P10 / The Political Economy of Global Digital Data Flows: Access, Opportunites and Constraints

Topic : SCIENCE AND TECHNOLOGY POLICY

Chair : Shishir Jha (Indian Institute of Technology, Bombay)

GENERAL OBJECTIVES, RESEARCH QUESTIONS AND SCIENTIFIC RELEVANCE

Digital data, unsurprisingly, has become a core concern of the rapidly growing digital economy. How is such digital data extracted, stored, processed, brokered, monetised and valued is an area of considerable and growing research interest? The governance of such big data is acquiring increasing global salience. Given the worldwide accumulation of data at a rapid pace, the large-scale extraction, processing, and monetization of such data by powerful entities, both corporations and governments, both for profit and control, is truly astonishing. There are enabling conditions that are shaping the contours of big data. Governments appear to be competing against each other to lure big data centers and processing companies. Intelligence gathered from aggregate digital data has now become a compelling part of the business model for digital platforms.

Data controllers, data processors, and data brokers are getting organized around the process of collection, processing, and leveraging of data at a global level. The varying interests of such entities (controllers, processors, and brokers) in setting up practices, rules and norms, is shaping big data development. Data brokerage has emerged as a cornerstone of the modern data infrastructure, occupying a critical space between the roles of data fiduciaries and data processors. This intermediary role is indispensable as data brokerage facilitates data monetisation while simultaneously enabling its movement through complex ecosystems. Programmatic advertising, an example of efficient data utilization, highlights the monetization of user information in real-time bidding systems. The growing sophistication of data-brokering platforms reveals a stark tension: while their mechanisms fuel economic innovation and value extraction, they often circumvent users' informed consent, exposing them to privacy violations. Policy frameworks such as the GDPR (EU), the DPDPA (India), and PADFA (US) aim to govern these complex dynamics. However, these frameworks still struggle to address the evolving nature of data ecosystems.

Another unfolding dynamic is within the Indian digital platforms such as Swiggy and Zomato. The digital data generated from the food and restaurant business is being used to create a ‘new digital supply chain’ beneficial to several participants in the business value chain. Such dominant Indian platforms leverage their digital supply chain and extract value from various stakeholders, including restaurant businesses, cloud kitchens, grocery suppliers, delivery partners, payment platforms/gateway, and consumers. Platform aggregators, based on the data acquired, optimise procurement, order fulfilment, pricing and delivery through algorithmic based decision making. These platforms collect real-time customer data that includes, search behaviour, past orders, location, time of purchase, price sensitivity, among others. In this data-oriented supply chain, there is an evident power asymmetry and platform aggregators clearly have an advantage over data they gather 24/7. They can rank restaurants based on pricing, quality of food, and the commissions paid. Those businesses that do not meet the platform demands for promotion and advertisement tend to get lower visibility. Such is the power of the overall business intelligence that is gleaned from the several stakeholders. The question to be asked how is the data extraction exercise (‘data flows’) allowing such digital platforms to leverage their data domination and increase their market share.

Finally, the increasing importance of legally vetted (copyright consent) digital data flows has caught the AI industry rather unaware. With reports stating that AI models could run out of training data as

early as 2026, the stakes are, unsurprisingly, rather high. Extracting data from individuals from their social media accounts did not prove to be legally constraining because the permission to extract was already baked into the individual’s attempt to sign onto the platforms consent form. However, such consent is a non-starter and legally not admissible when it comes to large swathes of already copyrighted content. Given this legal and now judicial hurdle (with several court cases the world over), with plaintiffs such as New York Times, ANI, Concord Music Group, Getty Images, having already taken different AI

engines to court, the question to ask is how does the ‘legal’ use of large databases portend for AI training and for data flows.

Through an extended literature review and case-based analysis, we intend to examine the emergence of such dominant players, both platform and other Big Tech goliaths, and to reflect upon the institutional arrangements that are contributing to big data flow and the growing concerns around data access, opportunities of growth and the constraints faced. How are these emerging entities jostling for power over continuous data flow and how is their relative influence shaping the very contours of the political economy of big data?

References:

· Evelyn Ruppert, Engin Isin, and Didier Bigo, Data politics, in Big Data and Society, July 3, 2017

· M Paterson, M McDonagh, ‘Data protection in an era of big data: The challenges posed by big personal data’ in Monash University Law Review, 2018.

· Mark Coté, Paolo Gerbaudo and Jennifer Pybus, ‘Politics of Big Data’ in Digital Culture and Society, December 24, 2016.

· Choi, J. P., Jeon, D. & Kim, B. (2019). Privacy and Personal Data Collection with Information Externalities. *Journal of Public Economics*, Vol 173, 113-124.

· Faroukhi, A. Z., Alaoui, I. E., Gahi, Y. & Amine A. (2020). Big data monetization throughout Big Data Value Chain: a comprehensive review. *Journal of Big Data*, 7:3.

· Pinchot, J., Chawdry, A. A. & Paullet, K. (2018). Data Privacy Issues in the Age of Data Brokerage: An Exploratory Literature Review, *Issues in Information Systems*, Volume 19, Issue 3, pp. 92-100

· Jenny Quang, ‘Does Training AI Violate Copyright Law?’, 36 Berkeley Tech. L. J. 1407, 2021.

· K Hristov, ‘Artificial Intelligence and the Copyright Dilemma’, 57 IDEA 431, 2017.

CALL FOR PAPERS

Digital data, unsurprisingly, has become a core concern of the rapidly growing digital economy. How is such digital data extracted, stored, processed, brokered, monetised and valued is an area of considerable and growing research interest? The governance of such big data is acquiring increasing global salience. Given the worldwide accumulation of data at a rapid pace, the large-scale extraction, processing, and monetization of such data by powerful entities, both corporations and governments, both for profit and control, is truly astonishing. There are enabling conditions that are shaping the contours of big data. Governments appear to be competing against each other to lure big data centers and processing companies. Intelligence gathered from aggregate digital data has now become a compelling part of the business model for digital platforms.

Data controllers, data processors, and data brokers are getting organized around the process of collection, processing, and leveraging of data at a global level. The varying interests of such entities (controllers, processors, and brokers) in setting up practices, rules and norms, is shaping big data development. Data brokerage has emerged as a cornerstone of the modern data infrastructure, occupying a critical space between the roles of data fiduciaries and data processors. This intermediary role is indispensable as data brokerage facilitates data monetisation while simultaneously enabling its movement through complex ecosystems. Programmatic advertising, an example of efficient data utilization, highlights the monetization of user information in real-time bidding systems. The growing sophistication of data-brokering platforms reveals a stark tension: while their mechanisms fuel economic innovation and value extraction, they often circumvent users' informed consent, exposing them to privacy violations. Policy frameworks such as the GDPR (EU), the DPDPA (India), and PADFA (US) aim to govern these complex dynamics. However, these frameworks still struggle to address the evolving nature of data ecosystems.

Another unfolding dynamic is within the Indian digital platforms such as Swiggy and Zomato. The digital data generated from the food and restaurant business is being used to create a ‘new digital supply chain’ beneficial to several participants in the business value chain. Such dominant Indian platforms leverage their digital supply chain and extract value from various stakeholders, including restaurant businesses, cloud kitchens, grocery suppliers, delivery partners, payment platforms/gateway, and consumers. Platform aggregators, based on the data acquired, optimise procurement, order fulfilment, pricing and delivery through algorithmic based decision making. These platforms collect real-time customer data that includes, search behaviour, past orders, location, time of purchase, price sensitivity, among others. In this data-oriented supply chain, there is an evident power asymmetry and platform aggregators clearly have an advantage over data they gather 24/7. They can rank restaurants based on pricing, quality of food, and the commissions paid. Those businesses that do not meet the platform demands for promotion and advertisement tend to get lower visibility. Such is the power of the overall business intelligence that is gleaned from the several stakeholders. The question to be asked how is the data extraction exercise (‘data flows’) allowing such digital platforms to leverage their data domination and increase their market share.

Finally, the increasing importance of legally vetted (copyright consent) digital data flows has caught the AI industry rather unaware. With reports stating that AI models could run out of training data as

early as 2026, the stakes are, unsurprisingly, rather high. Extracting data from individuals from their social media accounts did not prove to be legally constraining because the permission to extract was already baked into the individual’s attempt to sign onto the platforms consent form. However, such consent is a non-starter and legally not admissible when it comes to large swathes of already copyrighted content. Given this legal and now judicial hurdle (with several court cases the world over), with plaintiffs such as New York Times, ANI, Concord Music Group, Getty Images, having already taken different AI engines to court, the question to ask is how does the ‘legal’ use of large databases portend for AI training and for data flows.

Through an extended literature review and case-based analysis, we intend to examine the emergence of such dominant players, both platform and other Big Tech goliaths, and to reflect upon the institutional arrangements that are contributing to big data flow and the growing concerns around data access, opportunities of growth and the constraints faced. How are these emerging entities jostling for power over continuous data flow and how is their relative influence shaping the very contours of the political economy of big data?

References:

· Evelyn Ruppert, Engin Isin, and Didier Bigo, Data politics, in Big Data and Society, July 3, 2017
· M Paterson, M McDonagh, ‘Data protection in an era of big data: The challenges posed by big

personal data' in Monash University Law Review, 2018.
· Mark Coté, Paolo Gerbaudo and Jennifer Pybus, 'Politics of Big Data' in Digital Culture and Society, December 24, 2016.
· Choi, J. P., Jeon, D. & Kim, B. (2019). Privacy and Personal Data Collection with Information Externalities. *Journal of Public Economics*, Vol 173, 113-124.
· Faroukhi, A. Z., Alaoui, I. E., Gahi, Y. & Amine A. (2020). Big data monetization throughout Big Data Value Chain: a comprehensive review. *Journal of Big Data*, 7:3.
· Pinchot, J., Chawdry, A. A. & Paullet, K. (2018). Data Privacy Issues in the Age of Data Brokerage: An Exploratory Literature Review, *Issues in Information Systems*, Volume 19, Issue 3, pp. 92-100
· Jenny Quang, 'Does Training AI Violate Copyright Law?', 36 Berkeley Tech. L. J. 1407, 2021.
· K Hristov, 'Artificial Intelligence and the Copyright Dilemma', 57 IDEA 431, 2017.

P10 / The Political Economy of Global Digital Data Flows: Access, Opportunites and Constraints

Chair : Shishir Jha (Indian Institute of Technology, Bombay)

Session 1

Friday, July 4th 16:00 to 18:00 (D1)

Accessing Big Data: Is copyright emerging as a legal constraint for AI?

Shishir Jha (Indian Institute of Technology, Bombay)

The increasing importance of legally vetted (with copyright consent) information and knowledge for training, the ever-hungry AI engines had perhaps caught the AI industry unaware. With reports stating that AI models could run out of training data as early as 2026, the stakes are, unsurprisingly, rather high.

Extracting data from individuals from their social media accounts did not prove to be legally constraining because the permission to extract was already baked into the individual's attempt to sign onto the platforms consent form. However, such consent is a non-starter and legally not admissible when it comes to large swathes of already copyrighted content. Given this legal and now judicial hurdle (with several court cases), the question to ask is how does the use of large databases portend for AI training. Plaintiffs such as New York Times, ANI, Concord Music Group, Getty Images, along with several well-known authors, have already taken different AI engines to court. Other publishers such as, Rediff.com, Taylor and Francis group, LeMonde, The Atlantic, among several others are inking deals with AI companies to avoid legal trouble.

What, then, are the various cases against AI engines really amounting to? What is both broadly and more precisely at stake? Will access to data get constrained for the several AI models or is it merely a negotiating strategy by plaintiffs or is there perhaps much more at stake, in the very way that such AI companies get to access content? What are the issues on the copyright side of the debate? Will copyright prove to be an uncanny impediment in the march towards access to data for hungry AI models? These are a few of the questions that I'd like to raise and address in this paper.

References:

· Jenny Quang, 'Does Training AI Violate Copyright Law?', 36 Berkeley Tech. L. J. 1407, 2021.

· Kieala, D. et al, 'Plotting progress in AI', Contextual AI Blog, 2023.
· K Hristov, 'Artificial Intelligence and the Copyright Dilemma', 57 IDEA 431, 2017.
· Pablo Villalobos et al, 'Will we run out of data? Limits of LLM scaling based on human-generated data', EPOCH AI (2024).
· Shuja Asrar, 'AI cos' hunger for data ignites copyright wars', MINT, December 10, 2024.
· V K Ahuja, 'Artificial Intelligence and Copyright: Issues and Challenges', ILI Law Review Winter Issue 2020.

Locating data processors within big data: Understanding the interests and institutions

Ridhi Kakkar (Indian Institute of Technology, Bombay)

The governance of big data is acquiring increasing global salience. Given the worldwide accumulation of data taking place at a rampant pace, the large-scale extraction, processing, and monetization of such data by powerful entities, both corporations and governments, both for profit and control, is drawing increasing interest among scholars. There are enabling conditions that are shaping the contours of big data. Governments appear to be competing against each other to lure big data centers and processing companies.

Data controllers, data processors, and data brokers are getting organized around the process of collection, processing, and leveraging of data at a global level. The varying interests of such entities (controllers, processors, and brokers) in setting up practices, rules and norms, is likely to shape the contours of big data development. Through an extended literature review and case-based analysis, we intend to examine the emergence of such dominant players and to reflect upon the institutional arrangements that are contributing to big data growth and emergence, with a particular interest in the role of data processors within it. How are these emerging entities jostling for power and how is their relative influence indeed shaping the very contours of the political economy of big data requires more careful examination?

Regions across the world (GDPR, DPDPA) are either formulating or altering existing data governance structures to have jurisdictional control over data. Concerns over the privacy rights of individuals and communities over their data, which caters to their ability to access, control, use, and protect their data have become even more consequential.

References:

· Bruckner, M. A. (2018). The Promise and Perils of Algorithmic Lenders' Use of Big Data, *Chicago-Kent Law Review*, Vol 93 (1).

· Evelyn Ruppert, Engin Isin, and Didier Bigo, Data politics, in Big Data and Society, July 3, 2017

· Jose van Dijck, Datafication, 'Dataism and Dataveillance: Big data between scientific paradigm and ideology', in Surveillance and Society, Vol. 12 No. 2 (2014).

· M Paterson, M McDonagh, 'Data protection in an era of big data: The challenges posed by big personal data' in Monash University Law Review, 2018.

· Mark Coté, Paolo Gerbaudo and Jennifer Pybus, 'Politics of Big Data' in Digital Culture and Society, December 24, 2016.

(Virtual) The 'Indispensable' Role of Data Brokerage: Bridging Data Fiduciaries and Processors in Modern Data Infrastructure

Tuisha Sircar (IIT Bombay, India)

Data brokerage has emerged as a cornerstone of the modern data infrastructure, occupying a critical space between the roles of data fiduciaries and data processors. This intermediary role is indispensable as data brokerage facilitates data monetisation while simultaneously enabling its movement through complex ecosystems. As multi-sided platforms, data brokers maximize the utility of data through interoperability, which is vital to unlocking its latent value. Inherently dynamic, data gains meaning and utility as it traverses networks, feeding into applications such as programmatic advertising.

Programmatic advertising, an example of efficient data utilization, highlights the monetization of user information in real-time bidding systems. However, the opacity of such systems raises significant concerns, particularly regarding privacy and consent. The growing sophistication of data- brokering platforms reveals a stark tension: while their mechanisms fuel economic innovation and value extraction, they often circumvent users' informed consent, exposing them to privacy violations. High-profile cases, such as Oracle's legal settlement over unauthorized data sharing, underscore the systemic challenges in regulating data flow and protecting individual rights.

Policy frameworks such as the GDPR (EU), the DPDPA (India), and PADFA (US) aim to govern these complex dynamics. By emphasizing consent, data minimization, and cross-border restrictions, these regulations reflect an increasing focus on sovereignty and privacy. However, these frameworks struggle to address the evolving nature of data ecosystems. The dynamic recontextualization of data, where new interactions generate unforeseen value, complicates efforts to restrict data usage to predefined purposes.

The challenges extend to operational dimensions as well. Strict cross-border data-sharing restrictions may safeguard privacy but risk impeding innovation and efficiency in global data-driven business models. This calls for adaptive frameworks capable of reconciling privacy protection with the economic imperatives of data utilization. A promising approach is the implementation of a Data Trust framework. By embedding trustworthiness and accountability into data-sharing mechanisms, Data Trusts offer a solution to the persistent stalemate between market demands and regulatory concerns.

In conclusion, the indispensable role of data brokerage in modern infrastructure requires a nuanced approach to governance that transcends traditional binaries of ownership and control. By framing data trust as a core principle, policymakers can foster an environment that safeguards privacy while enabling innovation. This balance is essential for building equitable, transparent, and secure data ecosystems, ensuring that the transformative potential of data is harnessed responsibly.

References:

? Anthes, G. (2015). Data brokers are watching you. *Communications of the ACM*, 58(1), 28-30.

? Choi, J. P., Jeon, D. & Kim, B. (2019). Privacy and Personal Data Collection with Information Externalities. *Journal of Public Economics*, Vol 173, 113-124.

? Faroukhi, A. Z., Alaoui, I. E., Gahi, Y. & Amine A. (2020). Big data monetization throughout Big Data Value Chain: a comprehensive review. *Journal of Big Data*, 7:3.

? Mishra, S. (2021). The dark industry of data brokers: need for regulation?, *International Journal of Law and Information Technology*, Vol. 29, 395–410 https://doi.org/10.1093/ijlit/eaab012

? Pinchot, J., Chawdry, A. A. & Paullet, K. (2018). Data Privacy Issues in the Age of Data Brokerage: An Exploratory Literature Review, *Issues in Information Systems*, Volume 19, Issue 3, pp. 92-100

? Ramirez, E., Brill, J., Ohlhausen, M., Wright, J., & McSweeny, T. (2014). *Data Brokers: A Call for Transparency and Accountability*. Federal Trade Commission, i-110.

(Virtual) Digital Data Flows and the Re-organisation of the Food & Restaurant Industry

Rashmi Sinha (Indian Institute of Technology, Bombay)

Data brokerage has emerged as a cornerstone of the modern data infrastructure, occupying a critical space between the roles of data fiduciaries and data processors. This intermediary role is indispensable as data brokerage facilitates data monetisation while simultaneously enabling its movement through complex ecosystems. As multi-sided platforms, data brokers maximize the utility of data through interoperability, which is vital to unlocking its latent value. Inherently dynamic, data gains meaning and utility as it traverses networks, feeding into applications such as programmatic advertising.

Programmatic advertising, an example of efficient data utilization, highlights the monetization of user information in real-time bidding systems. However, the opacity of such systems raises significant concerns, particularly regarding privacy and consent. The growing sophistication of data- brokering platforms reveals a stark tension: while their mechanisms fuel economic innovation and value extraction, they often circumvent users' informed consent, exposing them to privacy violations. High-profile cases, such as Oracle's legal settlement over unauthorized data sharing, underscore the systemic challenges in regulating data flow and protecting individual rights.

Policy frameworks such as the GDPR (EU), the DPDPA (India), and PADFA (US) aim to govern these complex dynamics. By emphasizing consent, data minimization, and cross-border restrictions, these regulations reflect an increasing focus on sovereignty and privacy. However, these frameworks struggle to address the evolving nature of data ecosystems. The dynamic recontextualization of data, where new interactions generate unforeseen value, complicates efforts to restrict data usage to predefined purposes.

The challenges extend to operational dimensions as well. Strict cross-border data-sharing restrictions may safeguard privacy but risk impeding innovation and efficiency in global data-driven business models. This calls for adaptive frameworks capable of reconciling privacy protection with the economic imperatives of data utilization. A promising approach is the implementation of a Data Trust framework. By embedding trustworthiness and accountability into data-sharing mechanisms, Data Trusts offer a solution to the persistent stalemate between market demands and regulatory concerns.

In conclusion, the indispensable role of data brokerage in modern infrastructure requires a nuanced approach to governance that transcends traditional binaries of ownership and control. By framing data trust as a core principle, policymakers can foster an environment that safeguards privacy while enabling innovation. This balance is essential for building equitable, transparent, and secure data ecosystems, ensuring that the transformative potential of data is harnessed responsibly.

References:

? Anthes, G. (2015). Data brokers are watching you. *Communications of the ACM*, 58(1), 28-30.

? Choi, J. P., Jeon, D. & Kim, B. (2019). Privacy and Personal Data Collection with Information Externalities. *Journal of Public Economics*, Vol 173, 113-124.

? Faroukhi, A. Z., Alaoui, I. E., Gahi, Y. & Amine A. (2020). Big data monetization throughout Big Data Value Chain: a comprehensive review. *Journal of Big Data*, 7:3.

? Mishra, S. (2021). The dark industry of data brokers: need for regulation?, *International Journal of Law and Information Technology*, Vol. 29, 395–410 https://doi.org/10.1093/ijlit/eaab012

? Pinchot, J., Chawdry, A. A. & Paullet, K. (2018). Data Privacy Issues in the Age of Data Brokerage: An Exploratory Literature Review, *Issues in Information Systems*, Volume 19, Issue 3, pp. 92-100

? Ramirez, E., Brill, J., Ohlhausen, M., Wright, J., & McSweeny, T. (2014). *Data Brokers: A Call for Transparency and Accountability*. Federal Trade Commission, i-110.