmer were in place, most of these jurisdictions have either merged or renamed subsequently. The Princely States were governed by Rajpramukh who also took initiatives to make Census process successful in the regions under their control.

This coverage was not only numerically large but also multi-dimensional. In 1951, the policy planners of the newly independent country tried to dream about the futuristic world based on the best of both worlds – the technology of the West with ingrained thoughts of conventional wisdom of typically Indian practices. In this nascent stage of the Indian Republic, while the complete path of future course of development orient actions were yet to be envisioned, various Official statistics were clubbed with Census and the Census publications contained many aspects apart from the population figures.

3.2 State Reorganisation in 1950s to India We See Today

Subsequently, the teething period started getting over after bloodsheds of partition. State Re-organisation Act 1956, brought into reorganisation/recurving of the State boundary for many of the states as per the need of a newly independent country. In 1956

State Re-Organisation and next few years many states were freshly carved out through state specific Acts. Most of the territories under the Princely States and other governance merged with Indian Union. In West Bengal, the erstwhile Princely State of Kochbihar and French region of Chandannagar became a part of India around the early 1950s. Census data is collected and compiled as per Administrative Jurisdictions at different levels. All changes in jurisdictional boundaries during the intervening period of the two censuses are taken into consideration as a pre-requisite.

Although the outer boundary of the country did not change much since independence, there have been numerous internal changes. Table 1 bears testimony to the quantum of administrative changes that Census has taken care of over a period of five decades.

Table 1 transpires that the population grew steadily as the number of administrative units in the country also increased through the creation of new States and Union Territories and curving out of new districts internally. During the first few decades after independence, India's population registered a decadal growth rate of more than 20 per cent. Thus, the population almost doubled in the first 30 years (from 36.1 Crore in 1951 to 68.33 Crores in 1981). Gradually, with the overall societal changes and emphasis on smaller household sizes, the population growth rate has started slowing down in subsequent years. The momentum of population change, however, will take some more time to stabilize. With lower birth and death rates, longer life expectancy and better access to public health facilities, the overall population of the country can be expected to stabilize in the next few decades (Table 2).

Table 1: Coverage of Census in India

Census Year	No. of State and UTs	No. of Districts	Population
1961	14 & 12 (Including other areas)	303	43.92 Crores
1971	21 & 6	360	54.81 Crores
1981	22 & 8	412	68.33 Crores
1991	25 & 7	466	84.64 Crores
2001	28 & 7	593	102.87 Crores
2011	29 & 6	640	121.08 Crores

Source: Census of India 2011:

1. Provisional Population Totals Paper 1 of 2011, India Series 1,

2. Primary Census Abstract, India

Table 2: Population Growth in India

Census Year	Decadal Population Growth (Absolute)	Decadal Population Growth (Percentage)
1951-61	7,81,46,681	21.64
1961-71	10,89,24,881	24.80
1971-81	13,51,69,445	24.66
1981-91	16,30,91,942	23.87
1991-2001	18,23,16,397	21.54
2001-11	18,21,17,541	17.70

Source: Census of India 2011 –

1. Provisional Population Totals Paper 1 of 2011, India Series 1,

2. Primary Census Abstract, India

4. Adopting New Technologies and Data Management

Indian Census organisation is one of the leading Government organisations in the country to adopt and adapt scientific and systematic methods of collection, compilation and processing of large volume of data. As part of this effort, with an objective to conduct Census in a more efficient way, the administrative jurisdiction is divided into multiple smaller layers to collect and manage the large volume of data in a structured and logical manner. The States and Union Territories (UT) are divided into districts and districts into sub-districts (not sub-divisions). Census Enumeration Blocks (CEB) are the smallest units for collection of Census data. CEBs are demarcated within the boundary of the CD blocks or towns (as sub-districts) to ensure equitable distribution of work among each Census Enumerators. The Enumerators identify every possible location of human habitation in their respective jurisdiction and visit every household to collect the information. Crores of Census schedules are filled in every decade and are processed systematically. The collected data is then compiled and processed for preparation of data tables at various levels of jurisdiction.

During the earlier days of the era of manual compilation and processing, declaring the Census results and bringing out numerous publications for the data users had itself been a challenge. Method of 100 per cent enumeration of a few basic details and sample enumeration on some other parameters were practised till Census 1991 to manage the enormity of the task of covering so many aspects. Sample size varied from 10 to 20 per cent depending upon the need of the hour.

Statistical estimates were then made for these sampled parameters. For example, during Census 1951, migration, fertility, mortality (causes of death and IMR) etc. were estimated through data collection using Individual Slips for selected samples. Various sampling methods like Stratified Random Sampling, Systematic Sampling etc. were used to ensure proper representation of the heterogeneity of the population. The fact remains, though may sound strange, that till now many developed countries in the world like the United States, are continuing use of this technique of differential base of enumeration to overcome data handling challenges. Simultaneously, the state of the art technological innovations have been utilised for Indian Census data processing from time to time. From 1951 to 1991, specially trained Computational staff was engaged for mechanised coding of the filled-in schedules and sorting. For example, in 1971 –Key punching Machines (electrical and mechanical) were used for feeding the data in IBM 1401 Computer with Card readers. In 1981, designated data centres were set up across the country where paper-based information was converted to machine readable format through data entry in machines. FACIT (A calculator machine used worldwide for faster and accurate calculations till the 1980s. Gradually replaced by Electronic and other Calculators (Like the ones we used in our younger days) and other mechanical devices contemporaneous to the time were introduced for faster computational work. In Census 1991, Medha-930 mainframe Computer system was installed at the Head Quarter at New Delhi and Data Centres across the country were connected with it under Unix Operating system as Dumb Terminals. For the first time in the 1991 Census, Data Tables were published in digital Format besides conventional Books.

In the advent of the new millennium, the decision to conduct an enumeration of 100 per cent population with uniform Census schedules for both the House listing and Population Enumeration Phases in the entire country brought a new set of challenges. It necessitated a paradigm shift in the entire Census methodology. In Census 2001, Intelligent Character Recognition (Handwriting Recognition of scanned images through Computerised processing) ICR enabled scannable Census Schedules were introduced for the first time to digitise 100 per cent data collected from the field. All administrative units as well as all Census schedules for both phases of the Census were assigned with logical location codes. Computerised data processing was adopted. Although the volume of data increased manifold in the decade, the adoption of improved technology made the process of data management smoother and less cumbersome.

Continuing the same tradition, in Census 2011 also improved versions of the method was adopted. It ensured direct conversion of a large portion of the data to digital editable format, while, for the rest manual data entry from scanned and digitised Census schedules

were used. The use of more technology-oriented method of Census data collection and its application in data processing is thus a rational choice.

4.1 Coding and Meta Data of Census

Systematic identification of all administrative units in the country covering 7000 plus towns and 6.4 Lakh plus villages without missing a single jurisdictional unit has always remained a priority in Census information management. The entire process of Census every decade commences with the process of updating the jurisdictional changes happening during the period. For example, in a medium sized state like West Bengal, updation of every change for 40,000 plus administrative entities during a decade without any omission or duplication, is indeed a big ask.

Up to the 1981 Census, localised code structure was used as per the regional variation of the individual State/UT. Since Census 1991, a logical method was developed for coding across the country with an eye to standardise the code structure and ease out the post processing of collected data. Census 2001 Permanent Location Code Numbers (PLCN) for all administrative units in the country was assigned. The scanning and post processing of scanned schedules was conducted with this PLCNs. This was truly a huge step towards systematic data processing and tabulation. This Code structure not only envisaged a new era of Census Data Management but also worked as the base for numerous statistical surveys. This structure is heavily utilised in the representation of statistical data both by Government and Private bodies, researchers and many more.

However, the door-to-door fieldwork with 20 Lakh plus Census functionaries using 8 digit code structure created new sets of administrative issues. Cleaning of errors in Coding on the body of the Census Schedules became extremely rigorous in 2001. These Codes were central to the entire data processing and thus, ensuring their correctness was of prime importance. To get rid of this problem without affecting the procedure, the Code structure of Census 2011 was prepared in two sets - simpler codes for field level activities (Short Codes) and a more elaborate format for data processing. For the first time in Census 2011 Metadata and Data Standards (MDDS) Codes with one- to-one mapping with short codes of all administrative units of the country was introduced. MDDS is the data standard adopted for use across e-governance systems targeted to enable more efficient and easier standardised data structure. Indian Census has been the obvious choice for implementation of this novel project of linking diversified data structures in the country.

4.2 Management of Census: Challenges in the New Era and the Efforts to Overcome

The Census being an administrative exercise for collection of statistical information only

once in each decade, has its own set of managerial challenges which are unforeseen in any other statistical exercises and are unique in nature. Contrary to the regular statistical sample surveys, Census faces various non-sampling errors, which, by their own nature, are truly difficult to anticipate. Thus, every Census is like opening a bunch of gift boxes – some are very common and predictable; while the others unfold into unpredictable challenges, necessitating instant and novel actions. One basic requirement of the Census is the simultaneous deployment of large-scale manpower for covering each and every household of the country. But arrangement of twenty-five lakh enumerators at the same time is a huge exercise. Once this group is identified the next impending hurdle comes in the form of training them up suitably within a very short period of time. Most of these enumerators have very little prior exposure to statistical data collection. To tackle this problem, few interventions are made which have large bearings on the outcome of Census process. Special efforts are made to frame the Census questions either in close-ended or non-descriptive manner or inviting a numeric answer to minimise the possibility of un-avoidable enumerator bias emanating from lack of prior idea in statistical principles. On the other hand, well designed, properly executed and elaborate arrangement of training covering all minute details forms the foundation of a successful Census.

In a multilingual country like India where numerous languages are spoken, the enumerators hail from diverse linguistic backgrounds. Preparing all documents in multiple languages is thus another pre-requisite for successful Census. As mentioned earlier, the Census documents are translated into 18 languages. Translation works as one of the most important tool to percolate the guidelines and procedures – as a bridge for conveying all technical details to non-technical functionaries who board in an unfamiliar statistical domain. It is extremely challenging to translate the instructions in a language which is simple, easily relatable yet not missing any of the technical details.

Ensuring coverage of all population in every nook and corner of the vast expanse of this country is indeed the biggest focus of Census Operations and quite naturally it offers the largest hurdle. Process of identification of jurisdictional changes start 4 years prior to the actual fieldwork. Every factor is counted for and all issues go through various field trials and pre-testing so that no achievable goal is missed out. Since 1951, post enumeration checks or surveys are regularly conducted to improve the quality of coverage and the content of collected data. In Census 2011 the Net Omission Rate was 22.98 per thousand population, implying that out of 1000 persons less than 23 people were missed out. This rate is at par with what most of the developed countries could achieve.

Updating the Census questionnaire is another significant challenge. Identification of the

most important factors affecting the development and improvement of the standard of living in commensuration with decadal changes in the society are the first and foremost requirement. Next comes the meticulous task of framing unambiguous closed ended questions preferably with numeric answer choices. The 2011 Census had 54 items covered in the two different phases. Every decade sees new sets of issues pertinent to the changing society, emanating new topics for inclusion in Census and new choices for the existing set of questions based on analysis of previous data returns. For example, the inclusion of internet accessibility or consideration on whether or not to continue with 'usage of radio' is instances of such analysis to cater for the requirement of the changing society.

5. Prospects of Census Data in Era of Big Data

As has been discussed earlier, since inception, efforts have been made to apply the latest applicable technology in the process of conducting Census so that authentic and reliable information banks can be formed in a systematic manner. On the other hand, once the census data tables are released, a lot of efforts are made for the proper dissemination of the data. Census data tables in books have been published since the 1860s. Data availability in electronic format was initiated in Census 1991. Availability of electronic data became more elaborate and user friendly in subsequent decades. All Tables of Census 2001 and 2011 were released in electronic format. As per the latest Data Dissemination strategy, most of the published items are hosted on the website for use. Numerous publications since Census 1871 are also hosted on the website with free access. Dedicated Data Dissemination Units are set up in all Census Directorates across India.

There exist enormous possibilities in linking of the Census data sets to various other sources. Use of GIS based technology in preparation of maps is already in vogue for quite some time. Its extension to satellite imagery can open a new horizon in data interpretation. Time Series data on population changes happening along the coasts of a river with the changes in its course of the journey over a period of time can also lead to interesting insights. Impact of growth or shrinkage of forest cover on human habitation captured through satellite imagery can be an interesting topic for research in Social Sciences. Even impact of habitation through geotagged location lists with that of the natural calamities like floods or extremely heavy rainfall can be explored through analysis of big data. Prevalence of communicable diseases and availability of safe drinking water and their impact on population growth can also lead to Public Health Researches and can facilitate positive interventions.

Rigorous analysis of night time satellite imagery on the usage of electricity can also produce several inferences. In fact, time series information on the luminosity of the lights

captured at night time emitted by a locality can help us in interpreting the Census data in the usage of electricity. In the past Census remained the largest and perhaps the only source of data to estimate the loss of housing conditions after a natural calamity like a Super Cyclone or a disaster in the hills. Now time has come to carry forward the task, towards achieving a better society with proper use and interpretation of Census Database.

Quadrennial Heartbreak Called the Olympics

Subhasis Ray*

1. Introduction

Another Olympics came and passed by giving us more despair than enjoyment. Yes, there is a 250 per cent improvement in medals tally compared to 2016 but 7 medals for a population of 1.38 billion falls short of expectation. Every 4 years it comes and points out where India stands in the sporting arena. Criticism, discussion, planning and also adulation for the medal winners gradually will subside till a round of stirring just months before the next Olympics. This has been the story since 1896; feel more responsible after 1947 since India could have carved a niche for herself. Olympic analyst Gracenote's prediction of 19 medals for India in Tokyo now seems like a joke. Australian news channel ranked India at the 87th position (the last position among all countries that got at least a medal) on per capita medal count after Rio Olympics and this appears far more realistic assessment however unpalatable it may be. There are individual athletes whose personal cupboard is no less decorated than what India could achieve over 125 years! Can we be euphoric with just two individual golds in two decades with the kind of human resources at our disposal? Can you match the stamina of one who walks 10 kilometres every day for two matkas of water? Can you match the reflex of one who every day fights with tigers and crocodiles to catch a crab or collect honey? Can you match the strength and tenacity of people, especially at the hills who keep climbing mountains on a daily basis that too under low level of oxygen? Can you match the power and agility of Jalikattu players who tame bulls? The only thing missing is the proper planning to convert their abilities into Olympics related skills. If nurtured, these people can find 'struggle for medal' far easier than 'struggle for life' in all probabilities; after all self-actualization perches much above the need for roti, kapda and makaan in Maslow's pyramid.

2. Long Term Vision

Swamiji's outlook towards sports in nation-building is well documented. Sports in a child's daily life not only boosts health, self-confidence and leadership quality but also can teach gender equity, teamwork, quick decision making and most importantly, how to deal

* Management Consultant; Currently teaches at IISWBM

with losses. A few days ago, noted cardiologist Dr. Kunal Sarkar while releasing his book 'The Sickness of Health', commented that India should not probably be into the highest level of medical surgeries (e.g. 'complex neuro work, cardiac work, transplant work') given the socio-economic status of this country. Rather than judging the comment, I call this a consequence of our culture and tradition. Just the way he was inspired by his teachers, today all students aspiring to be the best cardiologist finds Dr. Sarkar as a reference point. Culturally, sports could never break the barrier of just being a pastime despite such a rich history of archery, horse riding, chariot racing, polo etc. generally played by the Kshatriyas. We are in a position where a healthy percentage of people know the pathway to reach MIT (Massachusetts Institute of Technology), CERN (European Council for Nuclear Research), JHU (John Hopkins University), NUS (National University of Singapore) or LSE (London School of Economics) to excel in any stream of studies. Can we say the same thing for players - football, volleyball, basketball or for that matter any discipline of sports (barring cricket and hockey where there is culture and expertise in-house)? The path to Santa Monica Club or Nick Bollettieri Academy is not only hard but also hardly known. The over dependence on academics has taken its toll. Till the 10th board exam, it is hard to find a family who accepts academics as a secondary skill with sports or any performing arts being the primary skill for their wards. A naturally talented athlete with an overdose of integration/differentiation lands up nowhere. By the time our athletes win junior Wimbledon, players from other countries are already into seniors' tournaments. For example, Ramanathan Krishnan, Ramesh Krishnan and Leander Paes had all won at 17 years of age whereas Michael Chang and Boris Becker won their first men's grand slam around that age. Pete Sampras turned professional in 1988, at the age of 16 and Roger Federer turned pro at 17. Sports and physical education are intertwined with the school curriculum in many western countries e.g. the USA. They have seen that students who play competitive sport in high school demonstrate more confidence, leadership with a better sense of morality in future. UNESCO is promoting Quality Physical Education (QPE) in partnership with the European Commission, International Bureau of Education (IBE), International Olympic Committee (IOC), United Nations Development Programme (UNDP), UNICEF, WHO and Nike in countries like Fiji, Mexico, South Africa and Zambia to inculcate sports and PE skills. St. Thomas Aquinas High School, Boston College High School, Christopher Columbus High School, Mater Dei High School are some of the schools in the USA known for providing a steady supply line to the national baseball, basketball, football, soccer, gymnastics, boxing ambit. Considering the socio-economic standing of India, I shall go one step ahead - can we have an education policy that promotes

specialization at an early stage – say class V? Can we design a course curriculum where Music, Dance or Sports of any kind gets the major share as primary skill and conventional subjects like science, mathematics, social studies aim to boost their primary skill and to prepare them as a responsible citizen? Once success kicks in, the guardians will be less condescending towards academic skills – look at the queue outside cricket coaching centres! However, building the necessary infrastructure including teachers to support such a policy and curriculum may be a farfetched dream – in management lingo, a long term vision.

3. Government Initiatives

We have seen many Government sponsored efforts. Samagra Shiksha Abhiyaan started in 2018, is aimed towards integrating sports within schools by ensuring a) supply of all sports related equipment and b) sports education as an integral part of the curriculum. These schools will participate in the annual Khelo India Youth Games (KIYG), an initiative by the Ministry of Youth Affairs and Sports in order to identify and nurture sporting talents. Besides, there is the Eklavya Model Residential School (EMRS), started by the Ministry of Tribal Affairs that aims not only to impart quality education to tribal children in remote areas but also to build their all-round development. Each such school is supposed to own a Centre of Excellence for Sports, a facility that will include sports kit, sports equipment, specialized training, boarding and lodging, competition exposure, medical expenses etc. However, the Annual Status of Education Report (ASER) from the 2018 survey reveals that less than 20 per cent of schools have got dedicated teachers for sports and physical education, around 66 per cent of schools have a designated period for physical education and only two-thirds of the schools have playground inside the school or adjoining places across all states. With another such initiative Fit India Movement, it seems there are dilution as well as confusion in designating the real target; is sports for excellence or is this just a pastime or a leisure activity or for fitness? Then there is the Target Olympic Podium Scheme (TOPS) from the Sports Authority of India whereby hand-picked athletes are provided financial and other assistance for succeeding in the Olympics and other such events. However, assigning foreign coaches or sending abroad a grown-up athlete for training may not yield the desired result. As the gap is already so big that catching up becomes untenable leading to mental stress and drainage of money. There are only sporadic successes from all these schemes. Policies, intent and investments are there but governance at such a big scale is posing challenges and it is yet to establish a continuous supply chain of athletes. Besides, can we say other initiatives like ISL or i-League where indigenous footballers get to play with and against some stalwarts of yesteryears to do any substantive improvement in India's football or its ranking? Rather India is on the verge of losing the

dominance it used to enjoy within the SAFF countries. Inspired by the success of IPL, cricket's franchise based competition, many other sports including football started the same. Over the last couple of years, badminton has closed down. There was a major legal tussle with the federation in volleyball.

4. The Prescription

'Catch them young' should be the mantra if one wants a sustainable result. The difference in skill and ability is nil at birth. As the age increases the gap between an Indian and a champion athlete increases too at a faster rate only to become insurmountable. There is the latest trend especially in football as many European clubs are selecting kids with merit as well as means from private boarding schools to join their nurseries. Even this is not a scalable model - can sports be considered a tool for empowering more vulnerable population? As a relatively short term measure, can we not pick up, say one or two kids (of around 10 years of age) per Lok Sabha constituency inheriting special genes described in the first paragraph and nurture them for the next 10 years? All this 'laboratory' needs is a supply of good food (which the kids probably don't get now), ambitious and worthy foreign or indigenous coaches and teachers (who has seen and done it all) and infrastructure (for chosen sports to specialize). More emphasis should be given on the knowhow rather than the infrastructure; Brazil footballers or middle distance runners from Kenya, for example, do not enjoy robust infrastructure. The recent success of Sikkimese boys and girls in the national U-19 one-day cricket tournament is largely being attributed to the knowhow of high performance coach, Atul Gaekwad. If money is an issue, sporting disciplines can be selected that does not warrant high infrastructure cost e.g. football, badminton, wrestling, volleyball etc. Another important thing is to ensure that the local community encourages the players during their formative years by cheering them during play. When these kids deliver in next 10 years, not only their parents will have anything to complain but also their entire locality will have their role models and the much desired culture of multiplying effect will kick off.

Scouting is a difficult task but the Government can surely provide the necessary bandwidth. There are many other challenges for nurturing young talents in India e.g. socio-economic, language, religious, cultural, dietary habits, social restriction, gender bias etc. The Public Private Partnership model involving NGOs, Businesses and Government can collectively deal with challenges.

5. Role of Data in Sports Management

Finally, in this data driven world, data and technology like any other industry should form the backbone for the management and governance of any such initiative. Developing sportspersons today has become extremely scientific. Presently, there is hardly any database on school physical education and sports in India including facilities, students

and their capabilities. Every game can now be captured, analyzed and played back many times – thanks to digital video capture. The coach should work with each ward; analyze their strengths and weaknesses to develop personal action plans for skill enhancement including a food chart, medicines, if any.

The sporting capabilities along with physical and mental parameters can be benchmarked against athletes of other countries in the same age group. Virtual reality modelling language (VRML) can create a virtual athletic training room as an accessory to train the trainers. The model has been found to be beneficial for teaching training room design, function and application of rehabilitative tools to trainers. Using the internet, coaches can easily get information on new trends, background on competitors' teams, players, coaches and competition locations, weather, transport, logistics etc. Analytics can then make it possible for coaches and athletes to analyse and integrate information and resources to improve training, decision-making and skill development. The following table outlines the data requirements in order to manage a facility described above:

Examples of various countries and games adopting data and technology are galore. Let's consider the ZXY sports tracking system for soccer which was originally developed by ZXY Sport Tracking, a Norway based firm founded in 2002. The system incorporates the use of video cameras to focus on each player individually and mandates the players to wear an ultra-light belt around their waistline. Personalized video measurements can provide data like positional updates at least 20 times per second, distance covered at different speeds (jogging, running, or sprinting), acceleration vs. retardation, pulse rate etc. synchronized in real time from body sensors with the sensors installed in the ground. Training are now hyper-customized

Sport Function	Database Content
1. Players	Demographic data - name, age, gender, contact information, medical details, food chart, performance history, restrictions, if any, achievements
2. Coaches	Batch information, training timetable, conditioning timetable, data from tools for practice and analysis, monitoring of evolving practices
3. Administration	The roster of volunteers, officials, timekeepers, coaches, counsellors, doctors, nutritionists and their contracts, equipment and inventory lists, facility maintenance, licenses for tools for practice and analysis
4. Public Relations	Donors for money or in-kind services, organizing tournaments, participating in other tournaments, liaison with the external world

according to the physiological structure of the player and the possibilities of results linked to different strategies can thus be studied – thanks to the wearable and Internet of Things (IoT) generated data and Artificial Intelligence and Machine Learning driven predictive analysis and detailed suggestions. Billy Beane, the General Manager of the Oakland Athletics baseball team, used algorithms based on statistical procedures to build a team of players who are cheap but collectively effective to fight budget crunch. The fact that the team went into the playoffs has been immortalized by Michael Lewis's book *Moneyball: The Art of Winning an Unfair Game* (2003) and later through the Brad Pitt starred film Moneyball (2011). Australian Open in tennis is one of the pioneers of the application of social listening and sentiment analysis by installing multimedia kiosks, touch screens and tennis totems to study the emotions and reactions of the players with respect to fans and fellow players. Closer at home in the 2009 IPL final, Anil Kumble started bowling for RCB against Adam Gilchrist of Deccan Chargers and got him out for a duck in the first over itself. The insight was provided by Bangalore based analytics company Mu Sigma after thorough video analysis of Adam Gilchrist's feet movement at the early stage of his innings. With continuous improvement in tool and analysis techniques, data and technology can enhance the sporting decision making process, thus elevating the team's sporting and business performances.

6. Conclusion

In a dance reality show on TV, I saw a few six to eight year old kids perform gymnastics in the name of a dance and they are as good as any floor exercise I had seen before. If we can quarantine them for six to eight months and hone their dancing skills based on their extreme physical fitness why can't the sports realm turn them into great athletes? 'Start small and scale fast' can be the way forward. However, financial and emotional support for athletes is essential for sports to be viewed as a sustainable career path. It is important to have alternate employment options for players limited by physical injury, mental fatigue or even skill preferably within the realm of sports itself.

Digital Transformation in Banking and Fintech Disruption

Santanu Mitra*

Santanu Mitra heads the coverage of digital disruptors in all domains. His area of focus is to grow the bank's franchisee with companies in ecommerce, payments, ride sharing, OTT, fintech, education technology etc. Prior to this he was a senior relationship manager and team lead within Telecom Media Technology Coverage. In a conversation with team Konnect, he shared his personal views regarding the digital transformation in banking and the challenges from the fintech disruptors.

IMI Kolkata: *How would you like to define the evolution and digital transformation of the financial services sector in the global context?*

SM: Digital transformation of the financial services sector started with just digitisation of the services, but in a true sense, it is going to change the way financial services are provided and consumed. This would imply a paradigm shift in consumer behaviour.

The first signs of change were around the way consumers accessed banking services, for example, mobile banking. This was followed by the emergence of wallets (Paytm etc.) which changed the way people paid for things, even for such payers and receivers who did not have the infrastructure to support credit cards.

This was followed by other areas like the selling of mutual funds and insurance products that moved from the domain of banks to online platforms. The move was primarily driven by the need to reduce distribution costs.

You may observe that the drivers in each case were different. One was driven by the necessity of reducing cash from the system and making cashless mode of transaction available to the credit unworthy consumer base; the other was driven by the necessity of reducing the distribution cost of the products. But the only common enabler was the ubiquity of mobile phones.

I envisage taking lessons from digital transformation that is happening across the world. For example, India would see the emergence of digital currency in a big way. We are likely to see a central bank backed digital currency (i.e. CBDC or digital INR) soon.

Facilitated by blockchain we would see a whole world of decentralised financing (DeFi) emerge. The primary reason institutions like

*Managing Director – Head of Digital Economy Group, DBS Bank Ltd

bank exists are to ensure trust in the system. If DeFi riding on blockchain based technology emerges in a big way this could also make the banks redundant to a large extent.

A whole world of trade finance exists because there is an endemic lack of trust amongst buyers and sellers located in different geographies. This has ensured that almost a couple of trillion dollars of trade financing happen through banks via issuance of (LCs) and confirmation of the same. This entire segment can move completely online as verification is being done using blockchain, thereby obviating the need for banks to play a role in the confirmation of documents and financing the same. The financing can then be done directly by other digitally enabled platforms.

The next step could be the emergence of metaverses in which the role and shape of financial institutions are yet to be conceptualised.

IMI Kolkata: *Do you think, the 'Digital India' campaign added an impetus to the digital transformation in the Indian Banking domain? If so, what are the aspects and initiatives that have assisted in such transition?*

SM: The 'Digital India' campaign has definitely added an impetus. One of the most important developments in this domain that is visible to us is the emergence of a government enabled payment rail called Unified Payments Interface (UPI). India was one of the first countries in the world to come up with a government enabled instant payment rail.

Some of the other aspects are around the centralised KYC registration. While these are building blocks and the impact may not be transformational, but the collective impact of all these building blocks are immense and would continue to be felt in the medium term.

IMI Kolkata: *What are the demand drivers of the digital disruption in banking sector?*

SM: The most important demand drivers are reductions in costs of distribution of products and services. I have touched upon some of these in the earlier answers. But another important demand driver is the ubiquity of mobile services and the way people are consuming services enabled by their handheld mobile devices. Banking is likely to be more and more invisible and embedded in our daily lives.

IMI Kolkata: *How did the Fintechs emerge?*

SM: The Fintechs emerged as there were opportunities within the whole gamut of banking and financial services in which banks were perceived to make a cut without really adding value or justifying the need for them to be there. With the emergence of e-commerce, a significant majority of the society felt cut away from this ecosystem due to them not having ways and means to pay for goods/services offered online. This led to the emergence of payment based fintech.

Subsequently, we saw the emergence of fintech in the insurance domain. Here the

sales function was largely left to individual agents and/or banks. This implied a high cost of distribution, without any meaningful value addition.

IMI Kolkata: *How are the fintechs connected to digital disruption?*

SM: Banks have a couple of core functions. One of which is maturity transformation which is basically the asset liability management, borrowing short from people and lending long to corporates and people. Second is the monitoring of opaque loans. When we talk of opaque loans there is a huge amount of information asymmetry that still exists. When banks lend to large corporates as well as customers, the amount of information that the banks may have based on their ability to do due diligence is not necessarily publicly available. If the information that even large corporates publish in their websites and reports were sufficient then everybody could have done an analysis of creditworthiness on the basis of public information that was put out but that is not the case. So, monitoring is one of the most important functions and the rest are payment and transaction services.

Now, payment and transaction services create a lot of information footprint which is verifiable, codifiable and then there is a soft aspect that is relationship based. For example, payments could mean remittances, cross border payments, or, a simple payment from party A to party B for a trade related transaction. The third type has the largest volume where a corporate A buys a service or a product from a corporate B, then there is an invoice raised which is processed and the invoice goes from the ERP of the buyer to the ERP of the seller. Then there is a payment that happens from a bank through a correspondent bank if it involves different currencies and different geographies. In case the payments take place within the same geographical location, it is a bank to bank transaction. Earlier this used to go through a clearing channel, wherein cheques, RTGS etc would be involved. So, we get an idea about the amount of data footprint that gets created which is all verifiable and codifiable because it goes through multiple touchpoints. So, if the payment is initiated on behalf of the buyer from the buyer's bank and goes to the seller's bank, then everybody in that loop, i.e. the buyer's bank, the clearing house, the seller's bank get the information footprint. It is called codifiable because with computerisation all of that information would be addressable by computational tools. Then there is a soft and relationship based approach that is not quantifiable but which is still codifiable.

IMI Kolkata: *How do we describe the demand and supply sides of such digital disruptions?*

SM: When we look into the supply side that is driving this digital disruption, we have application programming interfaces (APIs), cloud computing, smartphone, digital currency and blockchain technology which is being considered as potentially the most important technological development of the 21st century. Blockchain is basically a

technology that is a distributed ledger. So today when there is a clearance of any currency or any other transaction there is a centralized ledger to ensure its authenticity. But, blockchain just allows you to maintain that record across multiple nodes which is a distributed ledger and so the chances of fraud etc. as well are reduced.

Now, the drivers of the demand side of digital disruption include better service expectations. For example, the way we used to buy things is moving into e-commerce, the way food was consumed is getting transformed with food delivery, the traditional taxi rides are getting transformed into ride-sharing and associated with this digitisation of the common economy around us, there is an expectation of better payment mechanism and better financial transaction.

All of these exchanges of goods and services are getting digitized and with each exchange of goods and services, you have to make consideration for payment. Also, there is a general shift in user behaviour. The younger generation is comfortable using smartphones and would not want to use cash to make payments and live their lives within the mobile and that ecosystem. So, there is an expectation related to payments and remittances.

IMI Kolkata: *What are the fintech innovations with regards to payments?*

SM: Paypal emerged associated with eBay. eBay was the platform for people to sell from one customer to another customer (C2C), then B2C. When eBay started evolving they realized that they needed a better payment experience. It couldn't be that people would be swiping credit cards and so Paypal emerged. Then in the late 1990s and early 2000s, when the Chinese ecosystem started seeing traction around e-commerce they realized that there was a huge gap because the credit card penetration was not that great. Also, the banking network was not growing at the pace at which e-commerce was growing. WeChatPay had its origin in the gaming environment because it originated from Tencent and they realized that when gamers are playing, the way to make money was through in-app instant payment. So, this is when the wallet emerged. Similarly, Alipay was a wallet and thereafter we have now seen a preponderance of wallets as it gives one a much more seamless experience. So, the first level of fintech centred on payments because that was one of the most unregulated and it could be codified and verified. People were not too worried as it did not involve any investment decision and even a small amount could be put into the wallet. When the ecosystem further evolved into mobiles and international payments, then came up Apple Pay, Revolut etc.

The next step was driven by the demand drivers like peer to peer (P2P) lendings, insurtech, robo advisers etc. Once the platforms matured, they had the capability to see based on statistical database, customer profiling, performance based pricing etc. For

example, assume that there is an e-commerce platform. It processes so much data comprising of the customers who are buying from it, sellers who are selling through it, that even without having necessary access to the credit bureau score and the income generation ability of a person, it can profile basis this data about the ability to pay. In any kind of lending decision, there are two aspects, one is the ability to pay and the other is the willingness to pay. The ability to pay is linked to income so people would go to a credit bureau, ask for your financial data, bank data etc. but because of the data footprint that is accessible to the platforms, they realized that the same information can potentially be triangulated using the data footprint that is already easily accessible to them. So, unlike a bank based lending model where a certain percentage of loans are given and pricing of loans are decided based on loan size, fintech, based on the previous or prior performance, can offer dynamic pricing. This also meant a substantial reduction in the need for people for running such an organization. They had computers doing the entire credit profiling, collecting people when there are defaults but they did not need so many people who were originally needed to sanction consumer loans.

Financial inclusion was also another aspect that helped them. The fintech did not have any legacy technology to deal with so they did not have any issue like I am sitting on a core banking solution of Finacle so if I have to modify that to do any API based processing as opposed to batch processing; there is a huge amount of legacy technology that needed to be integrated. So, talking about the advantages, these were free of legacy systems, better consumer interface (i.e. UX) with much higher ROE part of the business. The disadvantages included the fact that fintech do not have a loyal customer base, higher cost of capital in comparison to the banks as they cannot take cheap deposits and lack of regulatory and risk management expertise which held them back from going too deep into the lending part of the business.

IMI Kolkata: *What according to you are the Fintech trends to look forward to, in the forthcoming years in the financial services sector?*

SM: The fintechs have entered into the distribution of investment products i.e. mutual funds. They are getting into the distribution of life insurance products. They are conceiving moves into the micro/instant insurance segment too.

The fintechs which were in the payment and gateway space are increasingly realising the digital payment rails across countries are taking away their basic rationale and hence will evolve into credit tech.

IMI Kolkata: *How has the fintech revolution impacted the traditional banking system?*

SM: The banking space has been impacted meaningfully as some of the functions which were primarily centred on the distribution of products and services, without taking any credit risk have moved to the fintechs. These were traditionally the value added services out

of which banks used to generate a lot of fee income. As these were primarily centred on information arbitrage, the fintechs saw an opportunity and started disrupting them.

As a result, the banks have started tightening their belts and are digitising their service delivery approaches. Most of the banks have ensured that they digitise or partner with the fintechs (that white label their services) to replicate the same customer journey. Today, most banks allow instant remittances on their online platform, subscriptions to mutual fund investments, buying insurance and other products.

IMI Kolkata: *How can the traditional banks strengthen their competitive advantage over the FinTechs and BigTechs?*

SM: The fintechs have the advantage of being fast and nimble in terms of developing newer tech paradigms and solutions. That has ensured that they have stolen a march on banks as articulated above. However, while the banks are slow due to their legacy IT development and manpower structures, at the same time, they have access to a much bigger source of capital and manpower, when they decide to invest resources. This is evident in some of the banks who are taking Fintechs head-on like Citibank, DBS, JPM Chase, HDFC, Santander as well as many others.

The strengths of the banks lie in their access to the liability balances which are not available to the fintechs. Even after 8 years of being, some of the payment banks or, fintechs do not have more than ₹ 500 in their wallets. Most of the banks have developed their tech teams and some banks have more software engineers than some of the medium-sized IT service companies. The banks are even acquiring fintechs to bolster their capabilities e.g. Goldman, Citi, JPM and Wells Fargo.

IMI Kolkata: *Do you think the challenges to traditional banking have put the customers in an advantageous position? If so, how?*

SM: As the banks digitise their services, the customers get the benefit of conducting all their transactions on one platform with the sense of security of being in a banking environment.

IMI Kolkata: *The digital transformation has led to ease of accessibility to customer data. Is the regulatory vigilance adequate in India to safeguard the interests and privacy of the customers?*

SM: The balance between efficiency and data protection would be an ever evolving situation. Data had maximum protection when it was in paper (or so we think) but even then any misuse could never be tracked. Digital transformation has led to a situation where there is a significant positive impact on the increase in efficiency of services. However, the question of data privacy has assumed a higher degree of importance. The regulatory regime is evolving to help this. With data localisation laws, supported by stringent protection clauses we would see some improvement in the ability to trace and plug leaks.

However, the use of data for the marketing of services will continue to impinge on individual concerns around privacy, and this would be a perennial issue.

IMI Kolkata: *How do you think Fintechs and the Banks can help in co-creating value for the customers?*

SM: Embracing agility in serving the customers efficiently is the call of the hour. To cater to this need many banks and fintechs partner to co-create value for the customers by maintaining the desired costs, offering more diversified and innovative products and services. This is usually done by mutual sharing of resources like the customer base of the traditional banks and the application programming interface (API) based platforms of the fintechs.

The fintechs have their own set of challenges. The biggest challenge being that the customers are apprehensive about fintechs. Also, the absence of physical establishments, sound knowledge about financial products and improper infrastructure are the other challenges. By partnering with the banks the fintechs get access to the bank's database and physical infrastructure which add to their advantage. On the other hand, banks have challenges like manpower and legacy systems which are robust and pose difficulty in faster adaptation to changes. By partnering with the fintechs, the banks, therefore, can encash on selling their products and services using the fintech platforms.

IMI Kolkata: *Financial inclusion is one of the major goals of financial policy makers. Is Fintech contributing to achieve this goal?*

SM: Financial inclusion refers to the provision for accessible, affordable and convenient financial services for one and all. Fintechs have application programming interface (API) based platforms which can help in building awareness regarding financial products which thereby enable financial literacy and inclusion. By partnering with the traditional banks, fintechs are able to offer better savings opportunities and more access to credit and insurance covers. This in turn helps to reduce income inequality.

IMI Kolkata: *Where does India stand vis-à-vis other emerging countries in terms of the implementations of Fintech?*

SM: India has witnessed much progress in adapting to digitalisation in the banking sector. However, there is further scope for building on digital financial awareness (DFA) and digital financial literacy (DFL). Currently, the digitalization in banking facilities is mostly limited to the section with access to the internet and smart devices. The study by ASSOCHAM-PwC in December, 2018 revealed that there would still be 50 crore non-smartphone users, approximately. The way forward would be to bring all such users under the umbrella of digitalisation. Also, to enhance customer protection against vulnerability due to breaches, robust cyber security measures need to be strengthened.

IMI Kolkata: *How do the regulators respond to the ongoing digital transformation?*

SM: The regulators want the bank incumbents to open up their APIs and allow for the fintechs to integrate to foster innovation and create a level playing field. On the other hand, the regulators want the fintechs to come under the purview of the banking regulation with reserve requirements, monitoring requirements and data protection regimes to protect customer privacy. Further, they would also take into account potential instability in the financial sector.

International Management Institute Kolkata
2/4C, Judges Court Road
Alipore, Kolkata 700027
Telephone: +91 33 6652 9664
website: www.imi-k.edu.in