DIGITAL FOOTPRINT, GLOBAL FAULT LINES: DATA EXTRACTION FROM THE GLOBAL SOUTH AND THE CASE FOR A BINDING INTERNATIONAL PRIVACY RULE

Piyush Rajpurohit, Jai Narain Vyas University

ABSTRACT

A single search query in Nairobi, a mobile money transfer in Singapore, or a photograph uploaded in Jodhpur each leaves behind a "digital footprint" that persists and travels far beyond its point of creation. This paper argues that the routine cross-border extraction of personal data generated in the Global South, and its processing for value by corporations headquartered in the jurisdictions that also write the governing rules, replicates the structural logic of historical resource colonialism. Drawing on Third World Approaches to International Law (TWAIL) scholarship and the "data colonialism" framework advanced by Nick Couldry and Ulises Mejías, the paper contends that existing national privacy statutes the EU's GDPR, India's Digital Personal Data Protection Act 2023, and China's PIPL are structurally incapable of correcting this imbalance because each operates within domestic jurisdictional limits while the harm they address is inherently transnational. Mechanisms intended to bridge this gap, such as the GDPR's adequacy regime, function less as neutral safeguards than as instruments of digital diplomacy that extend trust along existing lines of economic power, while international trade law treats cross-border data flows chiefly as a market-access question. Using India's "scale without leverage" position and the outsourced AI data-labour supply chain in the Global South as case studies, the paper proposes a binding international privacy rule modelled on the Nagoya Protocol's access-and-benefit-sharing architecture, built on three pillars: an affirmative recognition of data sovereignty as co-equal to free data flow, mandatory cross-border data impact assessments, and a benefit-sharing mechanism for value generated from data or data labour originating in the Global South.

Keywords: Data colonialism, TWAIL; data sovereignty; cross-border data flows; Global South; GDPR adequacy; Digital Personal Data Protection Act 2023; Nagoya Protocol; benefit-sharing; international privacy law.

Introduction

A search query typed into a phone in Nairobi, a mobile money transaction from Singapore, a photography uploaded from Jodhpur India; each of this act generate data that outlive the moment of its creation. Collectively these traces are Described as "digital Footprint" a cumulative record of behaviour, preference and identity that persists long after the user has moved on. What Is less often examined is where that footprint goes, who processes it, and who profit from it?

The answer, increasingly, is that footprint travel far from where it was made. Personal data generated by users in the Global south is routinely extracted, transferred by across the borders and processed by corporations headquartered. The jurisdiction that also happen to write the rules governing how the data may be uses the countries generating the data, by contrast, often have the least capacity to regulate its collection, terms and the share. This asymmetry is not accidental; it mirror with striking structural fidelity the extractive relationship characterized colonial era the resource economics: raw material drawn from the periphery, refined and monetized at the centre with the originating community left the control over either the process or the proceeds.

Scholars have begun calling this as "Data colonialism" *Nick Couldry* and *Ulises Mejías* Describe in the books *The costs of connection as a modern from of historical colonialism.*

This paper argues that the existing patchwork of national privacy status the European union's General data protection regulation (GDPR), India's Digital Personal Data Protection,2023(DPDPA), China's Personal Information Protection Law (PIPL) is structurally incapable of correcting this imbalance, reasoning is each operates within boundaries of its own jurisdiction while the harm it seeks to address in inherently transnational. Mechanism that went meant to bridge this gap such as the GDPR's adequacy decision regime, function less as neutral legal safeguard and more as instruments of digital diplomacy, extending a small circle of "Trusted" jurisdiction largely coextensive with existing economic power, international law. Meanwhile treats cross-border data flow primarily as a market access question under instruments like the WTO's e-commerce moratorium leaving human rights and sovereignty concerns as an afterthought.

The consequence is a governance vacuum precisely where oversight is needed most: at the point where data crosses borders from low-capacity, high-vulnerability jurisdictions into high-capacity, low-accountability ones.

The Historical Parallel- From Resource Colonialism to Data Colonialism

TWAIL and the Colonial Foundation of International Law

TWAIL scholar have argued that international law was not a natural set of rules that colonialism happened to violate, but rather then frame work substantiality conducted though and for colonial project. Sovereignty, property and contract; the core building block of international legal personality were defined in ways that permitted the extension of resource from colonized territories while denying those territories full legal agency to object.

Writers in this tradition, including Antony Anghie and B.S. Chimni, have outlined how doctrines developed to justify nineteenth-century resource extraction did not disappear with formal decolonization; they persisted in modified form through international economic law, trade agreements, and investment treaties that continued to structure unequal relationships between the Global North and South. Their broader argument is major of that the mechanisms of extraction adapt to new technologies and new industries while preserving the underlying power unevenness is directly relevant to the present inquiry, even though these scholars were writing primarily about physical resources rather than data.

The Emergence of "Data Colonialism" as a Framework

Recent scholar like Nick Couldry and Ulises have argued that the collection and commodification of personal data by large elephants of data forest replicate the structure of historical colonialism; the appropriation of resource modern Data rather than land or food from population that neither consent to the terms of extraction nor shares meaningfully in the value it generates.

Their central insight is that it is not merely an economic inequality but a relational one a new social order in which human life itself becomes an input into industrial production through continuous, largely invisible data capture. Under this framing, the "digital footprint" is not an incidental byproduct of internet use; it is the raw material of a new extractive economy, and

the infrastructure that captures it social media platforms, mobile applications, cloud services functions analogously to the trading companies and extraction concessions of the colonial era.

Why this Framing Matters for International Privacy Law?

Current scholar tends to analyse data protection through either a **rights based** lens data protection as an extension of privacy as a human right or a **market-based** lens data protection as a trust mechanism that enables digital commerce. Both are valuable, but neither fully captures the structural, North-South dimension of the problem. A rights based analysis, applied purely within domestic constitutional frameworks, cannot address harms that occur once data leaves the jurisdiction where the right was recognized. A market-based analysis tends to treat all jurisdictions as similarly positioned economic actors, concealing the reality that some jurisdictions write the rules of the market while others merely comply with them.

The data colonialism framework fills this gap. It reframes the question from "how do we protect an individual's data" to "how do we correct a structural imbalance in who controls, processes, and benefits from data at a civilizational scale." the harm being addressed is not a gap in any one country's law, but a gap in the architecture connecting all of them.

Mapping the Legal Architecture Patchwork Built for Data Importers

The European Union’s GDPR id often held up as the Global standard for data protection, and much of the World’s subsequent legislation Including the India’s DPDPA. The GDPR’s own mechanism permitting cross border transfers reveal its Northen centric Design. Under Article 45 European commission may recognize a third country as offering adequate level of protection, permitting data to flow there without additional safeguard. In the absence of such decision transfers require mechanism like standard contractual clauses or binding rules both of which impose significant compliance cost that fall disproportionality on smaller entities and developing country data processors.

Analysis the Domestic Statutes: Digital Personal Data Protection Act,

India’s DPDPA represent genuine advance in codifying data protection rights within there judication but by design each operate only within their own boundaries. The DPDPA regulates how data fiduciaries process the personal data of individuals within India, and imposes conditions on cross border transfer, but it has no mechanism to induce a foreign recipient of

that data a cloud provider headquartered in California, for instance to follow to Indian standards once the data has left Indian jurisdiction, beyond whatever contractual terms were agreed at the point of transfer. Enforcement in such cases depends heavily on the recipient jurisdiction's own willingness to cooperate, which is itself a function of relative bargaining power.

This is not a unique flaw to India's statute; it is an inherent limitation of any domestically legislated privacy framework operating in an international vacuum. The DPDPA's own broad exemptions for government processing under Section 17 further illustrate how domestic political considerations can dilute even a jurisdiction's own attempt at protection a dynamic replicated, in different forms, across many developing-country privacy statutes drafted under pressure to appear "GDPR equivalent" while preserving state access to data.

Case Studies

India: scale Without Corresponding leverage

India presents a particularly instructive case because it combines an enormous data generating population with a domestic regulatory framework that remains, by its own design, limited in reach. The Aadhaar biometric identification system, now linked to a vast range of government and private services, has created one of the largest centralized repositories of biometric and demographic data in the world. While Aadhaar itself is a domestic infrastructure project, the broader ecosystem it has enabled digital payment platforms, fintech applications, and data-driven service providers routes significant volumes of Indian user data through cloud infrastructure and analytics services operated by foreign technology firms.

The DPDPA regulate this processing only up to the point where data leave Indian jurisdiction under agreed transfer condition, beyond that point Indian data subjective have limited practical recourse against a foreign processor’s downstream use of their data, particularly where that processor operates from a jurisdiction with no reciprocal enforcement arrangement with India. The result is a scale paradox: India generates data volumes that rival or exceed those of any single Western economy, yet lacks the adequacy status, mutual enforcement treaties, or comparable leverage that would allow it to negotiate the terms of that data's international treatment on equal footing.

The Global AI Data labour Supply Chain

The current invisible concerns but rapidly growing form of extraction the human labour required to make AI systems function, which is disproportionately sourced from the Global South under conditions that mirror earlier extractive labour arrangements. Large AI companies have relied on outsourced content moderation and data labelling workforces including workers in Kenya and other lower-income countries to interpret, filter, and moderate the raw data used to train and refine AI models, frequently for low wages and with significant psychological exposure to disturbing content, while the resulting models and the value they generate are owned and monetized entirely by firms headquartered in wealthy jurisdictions.

This case study stronger the greatest anxiety which extends the data colonialism framework beyond personal data narrowly defined, to the broader data supply chain that increasingly supports the global AI economy. It illustrates that the extraction pattern is not limited to passive data collection from ordinary users; it also encompasses the active labour of Global South workers whose contribution to refining and validating that data receives a small fraction of the value it ultimately creates.

A Binding International Privacy Rule

The Nagoya Protocol: A Benefit Sharing Precedent.

International law has already confronted a structurally similar problem in a different domain the extraction of genetic resources and associated traditional knowledge, historically taken from biodiversity-rich developing countries by pharmaceutical and agricultural companies based in the Global North, without compensation flowing back to the countries or communities of origin. The Nagoya Protocol to the Convention on Biological Diversity, adopted in 2010, responded to this by establishing binding principles of access and benefit sharing country providing genetic resources is entitled to a share of the benefits arising from their commercial use, and prior informed consent is required before access is granted.

Genetic resources are limited and physically bounded in a way data is not but the Protocol's underlying legal architecture is transferable. It establishes that resources of significant value to global commerce, even when accessed voluntarily and legally, generate an obligation of

benefit-sharing running back to the point of origin. A binding international privacy regime could adopt an analogous structure.

Core Principle for the Proposed Rule

1. An affirmative recognition of data sovereignty as a standalone principle of international law not merely a permitted exception to trade liberalization, but a right co-equal to the free flow of data, requiring the two to be balanced rather than one treated as default and the other as departure.

2. mandatory cross-border data impact assessments, demonstrated loosely on environmental impact assessment regimes in international environmental law, requiring an entity transferring data out of a lower capacity jurisdiction to assess and disclose the human rights and developmental implications of that transfer before it occurs, rather than after harm has materialized.

3. a benefit-sharing mechanism, ensuring that value generated from data or data labour originating in the Global South is not captured entirely by the processing jurisdiction.

The Feasibility Critique

The most obvious objection to this proposal is political feasibility: the United States and China, the two jurisdictions with the greatest capacity to process global data, have historically resisted binding multilateral data governance instruments that would constrain their domestic technology sectors, and any treaty lacking their participation would address only a fraction of the problem. This is a genuine and serious limitation, not one to be dismissed. However, the same objection could have been raised and was raised against early efforts toward the Nagoya Protocol, the Paris Agreement, and other multilateral instruments eventually joined, however imperfectly, by major economic powers once a critical mass of developing and middle-power states coalesced around a shared normative framework.

Conclusion

We began with an image a search query typed in Nairobi, a mobile money transaction Singapore, a photograph uploaded from in Jodhpur ordinary acts that leave behind a digital footprint travelling far beyond the person who created it. However, journey is neither

accidental nor neutral. It follows a structural pattern in which data generated in the Global South is extracted, processed, and monetized predominantly by actors situated in jurisdictions that also write the rules governing that process a pattern that bears more than a passing resemblance to the extractive relationships that defined earlier centuries of resource colonialism. The theoretical foundation for this claim, situating it within TWAIL scholar's longstanding critique of international law as a structure historically built around, rather than despite, colonial extraction, and within the emerging "data colonialism" literature that applies a parallel critique to the digital economy. The Data is new gem, weapon and diplomatic talk the protection and legality is essential in the end, need not remain a passive metaphor a trace left behind, unowned and unclaimed once the moment of its creation has passed. It can instead be understood as an extension of the individual and the community that generated it, deserving of the same principles of self-determination that international law has, however imperfectly, extended to other resources. Building the legal architecture to make that a reality is the unfinished task this paper has sought to define, even where it cannot yet resolve every question of institutional design or political will required to complete it.

Reference

1. Antony Anghie, Imperialism, Sovereignty and the Making of International Law, Cambridge University Press 2005.

2. B.S. Chimni, 'Third World Approaches to International Law: A Manifesto' (2006) 8 International Community Law Review 3.

3. Nick Couldry and Ulises A. Mejias, The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism

4. Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological UNTS 3008

5. General Agreement on Trade in Services (adopted 15 April 1994, entered into force 1 January 1995) 1869 UNTS 183, art XIV

6. General Data Protection Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 [2016] OJ L119/1.

7. Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act 2016 (India)

8. Digital Personal Data Protection Act 2023 (India)

9. Personal Information Protection Law of the People's Republic of China (2021)

10. Reports on outsourced AI content-moderation and data-labelling labour in Kenya