Data Colonialism and Data Localization: Issues and Challenges

R. Subramani

Associate Professor, Department of Journalism and Mass Communication
Periyar University, Salem, Tamil Nadu, India
  https://orcid.org/0000-0003-2763-3396

Abstract

The zeitgeist of the 21st century is warped, managed, and altogether driven by the slew of technological applications. They have usurped every activity of our life in its hold in such a way that wherever we turn, we find them interfacing, navigating, and monitoring human life. The juggernaut of trade, commerce,and marketing made the entry that has taken this world in its firm grasp and the engine of marketing is consuming data of every citizen as its fuel. This function invariably brings in vast profits to large corporations and multi-national establishments which do not feel it necessary to take the consent of citizens and share the windfall they gain. This new order is literally thriving by harvesting, poaching, and appropriating citizens' life (data) which serve them as fodder or oil that spurs and keeps them financially in good stead. With the end of colonial era across the globe, gone are the evils associated with it, and here comes, new masters in the garb of multiple technologies, who are equally vicious, awfully exploitative, and with dour demeanour dominate our lives. Frankly speaking, now we are sort of re-colonized as our data is collected and used in the premise of bettering our lives. Of late, Global South's some of the developing governments in order to secure their local interest are pushing for localizing data, but this stance has many pitfalls attached to it as data localization will give increased access to the government, which may put its own people under surveillance for varied reasons. Therefore, at this juncture, human beings are caught up between the cleft stick. On one hand, data colonialism uses them as commodities (commoditization) while data localization makes them the subject of State surveillance. Hence, this paper comprehensively examines the corresponding complex issues and thoroughly peruses the embedded challenges straddling data colonialism and data localization.

Keywords: Data, Colonisation, Localisation, Commodification, Digital, Extraction.

Introduction

Colonialism is the systematic dynamics that generated colossal wealth including massive land across the world which in turn has bigoted the very concept of capitalism. Capitalism has evolved into a unilaterally powerful global social system that involves human beings in particular modes of economic production primarily for the benefit of elites. (Couldry and Mejias) adds that capitalism always lacks a sense of equity in terms of the distribution of labour and wealth global. Therefore, we find capitalism at the centre of colonialism and at its roots they are intertwined, hence, colonialism is an extension of capitalism where data unequivocally replaces land, labour, and natural resources across the globe. We can easily draw a parallel line between colonialism and capitalism, where colonialism found to be dispossessing, collecting, extracting and owning what belonged to others regardless of their rights. Hence, colonialism acts as colonizers who have always been taking, claiming and dispossessing what belonged to others by use of power and deception. Consequently, extraction of data is believed that nothing but poaching human life itself such as snatching and claiming the fragments of one's everyday existence in the form data.

When governments, non-governmental organizations, business houses and large corporations grip, own, poach, plunder, steal and acquire human information termed as data and claim ownership of it may be termed as the colonization of human life. (Thatcher et al.) opine now through the integration of technologies such as mobile phones, computers, and other portable devices into our everyday life, personal aspects of our lives are appropriated to use for economic gain (commodification of data). According to (Reardon and Tallbear), we are openly dispossessed of our resources. For instance, the DNA data of indigenous people are collected by various entities under the premise that it will be used to ameliorate their lives who are made to partake in so-called studies; yes, some researchers are sincere but there is no warranty that such appropriation of data will certainly not be misused in future. (Couldry and Mejias) strongly feels that every layer of human life has inexorably become the target of profitable extraction which is intrinsically built upon the capitalistic assumption that access to data is pure business and economic imperative. This very concept has made innumerable data barons across the globe menacingly predatory who do not care about the right to privacy while poaching people's data for economic gains. This blatant plundering of individual data can be equated with contemporary capitalism and this extraction can discriminate against societies most vulnerable through the automation of inequality. And today, further he adds that capitalism and colonialism are firmly spliced as data collection has become the default model of capitalism. A very recent example of Osaka Track – a global standoff on who stores our data, wherein June's G20 summit held in Japan, developing nations such as India, South Africa, and Singapore had boycotted the declaration as they discovered there was no opportunity to put their interest about data localization. Way back in 2016, data was touted as the new oil and in recent times global negotiations on data flow marred into controversies as some 'big nations' dislike the assertion of developing nations to claim over the data generated by their citizens. Now a government asking big tech companies to build data centres on their soil, but the cloud that powers much of the world's tech industry is grounded in vast data centres located around northern Europe and US coasts. If we take the case of Facebook, India might have the biggest amount of Facebook users but of the total 15 locations of Facebook's data centres, 10 are in North America, 4 are in Europe, and just one in Asia – Singapore. This disagreement between the location centres and new resources of data led to the accusation of data colonialism and the very logic of developed nations over digital industrialization to local cloud companies is unsatisfactory, hence, the controversy. Data colonialism is a double-edged sword, on one hand, it brings monetary value to businesses but on other hand, it facilitates heightened access to governments to sensitive personal information of people that further perpetrate surveillance opportunities on its citizens.

In this digitalized world, when it comes to the real privacy and security of their sensitive data, human beings are reduced to mannequins. Every time we swipe debit or credit cards at an ATM or store, purchase an air ticket, or use a GPS device our data gets commodified. Presently, every sphere of our life is governed by digital technology and whichever site we browse in asks us about key information like names, birth dates, addresses, and phone numbers, these things have become prerequisites without which we cannot move further. Today the temptation of internet is all pervasive and to gratify this alluring addiction one don't think twice in furnishing personal information to multiple sites that's where data appropriation happens that reduces us to a digital slave and the winners are large businesses. Companies engage in trans-border data transfer using data nodes located in different nations; the internet easily locates those modes and funnels the data through nearby centre. So, unless full-proof legislation is in place, it will be difficult to deal with this menace causing both national and international distress.

Data Colonialism: Issues and Challenges

The Canada-based erudite Professor Scott Prudham says, when the definitive purpose of production of 'something' is market exchange or sales and not use that invariably gets transformed into a commodity. He further adds it could be fun to be with a child hand-drawing a map of the

city, but hand-drawing a map for sale to tourists is nothing but producing commodity. It may not bring in humongous profit but when the data is global in scale, it may fetch uncountable gains for large businesses. Consequently, the very distinctive gap between the usage of any act that gets converted into a commodity serves as a function for the large corporations as they conceal the extractive nature of our association with spatial and digital technologies. Data produced by a single user at a given time and place such as posting a picture of a meal with a pal at the riverside on Facebook may have no value or worth unless it is linked to the user's past data, user's network of other users, location data, many users' data embedded with spatial patterns and sequential rhythms. Hence, the conversion of the individual to an aggregated, digital commodity dictates connecting data across users, times, and spaces, and therefore, the amalgamation of such data becomes big and it necessitates algorithmic selection, interpretation, and analysis as to what to include and exclude. It can be termed as the capital process through which different firms and big business establishments overtly poach and colonize our everyday life (data) for monetary gain which ultimately puts our privacy at peril (Thatcher et al.). In this highly technological world, at times, we might be feeling empowered but the harsh reality is that subliminally every action and reaction (data) is made saleable by umpteenth of tech and business houses. This very predatory nature of dispossessing individual's data with or without consent is metaphorically termed as 'data colonialism' and this newly coined euphemistic term - 'data colonialism,' clearly reveals and equates with the imperialistic ethos that systematically appropriates and confiscates people's lands, personal information, properties, identities, etc. Therefore, the profound arguments asking for a literal understanding of data colonialism by (Couldry and Mejias) ultimately insinuate at globally played out variegated ways and patterns where colonial legacies inexhaustibly underpin 'modern capitalism'. Data Colonialism is the new emerging order where human life and their data are appropriated for profit; colonialism and its modes, intensities, scales, and contexts are different and over but its heritage and legacy linger on. So, there are some left-over rationalities we are living with few of them has an economic dimension. Peruvian sociologist (Quijano) says that today colonialism is over and there are no colonies per se but its legacies and heritages linger on and that is what is being termed as coloniality of power. If we look into economic, racial, and terrorism all are said to be shaped by the heritage of colonialism; thus, the modes, intensities, scales, and contexts are different yet the core function remains the same and that is to dispossess and these realities carry with them some rationalities which include economic rationalities and if we think about the progression from cheap, nature, labour to data and the colonizers have always framed the world being made of natural resources which can be exploited. They floated the idea that data are exhaust, abundant, and free which can be used and appropriated. Therefore, there are technological rationalities that tell us more data is good for the progress of society, etc.

The idea of colonizers made to believe in the masses that the land they were invading was 'terra nullius' means no man's land, totally unclaimed, ready to be claimed, owned, and free to use, likewise along with it came the notion of free labour and abundant natural resources which was free to be used. So, they used appropriated land, properties, and natural resources to make their lives better. Ultimately, they would portray themselves as the saviour and benefactors of the people of that land annexed by them.

Likewise, when our data is extracted, taken, or asked for further use by different firms, they give the process of data accumulation a semblance of a neutral and natural phenomenon as if data is meant to be readily available for collection. The analysis of this queer and weird process speaks volumes about the exploitative nature exercised by large corporations which is nothing but merciless plundering of people's identities, behaviours, and sensitive personal information for economic gains while the owners of data get no benefits. (Schneier) says that the value produced by smart technologies does not necessarily come by buying it but rather you using it. At this juncture, opines that smart technologies can equally be viciously perilous; the function of a normal refrigerator is to keep food cold, but a smart refrigerator not only keeps food cold, but

it also knows what brands one consumes, at what time we eat and whether our food is expired or not so, it take our data and send targeted advertisement and recommend sponsored recipes; such fridge can also be used as surveillance device by legal agency if they want to sneak peek into our house. We all know that smart technologies are embedded with sensors, processors, and network connections.

The way technology firms captivate and conjure customers to acquiesce in to their complex terms and conditions by clicking to end-user licensing agreements which are less of consent and more of compliance is very unscrupulous. (Schneier) states that before the invention of the internet, social life was confined into company records, responses to direct marketing, government public records, credit bureau data, insurance bureau data, etc. Now within the complex framework of data processing, we have an inbuilt social quantification mechanism that exceedingly harvests data from everyday life from different social forums. It begins with the routine of self-tracking ourselves on various social platforms including those icons from across the social spectrum. It is digital platforms – various social media forums provide technological means to produce diversified data that are tracked, sorted, counted, and harvested for economic gains. So, social media is one form of platform where extracted data are commodified under data colonialism. Furthermore, (Cowen) adds data-driven logistics where management goods' movement within the supply chain is managed by human and non-human parties are the second conduit that allows continuous data collection. (Levy) says the third way in which human life is poached or appropriated is through social relations such as social security and insurance where contractual commitment is involved. It is digital platforms – various social media forums provide technological means to produce diversified data that are tracked, sorted, counted, and harvested for economic gains. So, social media is one form of platform where extracted data are commodified under data colonialism. Recent critical data studies state (Eubanks; O'Neil), that many people quite often voluntarily or consensually submit their data such self-data collection gives rise to the basis for discrimination and inequality. (Couldry and Mejias)

say, if we get deeper into Marx's social theory the palpable insight will make us understand the wider colonial appropriation and its social permutation. For him, commodification is labour, which was once just worked and productive enterprise that further gets commodified or metamorphosed as labour power. Hence, it will help us to understand the new function of commodification concerning data production and consumption. Retrospectively speaking, we know the fact that data is not inevitably extracted but at some points in time, we might have given consent or consented to browsing through the internet and a whole lot of social media platforms.

Data Localization

With the advent of the internet, the entire ensemble of life that includes global economy, trade, businesses, education, healthcare, and potentially life-related arenas got transformed for good. With rapid progress in the technological world, we start getting an answer to all impending queries, problems, doubts, and scepticism, and finally technologies get permeated in everything we do. From 1992-2017, worldwide internet networks started with 100 gigabytes (GB) per day traffic that has reached over 46.6 terabytes per second, and as of now, this traffic might have crossed the reach of 150.7 terabytes per second. The businesses that utilize the internet to trade both nationally and globally have had tremendous longevity and sustained success rates. Internet being the pivotal fulcrum, obviously allows large business houses to process colossal data across national borders which fetches promising returns. At this time, where the digital economy occupies the centre stage, we are bearing the brunt of the free flow of data in getting our privacy breached or overtly violated. This potential issue has a massive socio-economic dimension which if not addressed will reduce human life to zombies. Now all businesses are data-driven and multinational companies are generally from developed nations often telling underdeveloped and developing nations to part with the data of their citizens for their betterment. Data which is now referred to as new oil for businesses needs uninterrupted flow but of late many nations over their key concerns are vehemently pushing for an idea of localization of data that will optimize data

access to national security agencies. So, to arrest the loss of economic profits to large foreign might lead to violation of privacy rights and unintended foreign surveillance. These perceived risks made various countries agree to the idea of data flow regulation and in the last two decades most of them opted for the notion of localization of data. In such a situation, one school of thought says data localization will restrict the free flow of data thus limiting physical storage and further processing within the given jurisdictional boundaries. This idea of localization can be classified into two categories – soft and hard localization; hard localization needs local territory for storage and processing of data, which inhibits cross-border transfer of data; soft localization requires local storage lets data be transferred and treated outside national borders. These two broad categories further get sub-categorized into conditional, sector-specific, and general localization, which implies different sets of rules on bilateral and multilateral collaborations.

Data Localization Across the Globe

The earliest attempt to localize data was made way back in 2005 when Kazakhstan promulgated a law that made it mandatory to save data on domains called – 'kz' in Kazakhstan. In 2015, Russia introduced Federal law which required operators while processing personal datato use servers located in Russia. Thereafter, several countries floated the idea that technology companies must store data in centres located within the country. The need of data localization duly incorporated into digital governance frameworks is a much-debated issue that equally concerns autocratic and democratic nations alike and it varies from their respective standpoints. There are various national security issues such as information sharing, cybersecurity, and the struggle to battle against the tools of digital authoritarianism. These genuine mandates are routinely implemented under the pretence of national security as we have noticed even established democracies have faltered to strike a balance amid the allure of digital sovereignty facilitated by data localization with economic and security arguments all around. We know nations talking about threat to trade if data cannot flow freely across the borders but there has not been much discussion on national security in the wake of data-

flow strait jacket. Data localization means certain data of citizens of the country, which could be personal, health-related, financial, educational, or commercials to be stored on physical infrastructure within that nation's border. More importantly, data must be categorised such as personal data, sensitive personal data, and critical personal data, and different levels of restrictions and permissions must be applied to it. Data localization comes into different manifestations wherein most approaches can be classified into hard, soft, hybrid, or de-fact localization. China made security review an imperative step before data gets stored within its border and transferred as well. We will find the echo of this modus operandi in both India and Pakistan's Personal Data Protection Bill, wherein case of sensitive personal data transfer has certain preconditions. Business, human rights, and data privacy are the core issues that concern the stakeholders. Some democracies use security priorities like counter-terrorism and curbing foreign influence as a subterfuge to tighten control on their national digital infrastructure which ultimately enable them to sustain increased surveillance and censorship of their population. All five emerging markets referred to as BRICS (Brazil, Russia, India, China, and South Africa) are contemplating localizing their data. Brazil has allowed its legislation into force in that regard. Russia is aggressively going to win its internet infrastructure and under data mirroring policy it is cracking down on social media firms that are involved in business sans its physical presence in Russia. To counter the Brussels Effect, China has come up with the Beijing Effect which offers different models of data governance regulations. South Africa also started with its data governance frameworks with explicit data localization requirements.

Data Localization in India

Over the years, having passed sector-specific localization measures, the Indian government is now open to economy-wide proposals. And advocating for the location of data, it has emphasised that if consumer data is processed within the country, it will certainly fetch the perceived economic benefits. India has already approved some patchwork measures for data localization in the economic sector and it begins with the Reserve Bank of India

requiring all payment data to be stored in India but it is open for data to be taken out for processing. Telecommunication sector data needs to be localized and it does not allow subscribers information transfer of accounting information. Despite its vigorous push for data localization by opposing moves by Western businesses to make a profit, the Indian government wants international technological giants like Facebook and Google to store sensitive data of users within the nation. India had a very ambiguous response to data privacy but in 2017, Supreme Court ruled that the Indian Constitution guaranteed a fundamental right to privacy for everyone in the nation. During the discussion convened by the Joint Parliamentary Committee, where Indian government officials and executives of global tech giants like Facebook, Google, and Amazon discussed the issue of data localization which refers to the storage of data. In the discussion, one version says that data only can be stored in India and the other is that copies of the data must always be available in India. The saving and transferring of data depend on the three categories namely, personal, sensitive persona, and critical personal data. The Indian government on August 11, 2023, enacted its new privacy law – the Digital Personal Data Protection Act (DPDP Act) which will replace relevant provisions of IT Act 2000, IT Act 2008, and the Information Technology Rules 2011. It applies to digital personal data, it has extra-territorial application, and the act imposes various obligations on data controllers like consent, notice, record-keeping, data transfer, data breaches, data of children and persons with disabilities, exemptions, etc. The Act provides for various rights of data subjects, including rights to access, data correction, deletion, and grievance redressal, and prescribes penalties of up to 250 crore rupees ($30 million) for non-compliance. With regards to data localization, the Indian government is also contemplating some corresponding alternatives like granular data localization alternatives. At the outset, the first option called Conditional Mirroring Localization instead of asking for full local storage and processing would entail global storage and processing with a requirement for local mirrored copies (for critical personal data). Similarly, the second option, Conditional Soft Localization, asks for local storage and global processing. This option may allow personal data to be taken out of India for processing. Under the third optional variant, Conditional Hard localization would entail local storage and processing only for critical personal data.

Opportunities and Challenges of Datafication

Datafication offers a wide range of opportunities and advantages to organizations and individuals. It requires new skills that enable us to identify new trends, develop a new range of products, and services, and finetune decision-making. It also opens new vistas for social change as data can be used to address new social and environmental challenges in the areas of healthcare and climate change issues. While engaged in the above functions we must try to alleviate the chances of bias and discrimination that are likely to sneak in. It also needs security measures in place to discourage the chances of misuse and unethical practices. Datafication enables us to go beyond the traditional way of decision-making; it helps us to establish an innovative algorithm as a pivot to identify new emerging trends and patterns that very often sharpen and strengthen decision-making thus pushing us ahead of others. It helps us to know and understand preferences that further enable an organization to tailor-made products suiting the needs of customers and improve the quality of products. Datafication facilitates an option for data monetization by selling data to other organization. Artificial Intelligence and Machine Learning have changed the way an organization collects, stores, and processes data. The Internet of Things and Blockchain are the two most emerging technologies. IoT is related to things that are embedded with sensors, software, and other technologies that enable the collection and exchange of data with other devices over the Internet. Blockchain is associated with ledger technology that significantly influences datafication allowing the secure, transparent, and tamper-proof transaction.

Conclusion

Having scrutinized data colonialism, we could seemingly understand, grasp, and get acquainted with its inherent opportunities and underlying challenges

staring us in the face. Data colonialism's unbridled control that dispossesses us of our personal life has compelled different developing nations including India to go for data localization. In this contemporary socio-political environment, this concept is gaining traction which not only will establish sovereignty over own data, but also give liberty to use it for own benefit and create a conducive environment where we will be at ease about our sensitive, personal, and critical data. Most importantly, instead of launching a broadside against the Global North for the status quo, Global South nations must set their priorities in perfect order with a strategic strategy to iron out existing gaps. Nations from the Global South must map data gaps and prioritize their needs. A system involving clear guidance and standards for governing public-private data collaborations that will protect citizens' self-interest must be the priority. Now the time has come when governments and citizens of the Global South establish a shared understanding of the risks and benefits of their power of data and fast digitizing world. And put in place a system wherein local academics, government officials, and others must actively manage new data resources. Global firms and other data actors should commit to building sustainability in local knowledge generation and data production in low-income countries. We must support nations advancing in data-driven government strategies through good digital governance procedures. Global South's respective governments need to form regulatory policies stressing localizing citizens' data that must be collected and stored in a centre situated within the host country. We should find an opportunity to nationalize data and make large corporations pay for it, because presently technological giants are converting social life into revenue streams, and have no qualms about amassing huge profits while their products might be unwittingly spreading bias or discrimination. Not only Google, Facebook, Amazon, and Apple are neck-deep into data colonialism but China with its robust digital surveillance and Artificial Intelligence is another centre of power in data colonialism. Critics are of the view that US and Chinese firms to establish imperial control over the digital ecosystem is likely to cause lopsided influence over the economics, culture, and politics of the nations from the Global South. In this fast-paced world, various technological giants from developed nations are continuously spreading their footprints across the globe.

References

Cairns, David, and Shaun Richards. *Writing Ireland: Colonialism, Nationalism, and Culture*. Manchester University Press, 1988.

Couldry, Nick, and Ulises A. Mejias. "Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject." *Television & New Media,* vol. 20, no. 4, 2019, pp. 336-349.

Cowen, Deborah. *The Deadly Life of Logistics: Mapping Violence in Global Trade*. University of Minnesota Press, 2014.

Eubanks, Virginia. *Automating Inequality*. St. Martin's Press, 2018.

Levy, Karen. "The Contexts of Control: Information, Power, and Truck-driving Work." *The Information Society,* vol. 31, no. 2, 2015, pp. 160-74.

Lippold, Walter, and Deivison Faustino. "Digital Colonialism, Racism and the Primitive Accumulation of Data." *Germinal: Marxism and Education in Debate*, vol. 14, no. 2, 2022, pp. 56-78.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Penguin, 2016.

Quijano, Anibal. "Coloniality and Modernity / Rationality." *Cultural Studies*, vol. 21, no. 2, 2007, pp. 168-78.

Ramachandran, Sanjana, and K. Neetha. "What is Financial Data Extraction?" *Nanonets*, 2023, https://nanonets.com/blog/financial-data-extraction/

Rawat, Seema, et al. "Application of Machine Learning and Data Visualization Techniques for Decision Support in the Insurance Sector." *International Journal of Information Management Data Insights,* vol. 1, no. 2, 2021.

Reardon, Jenny, and Kim TallBear. "Your DNA is our History: Genomics, Anthropology, and the Construction of Whiteness as Property." *Current Anthropology*, vol. 53, 2012, pp. 233-245.

Schneier, Bruce. *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. W.W. Norton, 2015.

Schröttle, Monika, and Ksenia Meshkova. "Data Collection: Challenges and Opportunities." *Femicide across Europe: Theory, Research and Prevention*, Bristol University Press, 2018, pp. 33-52.

Thatcher, Jim, et al. "Data Colonialism through Accumulation by Dispossession: New Metaphors for Daily Data." *Environment and Planning D: Society and Space,* vol. 34, no. 6, 2016.

Yayboke, Erol, et al. "The Real National Security Concerns over Data Localization." *Center for Strategic & International Studies,* 2021.

Author Details

R. Subramani, *Associate Professor, Department of Journalism and Mass Communication, Periyar University, Salem, Tamil Nadu, India,* **Email ID***: erasubramani@gmail.com*