Data Sovereignty for Planetary Sustainability: Mapping the North-South Concerns

Sidiq Madya

Australian National University, Canberra ACT 2601, Australia

sidiq.madya@anu.edu.au

Abstract. The idea of data sovereignty has gained traction over the last decade. Data as a new resource is increasingly vital for decision-making in businesses, governments and societies. Proponents of data sovereignty resists exploitation of personal data especially by state and corporate power. The exploitation of data in a form of 'data colonialism' has created a range of risks that threaten the lives of people and the planet. This study seeks to examine how data sovereignty as a socio-political concept is discussed by civil society organisations in the Global North and South and how it relates to the agenda of planetary sustainability. Using textual data produced on websites and semantic network mapping as an analytical tool, this study explores and contextualises data sovereignty issues in the frame of the North-South divide. Comparing issues emerged online, the diffferent concerns were identified. This study found that the Northen organisations are more concerned on issues such as 'indigenousity', 'rights', and 'privacy'. Meanwhile, 'protection', 'trade', and 'development' are among prominent issues in the South. Understanding these differences requires recognition of unequal relation between North and South in a current data economy, even when the shifting context such as US-China domination in 'data realms' enters into debates. The finding suggests that to enhance the global resilience through realisation of data sovereignty, the notions of 'planetary sustainability' must be pushed forward in parallel with reconsideration of sovereignty of the data subjects in the datafied societies.

Keywords: Data Sovereignty, Planetary Sustainability, North-South Divide, Data Colonialism, Semantic Network Mapping

1 Introduction

Inequalities between the North and South are extended in the age of data-driven economies. Traditionally viewed as the wealthier-dominant global power, the North is the earlier player that capitalises the exponential growth of data as an economic resource through the creation of digital platforms mainly targeting the South users as a potential market (Arora 2019; Kwet 2019). Meanwhile, the Global South often get themselves the

A. R. Cangara and S. Muhammad (eds.), *Proceedings of the Asia Pacific Sociological Association Conference (APSA 2024)*, Atlantis Highlights in Social Sciences, Education and Humanities 35,
https://doi.org/10.2991/978-94-6463-680-2_6

poorer-marginalised which is losing control of making the most of their own data ecosystems. While people in the South are becoming and being the data producer, few big companies in the North maintain their dominance through data extraction and exploitation. Such divide recalls back the memory of traditional form of imperialism where natural resources were extracted from colonised lands to build 'civilisation' in other parts of the world. Resembling a form of traditional colonialism, data-driven economies are expressed neologically as a practice of 'data colonialism' (Thatcher, O'Sullivan, and Mahmoudi 2016; Couldry and Mejias 2019).

Data colonialism describes exploitation of the digital data of individuals and their social relations by few big entities mainly for power and profit. Concentration of power and profit mainly in 'big countries' leaves 'the rest of the world' behind. This new form of global dominance through control over data illustrates perpetuation of historical patterns of dependence and inequality. However, what is somewhat overlooked in the frame of 'data colonialism' is that data extraction by few big tech companies is not only depriving social relations but also creating harm for the planet. Planetary cost of data-driven innovation such as Artificial Intelligence is environmentally degrading (Markelius et al. 2024). It has been widely known that building and running computing infrastructures takes a high amount of energy that ends up in e-waste (Robinson 2009). What is metaphorically known as 'the cloud', for example, is in fact material as it requires grid electricity from coal, gas, nuclear, fossil, or renewable energy which affects the environment and climate (Crawford 2021). The widespread belief in data as a new resource poses a threat for the people and the planet if managed unwisely.

In light of this, the term 'data sovereignty' has gained prominence as a countermovement advocating the rights of data subjects (Hummel, Braun, and Dabrock 2021; IGF 2022), particularly in the Global South to govern their own data. As a state-domain, data sovereignty calls for political authority over the digital data to subject data flows into a national jurisdiction (Polatin-Reuben and Wright 2014). Proponents of data sovereignty challenges the existing dominant control exercised by transnational corporations, often headquartered in the Global North, which harvest and process vast amounts of data often in the absence of consent, regulation, or orientation for protecting data subjects and the environment. As an individual domain, the discourse on data sovereignty underlines the ability and rights of data subjects to access and control data they produce as data is increasingly intertwined with identity, behavior and citizenship. In an increasingly datafied society, data is tied to social lives of people, their interaction and behavior which shape their opportunity and constrain in the digital age (Beaulieu and Leonelli 2022).

The fight for data sovereignty, therefore, goes beyond mere state control or individual-person self-determination, it includes cultural and social dimensions. For many civil societies, such as data activists indigenous or marginalized groups, reclaiming control over their data is part of a broader struggle to assert their ability and rights against forces of globalized surveillance capitalism. Civil society elements and organizations play an influential role in advocating for data sovereignty, acting as critical voices in the fight for not only global data governance reforms but also its reconsideration for future

sustainability of the planet. These groups which often representing marginalized communities, activists, human rights organizations, and social justice advocates mobilise their concerns in many ways including through leveraging online technologies to raise awareness about the exploitation of data and the importance of global power shifting to control data.

Online technologies such as the web which are increasingly common to mobilise public concerns, are used by civil society organisations to fight back against the monopolisation and monetisation of data by powerful tech corporations and governments. Their advocacy is crucial in shaping public opinion, mobilising collective actions, and pressuring policymakers to create more just and sustainable data infrastructures and flows. In contemporary society, capturing the concerns and narratives of civil society organizations regarding data sovereignty is essential. These groups offer alternative perspectives on the consequences of unchecked data extraction, highlighting its impacts on a range of issues such as privacy, rights, security and data subjects' sovereignty. Understanding these narratives helps create a better picture of the data sovereignty movement, offering a bottom-up approach to the issue that centers the voices of those most affected.

This study seeks to address the following research questions: what are the issues around data sovereignty that matter most for organisations across the North and South? How the agendas of 'planetary sustainability' are related to the notions of data sovereignty as narrated by organisations online?

2 Methods

This study uses online data publicly available to map and explore issues around data sovereignty in relation to planetary sustainability produced by organisations' official websites that concerns with those issues. Organisation websites were in the first place identified through Google Search engine and followed by tracing their hyperlink networks. The key phrase 'data sovereignty' is used to identify organisations websites engaging with this issue. After information retrieval of search engine reaching a saturation, where the same result keeps appearing, searching was stopped and moved to crawling their hyperlinks. Tracing their hyperlink networks helps identify other potential organisations that might not be indexed in search engine result. A list of domain names of organisations then crosschecked using the Yearbook of International Organisations and the United Nations' list of NGOs (esango.un.org) to make sure that the websites are officially owned or managed by organisations.

Organisations must be global in scope (international organisations) is the initial inclusion criteria to recruit participants as this research seeks to gain an understanding of organisations' engagement with the issues in the global context. However, the process of identification of organisations was not restricted to global-scale organisations. Local organisations identified through the hyperlink networks were included. After completing the collection of data on actor/organisation as proxied by the domain name of the

organisations, data on issues as represented by the text produced on organisation websites or webpages were extracted. Presentation of the text on the websites have no standard format which creates a challenge to sort out. For example, organisations might post a short description about a seminar event on their web discussing ‘data sovereignty’ so that such content were identified during data collection process. Some websites mentioned the term ‘data soverignty’ but without further discussing it. This research only includes text content which is posted in a form of opinion or report format and must contain a ‘substantial’ discussion about ‘data sovereignty’ issues.

A compilation of text retrieved from organisation websites were treated as ‘official’ online documents as they were published on their websites regardless of the author. Relevant texts satisfy inclusion criteria were then pre-processed for analysis. There are 171 online documents produced by 126 organisations included for analysis in this research. Each document and organisation were then given their attributes based on classification of whether they belonged to the North or South. Countries where organisations headquartered become a proxy to base this categorisation. Assigning attributes results in an imbalance of proportion between North and South organisations.

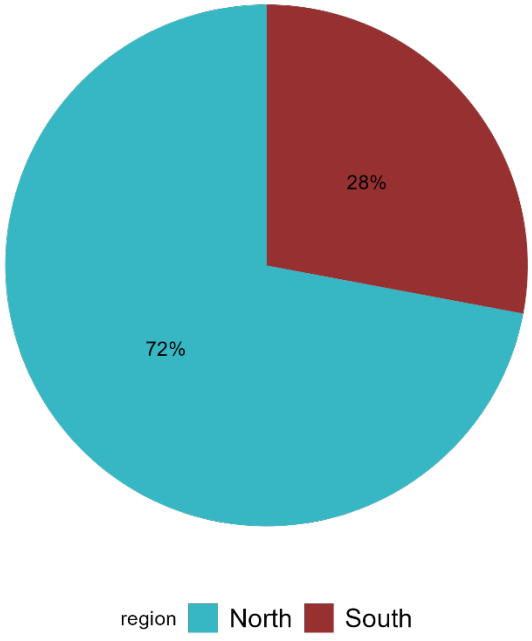

Fig. 1. Proportion of organisation based on regions.

Online documents that have been collected were then converted into a large corpus for text analysis. This research uses a collection of packages and a line of codes in R to pre-process and analyse data. Using a combination of text mining and semantic network analysis, this study identified and mapped what are the key terms emerged when data sovereignty and planetary sustainability issues were discussed online.

Table 1. Table captions should be placed above the tables.

	North	South	Corpus mean
Number of terms	1474.14754	1073.95238	1274.04996
Number of unique terms	679.26230	542.42857	644.21951
Percent of unique terms	46.07831	50.50769	54.33892
Number of hapax legomena	438.29508	363.57143	419.15854
Percent of hapax legomena	29.73210	33.85359	38.16745
Number of words	2432.40984	1741.85714	2087.13349

The initial stage of analysis is mapping the key terms used by organisations based on the frequency of words. This research uses a sub-corpus on ‘data sovereignty’ documents discussing ‘planetary sustainability’ issues. Collecting the North and South documents results different composition of the terms. Documents published by the North organisations have more terms than the South. Each document is composed of 2087 words on average which is safe to claim as a typical website article or a blog post. Frequency analysis allows identification of individual terms deemed prominent in a large corpora. Semantic network analysis makes it possible to capture how these key terms are interrelated and linked to each other which later provides the general picture of opinion formation in the corpus.

Using exact word or phrase to map the emergence of issues, semantic network analysis in this research takes what (Yang and González-Bailón 2016) called a lower-order semantic analysis which are by nature inductive and exploratory. As such, issues and frames raised by organisations when discussing ‘data sovereignty’ and ‘planetary sustainability’ which were often hidden can be revealed and mapped. To be able to identify clusters of issues around planetary sustainability, several prominent terms discussed were selected in advance through selection of relevant key terms such as ‘climate’, ‘energy’, ‘health’, ‘sustainability’, ‘environment’, and ‘resilience’ making up the general concept of ‘planetary sustainability’. This deductive approach allows to capture more specific terms set up in advance to define which issues identical to the agenda of ‘planetary sustainability’.

3 Findings

3.1 Data sovereignty in the North and South

Discourse on data sovereignty brings to light a range of issues as different actors engage with it promote different concerns. Figure 2 presents a visual representation with a wordcloud comparison on data sovereignty issues based on key terms used by organisations in different regions. Using terms extracted from the websites, size of the terms is proportional to the frequency of occurrence indicating the most prominent keyword in each region. It shows that the South organisations point to more diverse issues than the North. However, it does not mean that words that appear in the South were totally absent in the North. For example, the terms 'digital' and 'protection' which belong to the South means that they were being used more frequently by the South organisations compared with their North counterpart. Clustering of the terms based on region allows an interpretation on relative concerns between different actors when discussing the same concept, namely 'data sovereignty'. It appears that the North organisations in this study pay more attention to issues such as 'indigeneous', 'rights', and 'cloud'. Meanwhile, the South more focus on issues such as 'digital', 'protection' and 'trade'.

Fig. 2. Comparison of terms mentioned based on regions

Although it explicitly indicates the presence of words appeared in the same issue cluster such as 'rights' and 'privacy' form the North and 'protection' and 'development' from the South, there is no further information that can be gained from visual representation of the wordcloud comparison. What can be further described is that these

divergent issues of concern underlying complication of relation between the North and South when discussing 'data sovereignty' at least among civil society organisations. As a result, data sovereignty becomes a field of issue contestation among different organisations shaping different priorities across regions. By inductively mapping the terms as shown above, it is clear that issues related to the problem of 'planetary sustainability' (e.g climate, health, environment, etc) were not prominent in the entire corpus. Indeed, planetary sustainability was discussed or at least mentioned as these text were based on the corpus on the broad topic of data sovereignty. To make issues around 'planetary sustainability' visible, manually selected the terms related to 'planetary sustainability' were summarised to see their position in comparison between the North and South Corpus.

3.2 Planetary sustainability in the North and South

Table 2. Table captions should be placed above the tables.

	North		South	
Issue	% Term/Level	% Level/Term	% Term/Level	% Level/Term
climate	0.04	30.39	0.36	69.61
energy	0.02	68.42	0.03	31.58
health	0.11	79.13	0.12	20.87
sustainability	0.07	60.67	0.18	39.33
environment	0.12	88.50	0.07	11.50
resilience	0.03	88.46	0.02	11.54

Presenting the frequency terms on specific theme using region as a variable provides a comparison between the North and South in the topic around 'planetary sustainability' in data sovereignty corpus. The table above shows that the word 'climate' is more prominent in the South. It represents 0.36% of all occurrences of words mentioned by the South organisation websites. More than half (69.61%) of occurrences of the term 'climate' are cited by the South organisations. While the rest of 'planetary sustainability' terms appeared more often in the North, some terms are proportionally used much more in the South such as 'health' and 'sustainability'. Meanwhile terms 'energy', 'environment' and 'resilience' are used much less by the South organisations.

There have been some studies that diagnose the link between data-driven development and a current threat on the environment (e.g. Crawford 2021; Becker 2023). The specific term on 'climate' which was more frequently mentioned by the South recalls the existing climate crisis which disproportionately affects the Global South. Regardless the assumption that the South region is more dependent on natural resource, issues on climate crisis exacerbates already existing vulnerabilities related to energy, health, sustainability, and environment experienced especially by more people in the South. However, there is also an optimistic narrative around the belief that the wealth of data and information in

contemporary societies is capable of solving global issues including climate change and public health issues.

3.3 Mapping of issues around 'planetary sustainability'

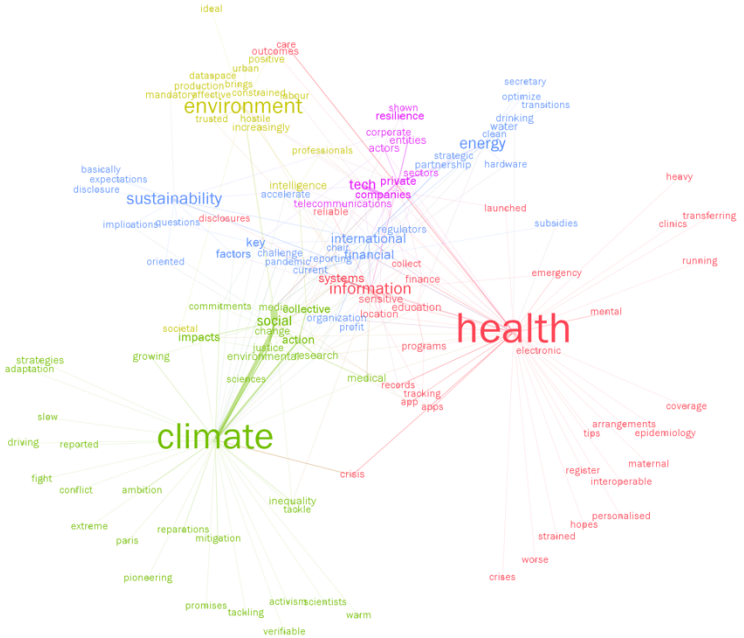

Fig. 3. Semantic mapping of the key terms on 'planetary sustainability'

To explore interrelation between key terms surrounding the discourses around 'planetary sustainability', semantic mapping is performed by selecting specific words to view how they are clustered and interrelated. The semantic graph above provides community clusters on each term differentiated by colour. The term size is proportional to centrality of position as intermediary or connector between words or in the language of social network analysis known as 'betweeness centrality' (Freeman 1978). From the graph, it is possible to identify the use of the term in relation to other terms shaping cluster of issues on planetary sustainability. Issues such as 'climate' and 'health' occupy core position in the network in which both issue-clausters are undirectly linked.

There are five algorithmically detected communities of words in planetary sustainability issues where 'climate' and 'health' are situated in the core of the network component. It is unsurprising to see that those clusters are centered around 'climate', 'health', 'environment', resilience', 'sustainability' and 'energy' issues where

sustainability and energy fall into the same cluster indicating the seemingly inseperable concerns between the two. Material aspects of data sovereignty such as building sustainable data center infrastructures require a huge amount of energy which are increasingly demanded to be clean and renewable. It is interesting to see a linkage between 'climate' and 'health' issue clusters which are mediated to the key term 'crisis' indicating a shared concern when planetary sustainability discussed in the context of health and climate.

When the global economy is increasingly digital-based and data-driven, it is crucial to limit the scope of digital platforms that base their sustainability mainly on users' data extraction. Term 'sustainability' which directly connect with 'oriented' and 'international', for instance, form a message that it is important to reorient sustainability model that are not defined by individual platform companies but having more global orientation where the users, public or commons can take a lead to shape and exercise the model. In data governance debates, the notions of 'data sovereignty', 'data common', 'data trust' and those related concepts go hand in hand with this spirit of reorienting sustainability model (Renieris 2023; Becker 2023).

4 Conclusion

Planetary sustainability becomes one of the increasing concerns for organisations and civil society groups working on data sovereignty issues. The fact that data ownership and control are intrinsically linked to how environmental resources are managed, data sovereignty for planetary sustainability are promoted as a shared agenda by civil society organisations. Organisations focused on data sovereignty advocate for more just, responsible, and equitable access to data that can drive sustainable practices (Hummel et al. 2018; Hummel, Braun, and Dabrock 2021; Polatin-Reuben and Wright 2014). The existing North-South divide potentially exacerbates the environmental impacts of datafication, in particular for the Global South given the North often benefits from data extraction as a result of their domination in data infrastructures, access and control. The move towards planetary sustainability goes beyond 'greening' data governance, but also creating a condition where data subjects particularly in the South have the power to use their data to fight against the existing practice of 'data colonialism'. This is a crucial aspect of how civil society groups envision a sustainable future where data sovereignty as a framework of data governance serves the need of both humans and the planet.

In the Global South, organisations focusing on data sovereignty face a broader array of issues compared to their North counterpart. While the North organisations often promote issues around individual rights including in relation to privacy and indigenousity, Southern organisations, on the other hand, address more diverse issues, ranging from digital protection and trade to development, gender, and cybersecurity. Climate crisis remains a pressing issue globally, but its impacts arguably affect more acutely and immediately in the South. For many organisations in the South, more than protecting

marginalised groups and individual users from the risk of excessive datafication by big techs, data sovereignty is also conceived of ensuring that existing global data governance can be reformed and reoriented to address existential threats to the planet and the living environment. Based on the semantic mapping, organisations in both regions express shared concerns which illustrate the current and future crisis contextualised in multiple areas primarily in health and climate change.

Reorienting data-driven development toward more just sustainability is crucial in achieving equitable outcomes for all regions. This involves shifting the focus from profit-driven data extraction and monopolisation that benefit few large entities in the Global North and increasingly China, toward the creation of systems that prioritise social equity, environmental justice, and fairness. To do this, it is important in the first place to put limit on the role of profit-oriented machines of data extraction such as big tech companies. As argued by (Couldry and Mejias 2019), the problem today is not only in data neither a particular platform exploiting data, but interlocking combination ranging from infrastructure, order, system, knowledge to models of governance which use data as the new means to reshape the world according to capitalist ideals. Data sovereignty as concerned by civil society organisations address the problem beyond data where the notions of 'planetary sustainability' are put in place in parallel with reconsideration of sovereignty of the data subjects in order to realise globally just societies in the age of data-driven economies.

References

1. Arora, Payal. 2019. *The Next Billion Users: Digital Life Beyond the West*. Cambridge, Massachusetts: Harvard University Press.
2. Beaulieu, Anne, and Sabina Leonelli. 2022. *Data and society: a critical introduction*. London: Sage.
3. Becker, Christoph. 2023. *Insolvent: How to Reorient Computing for Just Sustainability*. Cambridge, Massachusetts: The MIT Press.
4. Couldry, Nick, and Ulises Ali Mejias. 2019. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Culture and Economic Life. Stanford, California: Stanford University Press.
5. Crawford, Kate. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.
6. Freeman, Linton C. 1978. "Centrality in Social Networks Conceptual Clarification." *Social Networks* 1 (3): 215–39. https://doi.org/10.1016/0378-8733(78)90021-7.
7. Hummel, Patrik, Matthias Braun, Steffen Augsberg, and Peter Dabrock. 2018. "Sovereignty and Data Sharing." *ITU Journal: ICT Discoveries, Special Issue*, no. 2: 1–10. https://www.itu.int/en/journal/002/Pages/default.aspx.
8. Hummel, Patrik, Matthias Braun, and Peter Dabrock. 2021. "Data Sovereignty: A Review." *Big Data & Society* 8 (1). https://doi.org/https://doi.org/10.1177/2053951720982012.
9. IGF. 2022. "Data Sovereignty in Asia Pacific and Beyond." https://www.intgovforum.org/en/content/igf-2022-ws-363-data-sovereignty-in-asia-pacific-and-beyond.
10. Kwet, Michael. 2019. "Digital Colonialism: US Empire and the New Imperialism in the Global South." *Race & Class* 60 (4): 3–26. https://doi.org/10.1177/0306396818823172.

11. Markelius, Alva, Connor Wright, Joahna Kuiper, Natalie Delille, and Yu-Ting Kuo. 2024. "The Mechanisms of AI Hype and Its Planetary and Social Costs." *AI and Ethics* 4 (3): 727–42. https://doi.org/10.1007/s43681-024-00461-2.
12. Polatin-Reuben, Dana, and Joss Wright. 2014. "An Internet with BRICS Characteristics: Data Sovereignty and the Balkanisation of the Internet." In. FOCI 2014.
13. Renieris, Elizabeth M. 2023. *Beyond Data: Reclaiming Human Rights at the Dawn of the Metaverse*. Cambridge, Massachusetts: The MIT Press.
14. Robinson, Brett H. 2009. "E-Waste: An Assessment of Global Production and Environmental Impacts." *Science of The Total Environment* 408 (2): 183–91. https://doi.org/10.1016/j.scitotenv.2009.09.044.
15. Thatcher, Jim, David O'Sullivan, and Dillon Mahmoudi. 2016. "Data Colonialism Through Accumulation by Dispossession: New Metaphors for Daily Data." *Environment and Planning D: Society and Space* 34 (6): 990–1006. https://doi.org/10.1177/0263775816633195.
16. Yang, Sijia, and Sandra González-Bailón. 2016. "Semantic Networks and Applications in Public Opinion Research." In, edited by Jennifer Nicoll Victor, Alexander H. Montgomery, and Mark Lubell. Vol. 1. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190228217.013.14.