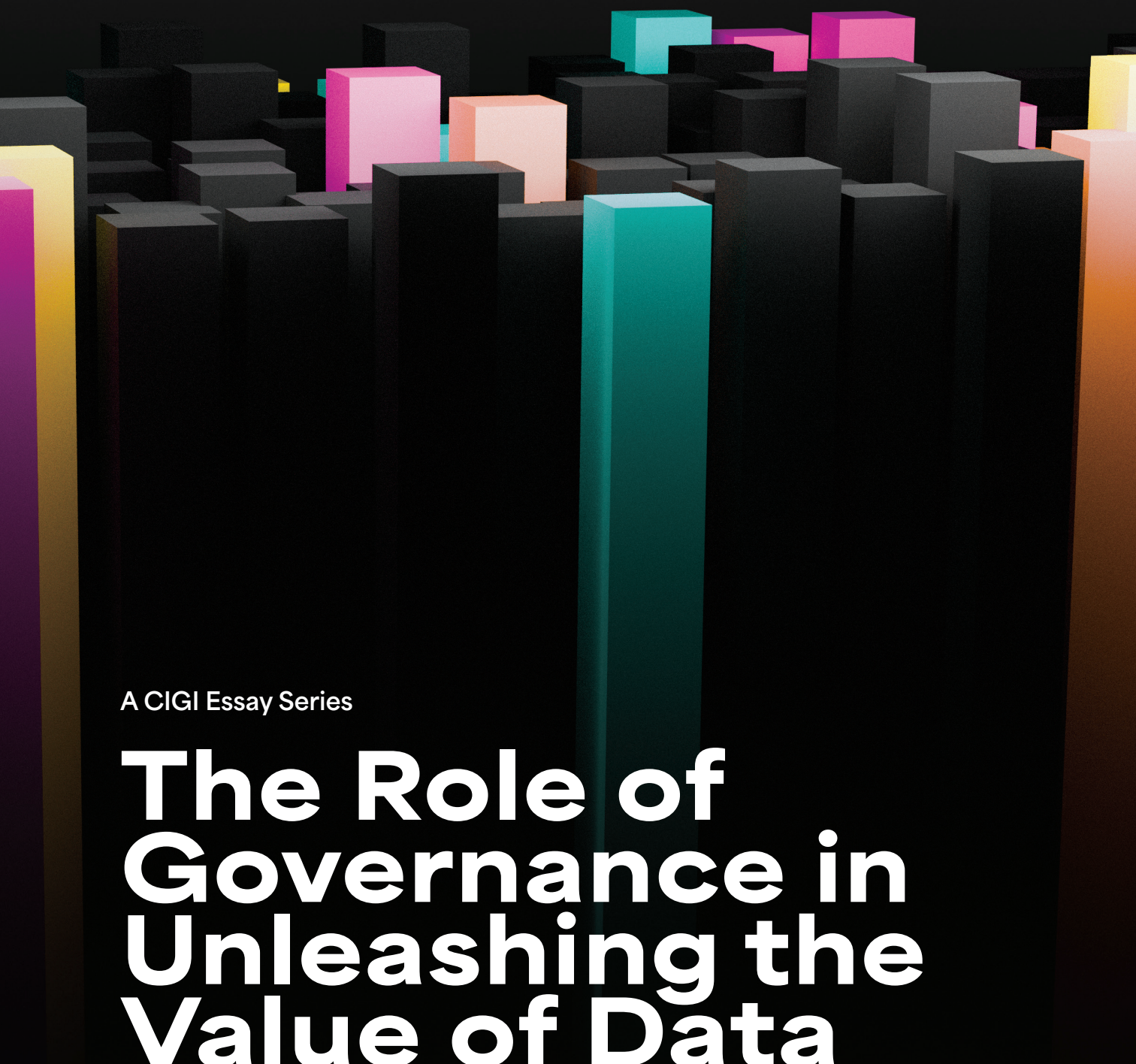
Centre for International Governance Innovation



A CIGI Essay Series

The Role of Governance in Unleashing the Value of Data

Credits

SERIES EDITORS
Paul Samson
Robert Fay

PROGRAM MANAGER
Jenny Thiel

GRAPHIC DESIGNER
Sami Chouhdary

SENIOR PUBLICATIONS EDITOR
Jennifer Goyder

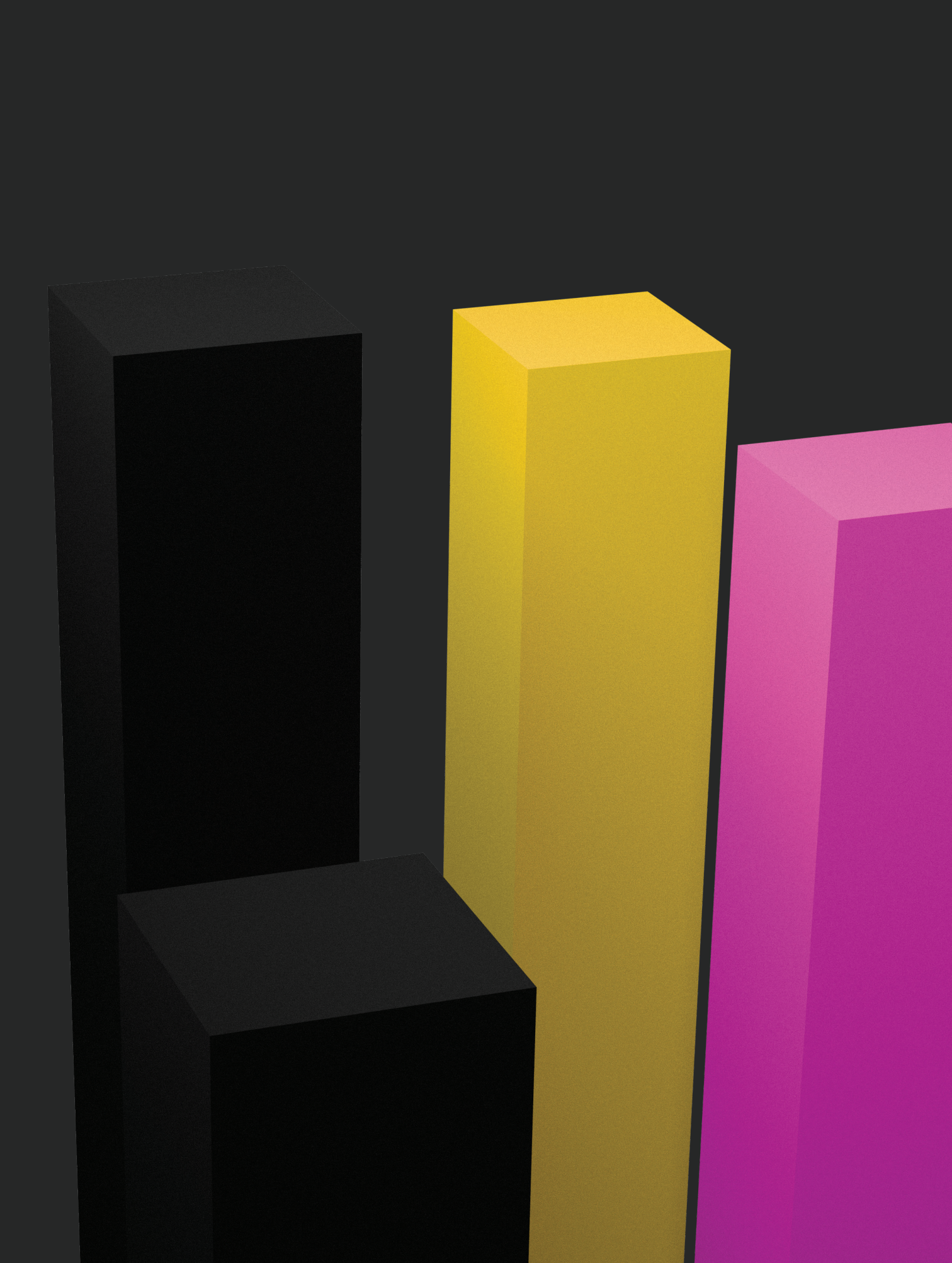
PUBLICATIONS EDITORS
Susan Bubak
Christine Robertson

Digital version available at:
cigionline.org/value-of-data

This essay series was made possible thanks to generous support from Omidyar Network.

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

Table of Contents



Introduction

The Role of Governance in Unleashing the Value of Data

Robert Fay

In 2018, the Centre for International Governance Innovation (CIGI) released an essay series titled Data Governance in the Digital Age. It was a far-reaching compendium covering topics such as the rationale of a data strategy, how to balance privacy and commercial values, and international policy considerations. And it anticipated many of the issues that have emerged, such as surveillance capitalism. One area that was not covered in depth, however, was data valuation.

Against that background, in November 2023, CIGI co-hosted an international conference, in conjunction with the International Association for Research in Income and Wealth, to advance discussion on the valuation of data as an asset.[1] Despite some meaningful progress by national account statisticians to value data as an asset, this value is still not included in national balance sheets, nor is it included in corporate balance sheets. No ideal or agreed-upon methodology has yet emerged to measure data's value, largely because its value depends on its usefulness in a particular context (Coyle and Manley 2022; Mitchell, Ker and Lesher 2021). That context is framed by governance — the rules and regulations that determine how data, especially personal data, can or should be used — and includes standards (such as those set by accounting and other regulatory bodies), intellectual property rights, trade treaties, competition, privacy and other frameworks that will vary across jurisdictions and even within them. Achieving a coherent framework that encompasses these areas — and others — is a substantial challenge for any country and obtaining coherence globally is even more difficult.

CIGI therefore commissioned some global thought leaders to share their ideas on how to advance data governance to unleash the value of data. This essay series explores four themes: the current state of global data governance; different perspectives on notions of value; governance frameworks to unleash the value of data; and mechanisms for governance cooperation.

The Current State of Global Data Governance

Establishing country-level data governance frameworks is an ongoing and complex undertaking. Countries have very different capacities to develop these frameworks, implement them and then enforce them. They face a variety of challenges, including the interconnected nature of this governance; the digital divide in creating and enforcing frameworks; an ever-changing legislative landscape that includes laws, regulations and standards at both the national and international levels; and constantly evolving data-intensive technologies. Yet these frameworks are essential for the trustworthy sharing of data that can drive economic activity.

Drawing upon findings from the Global Data Barometer, Silvana Fumega examines the progress that has been made by countries globally in establishing legal frameworks and data policies as well as documenting the many challenges that exist while offering some potential solutions. Susan Ariel Aaronson looks specifically at the interplay of data and artificial intelligence (AI) governance in country policies, the apparent disconnect between data governance and AI governance, and the risks that this gap presents nationally and internationally. Against this background, Lorrayne Porciuncula discusses ways to put data governance at the forefront of policy discussions and redefine value creation in ways that prioritize societal well-being and sustainability over the short-term profits of firms.

Different Perspectives on Notions of Value

The focus of data governance has typically been on the individual and their personal data, but data may also be inferred from groups of individuals and their activities. Meanwhile, individuals and groups may have different perspectives on how their data should be governed and the value attached to it. For example, aggregated data may bring value to society that may not be part of an individual's perspective. It can also raise a variety of risks. This situation is evolving over time as new technologies emerge and harness different types of data. How to represent the diverse perspectives and interests and the weights to place upon them is an ongoing challenge. To address that challenge, various data stewardship models have been developed, each with strengths and weaknesses. Nevertheless, these models can allow individuals and groups to control how their data will be used, how to derive value from it and how to share that value.

Teresa Scassa discusses the evolution of individual and collective privacy and emerging data rights that give individuals — and perhaps communities — more control over both the personal and non-personal data that they generate. Sean Martin McDonald points out that the value of data reflects whose interests are being represented and the integrity of the supply chain by which the data is produced. He discusses how fiduciary models provide one mechanism for rights holders to participate in the governance of their data and to protect the supply chain that creates the data. Kean Birch argues that such governance models, including not only data trusts but also data commons and national statistical agencies, could be used as models to create a data wealth fund along the lines of those based on commodities, although doing so in practice is not straightforward.

Governance Frameworks to Unleash the Value of Data

Data forms a value chain: data, and especially big data, can be used with AI technologies to create powerful analytics that can vastly improve policy and business decision making. Digital technology firms that are first movers in big data have tremendous advantages through economies of scale and scope, network effects and information asymmetries; these characteristics have not only driven up their value, but also give these firms tremendous market power, sometimes in multiple markets. This situation has led to different proposals on how to address this power since it can stifle innovation, be privacy invasive and impact how value is created and distributed.

Keldon Bester discusses the renewed focus on competition policy and how essential its role is in unleashing the value of data, and how it can capture non-monetary issues such as privacy. The monopolization of key elements of infrastructure in the "tech stack" by big firms has also led to an exploration of the role that digital public infrastructure (DPI) might play. Soujanya Sridharan, Vinay Narayan and Jack Hardinges describe how DPI can be viewed as an alternative to private monopolies and explain that how and why DPI will be used — its orientation — is critical to ensure proper governance. Market-based mechanisms are an obvious way to unleash the value of data. Alex He and Rebecca Arcesati document the experiences of China with local data exchanges that allow the public trading of various types of data, including personal data, and note that while trading exchanges are usually seen as a primary means of value discovery, even such exchanges require solid data governance to run effectively.

Mechanisms for Governance Cooperation

Multi-stakeholder input is an oft-mentioned approach to create inclusive data governance, although how to achieve it in practice is not straightforward, especially since the term multi-stakeholder can take on a variety of different meanings whereas stakeholders likely have different objectives and capacities to participate. CIGI's Global Platform Governance Network (GPGN) was created to bring different perspectives to platform governance issues in a multi-stakeholder environment, recognizing that this governance needs to be multidisciplinary, representative and transnational. More formally, given that data flows globally, trade agreements are already being used in various ways to deal with elements of data governance, but these agreements are not comprehensive, and may reflect power imbalances among signatories as well as impact those who are not party to the agreement(s).

Jeni Tennison makes the case for the necessity to have multi-stakeholder representation in data governance and discusses how the inclusion of civil society is not only democratic, but also creates a shared understanding that breaks down barriers, generates trust, boosts literacy and encourages adoption of digital technologies. Chris Beall documents CIGI's experience with the GPGN and offers concrete recommendations on how this type of network can be applied to other areas, including data governance. Finally, Patrick Leblond explores the growing "digital noodle bowl" of regulations in trade agreements related to cross-border data flows and suggests some ways forward to create effective global governance.

In summary, these essays reveal the complicated governance background that lies behind more technical discussions on how to measure the value of data. Countries are making substantive progress on data governance frameworks, but there are still gaps and silos at the national and international levels that need to be addressed so that decisions made on data governance are representative and inclusive and reflect the values of various stakeholders. In doing so, data governance can create a trusted environment to share data that, in turn, can create value for the individual and for society.

Note

1 Funding from Omidyar Network and the Balsillie Family Foundation is gratefully acknowledged.

Works Cited

Coyle, Diane and Annabel Manley. 2022. "What is the Value of Data? A review of empirical methods." Policy Brief. July. Bennett Institute for Public Policy, University of Cambridge. www.bennettinstitute.cam.ac.uk/publications/value-of-data/.

Mitchell, John, Daniel Ker and Molly Lesher. 2021. "Measuring the economic value of data." OECD Going Digital Toolkit Notes, No. 20. Paris, France: OECD. www.oecd-ilibrary.org/science-and-technology/measuring-the-economic-value-of-data_f46b3691-en.

About the Author

Robert (Bob) Fay is a highly accomplished and respected leader in the field of digital economy research. With more than 30 years of experience working in the public and private sectors, he has developed expertise in economics, policy analysis and strategic planning.

Bob served as managing director of digital economy at CIGI, where he led a network of researchers focused on the intersection of technology, trade, innovation and governance. In this position, he has played a key role in shaping the discourse around the digital economy and has contributed to numerous policy debates and research initiatives on topics such as data governance, digital innovation and the future of work.

The Global Landscape of Data Governance

Silvana Fumega

In the past few decades, data has emerged as an invaluable asset, offering the promise of shaping our collective future through informed decision making and transformative solutions. Its potential spans crucial domains such as climate action, health care and economic development, where data-driven approaches hold the key to unlocking progress for the greater good. However, the use of data for public good[1] is not without its hurdles. While it fuels innovation and progress, it also presents challenges, such as widening disparities and privacy concerns. Robust legal frameworks and ethical guidelines governing data collection, storage and accessibility need to be established in order to achieve a balance between advancing shared interests and safeguarding individual rights.

Governments play a pivotal role in shaping this framework, enacting clear regulations, adopting international standards and establishing oversight mechanisms. Effective data governance, particularly concerning government-held data, ensures accessibility while safeguarding privacy. Moreover, despite the increasing attention to artificial intelligence (AI), the fundamental role of data governance in this landscape is often overlooked, as stated by Stefaan G. Verhulst and Friederike Schüür (2023). AI governance relies inherently on the principles and practices of data governance, and neglecting this synergy leads to fragmented approaches and missed opportunities for collaboration.

The March 2024 UN General Assembly draft resolution A/78/L.49[2] emphasizes the critical role of data and data governance in advancing "safe, secure and trustworthy" AI systems (UN General Assembly 2024), which are fundamental for driving sustainable development globally. Acknowledging data as the cornerstone for both developing and operating AI systems, the resolution[3] urges the adoption of fair, inclusive, responsible and effective data governance practices. It calls for enhancing data generation, accessibility and infrastructure while maximizing the utilization of digital public goods. Furthermore, member states are urged to exchange best practices on data governance and to foster robust international cooperation, collaboration and assistance efforts.

Following all those ideas, failure in data governance may result in AI systems falling short of legal and regulatory compliance, posing risks to data integrity and privacy. This underscores the importance of establishing clear frameworks, particularly for public data. According to the *2021 World Development Report: Data for Better Lives*, unlocking the full potential of data for development requires the establishment of a comprehensive

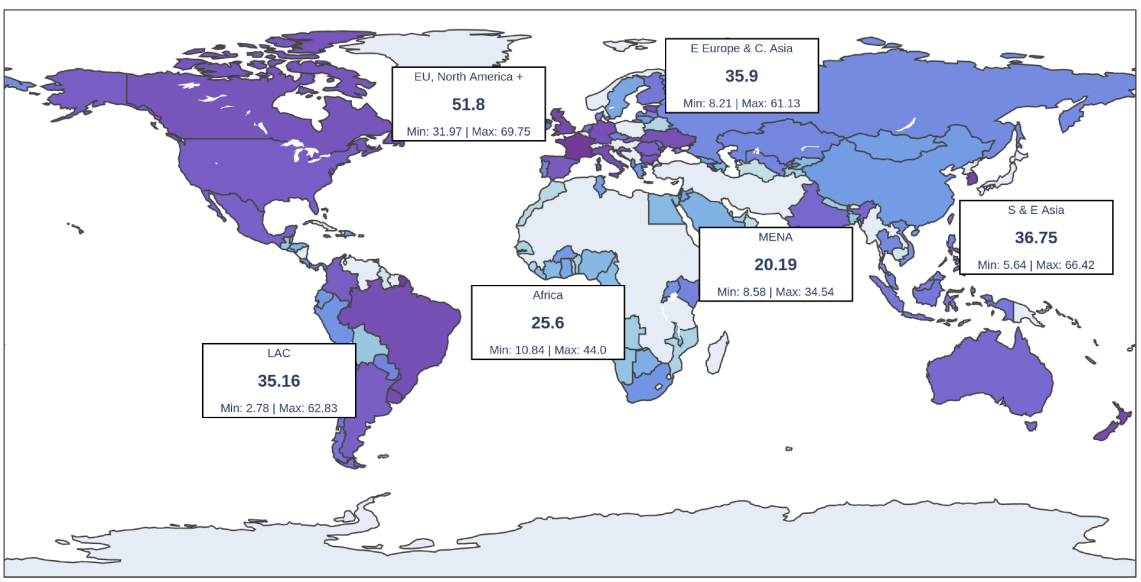
national data system supported by data governance frameworks, and addressing factors such as data quality, technical standards and transparent processes (World Bank 2021).

This essay explores the landscape of public data governance, drawing insights from the first edition of the Global Data Barometer (GDB).[4] It delves into the complexities, challenges and opportunities inherent in this domain, emphasizing the potential of data governance as a catalyst for positive societal change and inclusive development.

A Global Overview of Data Governance

Effective data governance plays a paramount role in ensuring the reliability and integrity of data throughout its life cycle. At its core, data governance entails maintaining high data quality and implementing robust data controls, which are crucial for training AI models and making informed decisions. In essence, data governance emerges not only as a mechanism for responsible data management but also as a tool for promoting positive societal change and inclusive development. Ultimately, navigating the data governance landscape requires an approach that integrates legal frameworks, technological standards and collaborative efforts to ensure responsible and effective data management for the benefit of society and the ethical advancement of AI technologies (see Figure 1). This encompasses various key focus areas such as consistency, data integrity, security and standards compliance.

Figure 1: Governance Pillar Regional Scores



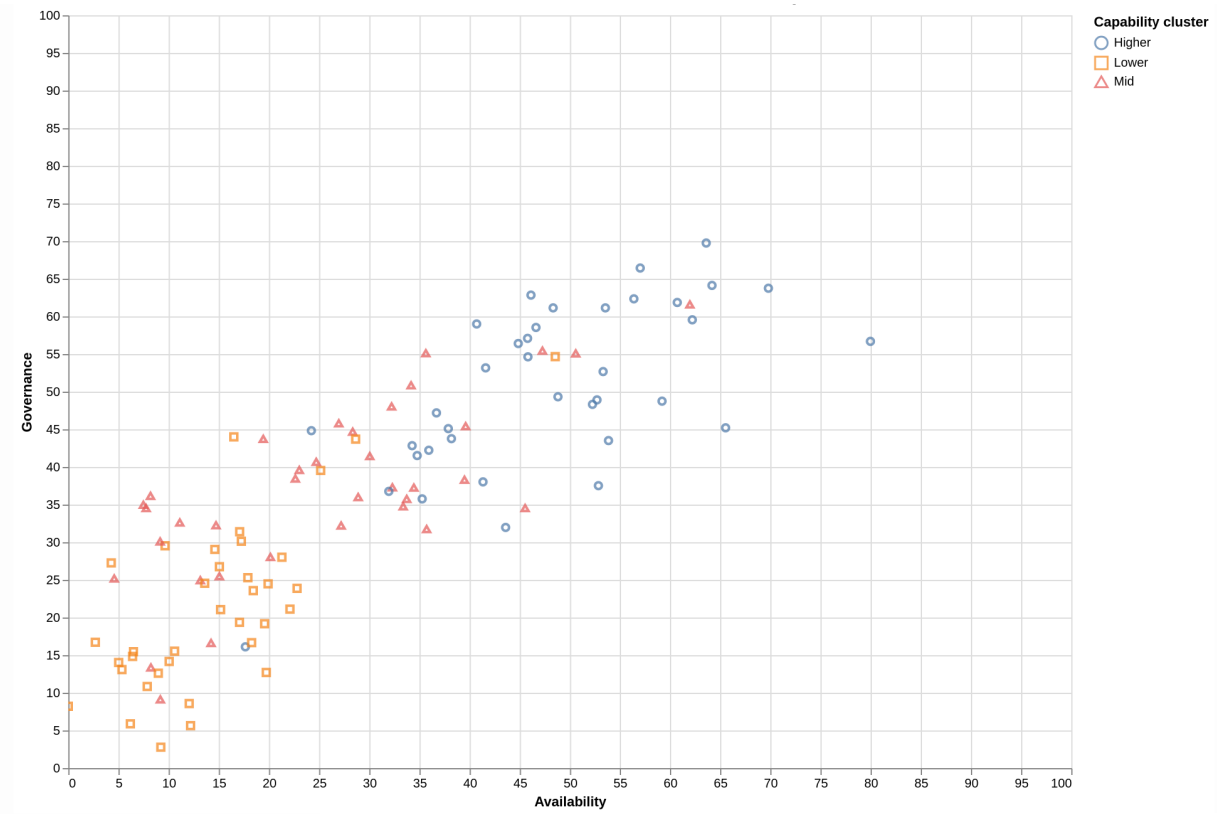
Governance pillar regional scores: The EU, North America+ regional grouping achieves the highest scores on the governance pillar. Countries in the Middle East and North Africa have the lowest average score.

Source: GDB (2022, 23).
Note: LAC = Latin America and the Caribbean; MENA = Middle East and North Africa.

The significance of data governance is underscored by insights from the GDB. For example, the positive correlation between governance and public availability[5] scores in the GDB (2022) highlights the essential role of robust data governance frameworks in making data readily available and fostering progress (see Figure 2). Thus, the correlation

between laws mandating data publication and actual data availability is evident, yet disparities persist in the implementation of these requirements across various sectors, highlighting the existence of "implementation gaps" that warrant attention and remediation.

Figure 2: Correlation of Governance and Availability Pillars



Correlation of governance and availability pillars: There is a positive relationship between governance and availability scores.

Source: GDB (2022, 30).

The next section explores various challenges within the realm of data governance, including privacy concerns, open data policies and the dynamics of data sharing.

Navigating the Complexities: Challenges in Data Governance

The global landscape of data governance presents a nuanced picture, marked by both progress and challenges. While advancements have been made in establishing legal frameworks and data policies, significant hurdles remain.

Uneven Progress

According to the main findings of the first edition of the GDB, countries around the globe are increasingly recognizing the importance of safeguarding personal data through regulations. However, in some cases, these frameworks provide limited protections, primarily focusing on specific sectors rather than offering comprehensive coverage. There is therefore diversity in the implementation of data governance frameworks across countries, with many lacking comprehensive regulations. This creates a patchwork of data protection, privacy and sharing frameworks, which can hinder collaboration and potentially exacerbate existing inequalities.

In that sense, although 98 out of 109 surveyed countries by the GDB present some sort of data protection framework, only 46 percent of them boast robust data protection frameworks, leaving a significant portion of the global population vulnerable to data misuse and privacy violations. According to Keziah Munyao of the Local Development Research Institute, who provided an overview for Sub-Saharan Africa in the *Global Report*, "The fieldwork also identified gaps with respect to data protection or privacy standards in a number of countries, even where efforts are underway to promote wider data usage and openness. The absence of strong legal frameworks alongside new technological advancements seems to be a developing concern, particularly in countries where no frameworks exist to oversee the use of emerging technologies such as artificial intelligence (AI)" (GDB 2022, 74).

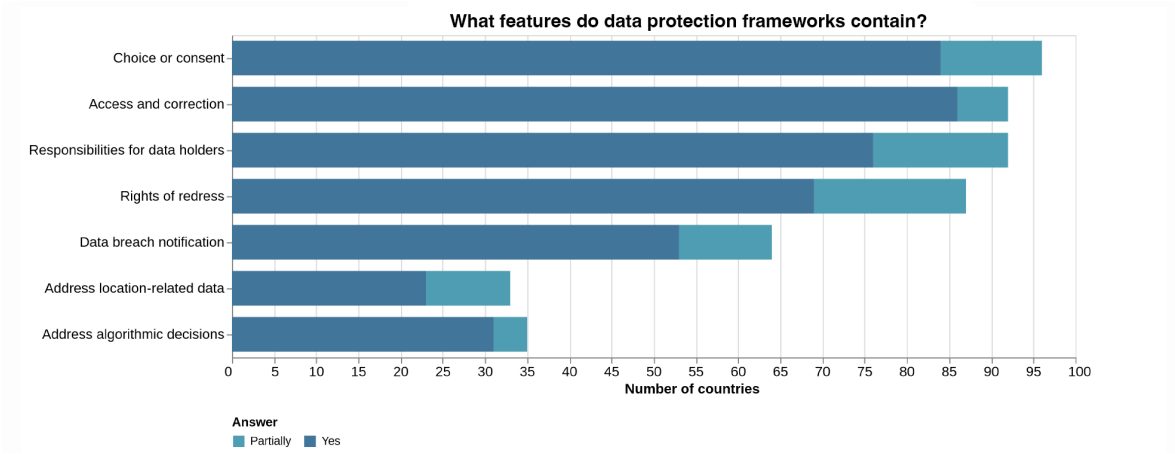
Balancing Privacy and Innovation

Balancing individual privacy rights with the vast potential of data sharing for innovation and the greater public good poses a multi-faceted challenge. Persistent concerns surrounding data collection, storage and utilization continue to evoke ethical and legal dilemmas, complicating the quest for a sustainable equilibrium. Following Uma Kalkar and Natalia González Alarcon (2023), the strategy that finds a balance between regulating and encouraging innovation considers both data protection and the advancement of data reuse.

The journey toward establishing comprehensive data governance requires continual refinement and adaptability. Insights from the *Global Report* (GDB 2022) reveal a significant gap in data protection regulations across countries. Notably, while certain aspects are widely covered, 45.9 percent of nations lack robust provisions for data breach notifications, and 29.6 percent offer limited redress for harm caused by data misuse (ibid., 24). Moreover, only 23.5 percent of existing frameworks effectively address location data concerns, with a slightly higher percentage (31.6 percent) tackling issues related to algorithmic decision making (ibid.) (see Figure 3). Nevertheless, recent advancements in global standards[6] underscore a commitment to tackling emerging challenges. Efforts are increasingly focused on improving breach notification protocols and recognizing the sensitivity of location data. In alignment with UN resolution A/78/L.49, safeguarding privacy and personal data integrity during AI system testing and evaluation is paramount.

Nevertheless, recent advancements in global standards underscore a commitment to tackling emerging challenges. Efforts are increasingly focused on improving breach notification protocols and recognizing the sensitivity of location data. In alignment with UN resolution A/78/L.49, safeguarding privacy and personal data integrity during AI system testing and evaluation is paramount.

Figure 3: Features in Data Protection Frameworks



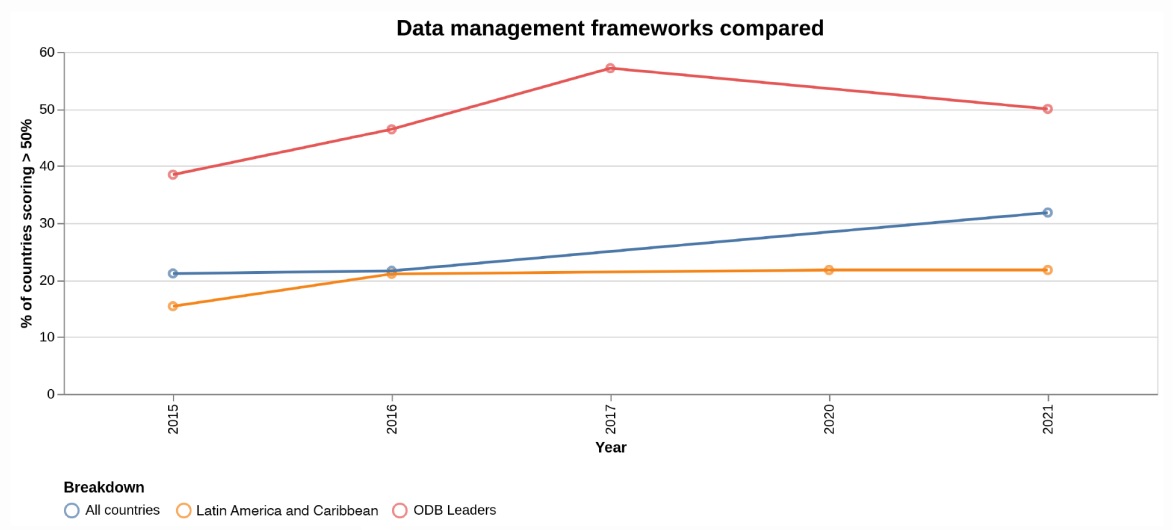
Source: GDB (2022, 24).

Untapped Potential of Open Data

Open data, when readily accessible and usable, can fuel innovation, empower individuals and foster transparency. In this context, the proliferation of open data policies emerges as a beacon of progress. The GDB *Global Report* identified 74 countries with open data policies, 30 of which have legally enforceable regulations (GDB 2022, 25). However, this potential remains largely untapped due to issues of standardization and interoperability (only 47.3 percent of policies address common data standards) (ibid.). Without consistent standards and seamless interoperability, open data initiatives struggle to achieve their full impact.

Ensuring both the quality and protection of data, while also harnessing its potential for public good initiatives, demands careful consideration in terms of data management. In this regard, juxtaposing data from the GDB, the 2015 Open Data Barometer (ODB), the 2018 ODB Leaders Edition and the 2020 Latin America and the Caribbean edition provide valuable insights. These comparisons hint at a modest global trend toward more robust data management, notably emanating from countries beyond the ODB Leaders Edition (see Figure 4).

Figure 4: Comparison of Data Management Frameworks



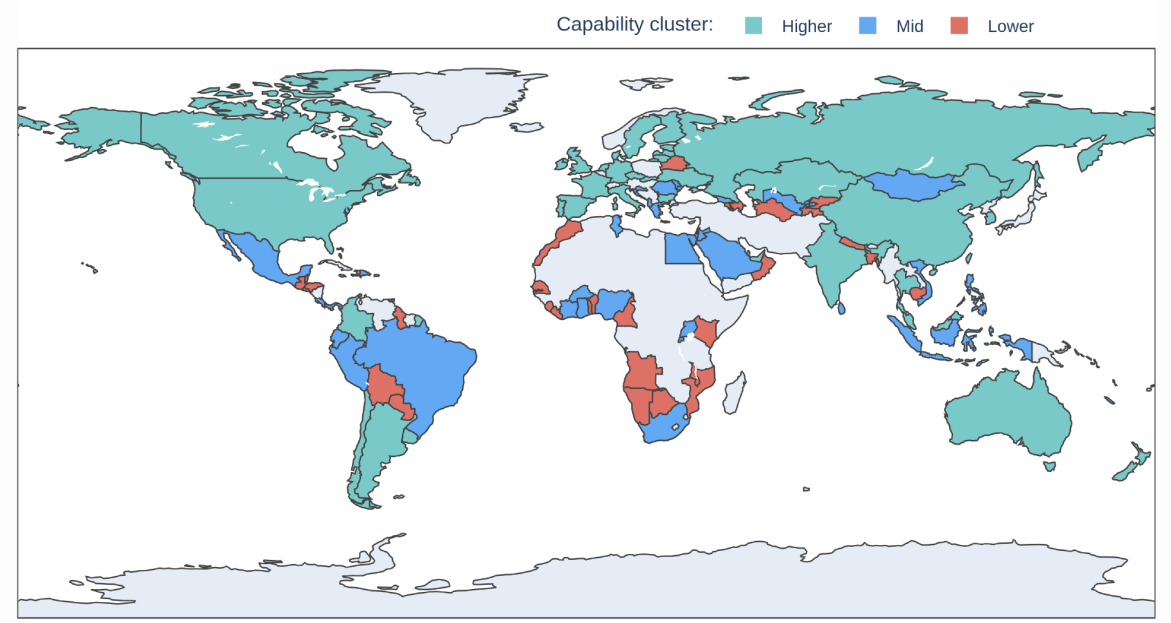
Source: GDB (2022, 25).

Capacity Constraints

National governments play a pivotal role in data governance and stewardship, with various responsibilities including shaping data strategies, establishing and funding crucial governance bodies such as data protection authorities, offering digital services for data collection and access, and defining and adopting precise data standards. Secondary indicators within the GDB highlight the distinct hurdles encountered by countries with lower capabilities, especially concerning the presence of governing institutions responsible for data management, governance and protection. Additionally, these countries often lack essential infrastructure such as government cloud platforms and comprehensive strategies, including technology and interoperability strategies, further complicating their data management efforts.

Thus, many countries struggle with capacity constraints (see Figure 5), since they lack the resources and expertise necessary to effectively implement their data governance frameworks. This lack of skilled personnel and clear accountability mechanisms hinders their ability to establish robust data protection systems, manage data sharing effectively and leverage open data for public good initiatives. Unless these capacity constraints are addressed, the digital divide will continue to widen, and the benefits of data governance will remain out of reach for many.

Figure 5: Capability Cluster



Source: GDB (2022, 34).

Emerging Technologies

The rapid pace of technological advancements presents a unique challenge for data governance frameworks. New technological developments such as the advances in AI constantly raise new risks and opportunities, demanding continuous adaptation and revision of existing regulations. Data governance applies to any form of technology that collects, uses or processes data, offering a holistic and adaptable framework that can evolve with changing technologies (Verhulst and Schüür 2023). Failing to keep pace with this evolving landscape can leave data vulnerable to misuse and create unforeseen ethical dilemmas, potentially hindering responsible innovation.

Data Sharing: Trustworthy Exchange in Limbo

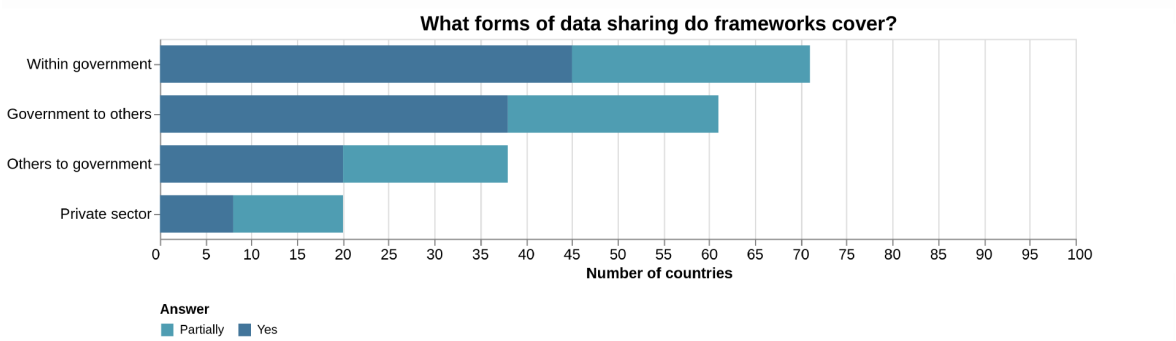
Access to government data has the potential to spur AI development.[7] To achieve this, a multi-faceted approach that enhances data sharing is necessary. However, data-sharing, crucial for collaborative problem solving and innovation, remains hindered by underdeveloped legal frameworks.

There is still a notable absence of robust frameworks governing data-sharing practices in many countries, which could limit the accessibility of diverse data sets essential for training AI models. While 30 countries have enacted legally binding open data policies, challenges persist in ensuring standardized and interoperable data publication, which is essential for facilitating data sharing and collaboration.

This lack of robust legal structures creates an environment of uncertainty and mistrust, discouraging the exchange of information for public good initiatives. Particularly concerning is its impact on areas such as public health and environmental sustainability, where data collaboration holds immense potential for positive impact.

Delving into the various types of data sharing governed by existing frameworks, according to GDB data, the majority (92.6 percent of the 68 frameworks identified[8]) pertains to data sharing within governmental entities (GDB 2022, 27). Almost 80 percent (79.4 percent) address the sharing of data between the government and other sectors, while 51.5 percent outline protocols for data sharing from other sectors to the government (ibid.) (see Figure 6). A mere 16.2 percent explicitly cover the utilization of data in AI applications and just 26.5 percent focus on data sharing within the private sector (ibid.).

Figure 6: Forms of Data Covered by Frameworks



Source: GDB (2022, 27).

Looking ahead, the next decade is likely to see increased voluntary and mandated data-sharing arrangements between businesses in industry sectors, between business and government, and in support of data collaborative arrangements oriented toward addressing humanitarian and development challenges. Without clear frameworks that facilitate and govern such arrangements, there are risks that positive uses of data will be missed, and that abuses of data will proceed unchecked.

Digital Divide

While there have been notable advancements in global data governance, persistent challenges remain, particularly for lower-income countries. These nations often lack the necessary resources and expertise, exacerbating the digital gap. Balancing privacy with innovation, overcoming capacity constraints and staying abreast of emerging technologies are pivotal for fostering responsible and impactful data governance going forward.

Efforts to address these challenges are imperative to unlock the full potential of data for a more equitable and prosperous society. UN resolution A/78/L.49 recognizes the diverse levels of technological development among nations and underscores the need for cooperation to ensure inclusive access, close the digital divide and enhance digital literacy, particularly in developing countries.

In that same line, the GDB highlights a widening digital divide in data governance. Lower-capability countries struggle to establish robust data protection frameworks and effective data-sharing mechanisms due to limited resources and expertise. This hampers their ability to safeguard citizen privacy and participate fully in the data revolution.

In Africa, for instance, the GDB data reveals that the region scores below the global average across all pillars (governance, capabilities, availability and use). There is a pressing need for investment in institutional frameworks, comprehensive data infrastructure and capacity building to harness data for public good effectively.

Addressing these disparities is crucial to prevent the entrenchment of existing inequalities and to ensure that the transformative potential of data is accessible to all.

Addressing the Gaps

Bridging the existing data governance gaps demands a comprehensive strategy that encompasses international collaboration, targeted support for lower-capability nations and initiatives focusing on knowledge sharing and capacity building. Empowering countries to develop robust data protection frameworks and fostering expertise in navigating complex data governance landscapes paves the way for more inclusive participation in the global data ecosystem.

Moreover, the establishment of clear and consistent legal frameworks for data sharing, including the creation of data governance bodies and the implementation of ethical guidelines, is essential for building trust and facilitating collaboration. Strengthening data security and privacy protection mechanisms further reinforces this foundation of trust.

Prioritizing standardization and interoperability is crucial to unlocking the full potential of open data, and ensuring its accessibility and usability across different platforms. Establishing common data formats and enabling seamless data exchange would empower individuals and organizations to harness the benefits of open data for public good initiatives.

Additionally, updating and reinforcing right-to-information legislation, coupled with capacity-building efforts, ensures that citizens have the necessary tools to hold governments accountable and actively participate in decision-making processes. This commitment to transparency and accountability underpins a democratic approach to data governance.

In embracing these measures, we not only pave the way for a more inclusive and equitable data landscape but also unlock the immense potential of data. By integrating the broad principles of data governance with the specific requirements of emerging technologies such as AI tools, we establish a robust framework that aligns data practices with ethical standards, legal requirements and societal expectations. This integrated approach forms the cornerstone of effective data governance, ensuring that data-driven innovation serves the collective good while upholding fundamental rights and values.

Notes

1 Public good is a contested concept. There are many publics, many different visions of how society should be organized and many views on the goals we should individually and collectively work toward. In the GDB, the Sustainable Development Goals, agreed through a broad international process, provide a common point of reference for identifying a set of particular public goods that data might help deliver, and that we can provide some global assessment against – from good health for all, to climate action, to just and strong institutions.

2 See UN News (2024).

3 This essay was developed before the Summit for the Future where world leaders adopted a Pact for the Future that includes a Global Digital Compact.

4 The Barometer is a multi-dimensional and multi-layered study that assesses the state of data for public good in 109 countries. An expert survey was conducted from May 2019 to May 2021 to create a new global benchmark that looks at data governance, capability, availability, and use and impact of data for public good. The data can be explored and viewed, through its core (for example, governance and capability) or thematic modules (for example, climate action, land and public procurement) as well as on a country level. Learn more about the data structure at https://globaldatabarometer.org/. In 2024, the second edition is in development stages (see https://globaldatabarometer.org/2023/11/exciting-news-the-second-edition-of-the-global-data-barometer-is-underway/).

5 In the first edition of the GDB *Global Report*, the availability pillar surveys the presence, openness and key features of selected data sets in order to understand to what extent each country is making key data sets accessible in structured online forms that are fit for purpose for public good use cases (GDB 2022).

6 For more information, see Council of Europe (2019).

7 For more information, see The Global Partnership on Artificial Intelligence (2023).

8 In instances where researchers could not locate specific data-sharing frameworks, they examined data-sharing provisions within data protection legislation.

Works Cited

Council of Europe. 2019. *Artificial Intelligence and Data Protection*. November. https://rm.coe.int/2018-lignes-directrices-sur-l-intelligence-artificielle-et-la-protecti/168098e1b7.

GDB. 2022. *Global Report*. First edition. https://globaldatabarometer.org/wp-content/uploads/2022/05/GDB-Report-English.pdf.

Kalkar, Uma and Natalia González Alarcón. 2023. *Facilitating Data Flows through Data Collaboratives: A Practical Guide to Designing Valuable, Accessible, and Responsible Data Collaboratives*, edited by Arturo Muente Kunigami and Stefaan Verhulst. http://dx.doi.org/10.18235/0005185.

Mishra, Vibhu. 2024. "General Assembly adopts landmark resolution on artificial intelligence." UN News, March 21. https://news.un.org/en/story/2024/03/1147831.

The Global Partnership on Artificial Intelligence. 2023. *The Role of Government as a Provider of Data for Artificial Intelligence*. Interim Report. November. https://gpai.ai/projects/data-governance/DG08%20-%20The%20Role%20of%20Government%20as%20a%20Provider%20of%20Data%20for%20Artificial%20Intelligence%20-%20Interim%20Report.pdf.

UN General Assembly. 2024. "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development." A/78/L.49. March 11. https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf.

Verhulst, Stefaan G. and Friederike Schüür. 2023. "Interwoven Realms: Data Governance as the Bedrock for AI Governance." *Data & Policy Blog*, November 20. https://medium.com/data-policy/interwoven-realms-data-governance-as-the-bedrock-for-ai-governance-ffd56a6a4543.

World Bank. 2021. *World Development Report 2021: Data for Better Lives*. Washington, DC: International Bank for Reconstruction and Development/The World Bank. www.worldbank.org/en/publication/wdr2021.

About the Author

Silvana Fumega is a specialist on the intersection between data, policy and inclusion. She is the co-founder of Data Against Feminicide (together with Catherine D'Ignazio and Helena Suárez Val) as well as a research affiliate at the Data + Feminism Lab, MIT and member of The Data Tank's Global Advisory Council.

During her career, she has served as a consultant for numerous international organizations, governments and civil society groups. She also acted as research and policy director of The Latin American Initiative for Open Data until December 2022. Since January 2023, she has been an independent consultant while also acting as the Global Data Barometer project's director.

Silvana holds a Ph.D. from the University of Tasmania, Australia. She also holds a master's degree in public policy from Victoria University of Wellington, New Zealand, and a degree in political science from the University of Buenos Aires, Argentina.

Data Governance Is Not Ready for AI

Susan Ariel Aaronson

There is no artificial intelligence (AI) without data. Yet policy makers around the world struggle to govern the data that underpins various types of AI (Office of the Privacy Commissioner of Canada 2023). At the national level, government officials in many countries have not yet figured out how to ensure that the large and often global data sets that underpin various types of AI are governed in an effective, interoperable, internationally accepted and accountable manner. At the international level, policy makers have engaged in negotiations but have made little progress. As a result, despite the centrality of data to AI, data governance and AI governance are disconnected.

This essay[1] examines the implications of this incoherence. Starting with an overview, the author then focuses on why data for AI is so difficult to govern. Next, the author examines the data governance challenges presented by AI and discusses why international data governance is a work in progress.

Most of the efforts to govern AI say relatively little about data, including the EU AI Act (Hunton Andrews Kurth 2024) and US President Joe Biden's executive order on AI (The White House 2023a). Given the importance of data to economic growth, data governance is a key component of twentieth-century governance. Moreover, how nations govern data has implications for the achievement of other important policy objectives, from protecting national security to advancing human rights (Jakubowska and Chander 2024; Aaronson 2018, 2022).

There is no internationally accepted definition of data governance. The United Nations defines data governance as "a systemic and multi-dimensional approach to setting policies and regulations, establishing leadership for institutional coordination and national strategy, nurturing an enabling data ecosystem, and streamlining data management" (Yao and Park 2020). The World Bank (2021) notes that data governance consists of four main tasks: strategic planning; developing rules and standards; creating mechanisms of compliance and enforcement; and generating the learning and evidence needed to gain insights and address emerging challenges.

Policy makers have been governing various types of data for centuries. Recent research by the Digital Trade and Data Governance Hub examined 68 countries and the European Union from 2019 to 2022. The authors found that data governance is a work in progress. Most nations protect specific types of data, such as intellectual property

(IP) or personal data, but are in the early stages of creating institutions and enforcement mechanisms to ensure that governance of data is accountable, democratically determined and effective. Additionally, many developing countries struggle to implement existing data laws and regulations (LaCasse 2024). Finally, countries have few binding data governance mechanisms at the international level (Struett, Aaronson and Zable 2023). These enforcement problems and governance gaps have become more visible since the popularization of generative AI, which is built on data scraped from around the Web (global data sets). Policy makers have struggled to protect personal and proprietary data taken from Web scraping, yet they have no means of ensuring that the globally scraped data is as accurate, complete and representative as possible (Aaronson 2024a).

Why Is Data Used in AI So Difficult to Govern?

Data Is Multidimensional

Data can simultaneously be a good and a service, an import and an export, a commercial asset and a public good. There are many different types of data, and policy makers must figure out how to protect certain types of data (such as personal or proprietary data) from misuse or oversharing while simultaneously encouraging such sharing in the interests of mitigating "wicked problems" — problems that are difficult for one nation alone to address because they transcend borders and generations (Aaronson 2022). When raw data is organized, it becomes information — information that society uses to grow economies, hold governments to account and solve wicked problems. Researchers see tremendous potential in the use of AI built on data to address such problems, but only if data is shared across borders.

Data for AI Is Multinational

Large language model (LLM) applications such as the chatbot ChatGPT are built on different sources of data. Moreover, data and algorithm production, deployment and use are distributed among a wide range of actors from many different countries and sectors of society who together produce the system's outcomes and functionality. These LLMs are at the bottom of a global product built on a global supply chain with numerous interdependencies among those who supply data, those who control data, and those who are data subjects or content creators (Cobbe, Veale and Singh 2023).

Data Markets Are Opaque

Researchers and policy makers have little information about the demand, supply or value of much of the data that underpins the data-driven economy. In addition, most entities collect personal and non-personal data yet reveal very little about the data they collect. Here, again, generative AI provides a good example. LLMs are generally constructed from two main pools of data (pre-filtered data sets). The first pool is comprised of data sets created, collected or acquired by the model developers. This pool of data can be considered proprietary because it is owned and controlled by the LLM developer. It may include many different types of data from many different sources, as well as computer-generated (synthetic) data created to augment or replace real data to improve AI models, protect sensitive data and mitigate bias (Martineau and Feris 2023). The second pool is made up of Web-scraped data, which is essentially a snapshot of a sample of the Web at a given moment in time. Although these scrapes provide a broad data sample, it is hard to determine if the sample is accurate, complete and representative of the world's data, a particular problem for generative AI.

Data Is Both Plentiful and Precious

On one hand, data is plentiful because almost every entity today, whether a government, a non-governmental organization such as Save the Children or a business such as Spotify, collects data about its stakeholders.[2] These same entities often use AI to analyze the data they have collected. On the other hand, governments and firms are taking steps to make data less plentiful. For example, policy makers increasingly recognize that large pools of data can be used to make predictions about country behaviour or to manipulate their citizens. As a result, countries such as Australia (Hammond-Errey 2022), Canada,[3] China (Cai 2021), the United Kingdom (Geropoulos 2023) and the United States (Busch 2023) now see such pools of data as a security risk as well as a privacy risk.

Data-Driven Sectors Are Built on Information Asymmetries

Firms with more computing power are better positioned to extract and use data. They have the expertise, the finances and generally the data to utilize AI. Moreover, firms with more data are more likely to create new data-driven goods and services, which, in turn, generate more data and more market power. This phenomenon also applies across countries. Only some 20 firms possess cloud infrastructure, computing power, access to capital and vast troves of data to develop and deploy tools to create LLMs (Staff in the Bureau of Competition & Office of Technology 2023). These firms are also concentrated in a few advanced developed countries — in Asia, Europe and North America. As a result, a few companies with expertise in generative AI could hold outsized influence over a significant swath of economic activity (Staff in the Bureau of Competition & Office of Technology 2023; Hacker, Engel and Mauer 2023; Khan and Hanna 2023). Without incentives, these companies may not be motivated to ensure that their data sets are broadly representative of the people and data of the world.

How Is AI Altering Data Governance?

AI is constantly evolving and has become a key element of many goods and services (Wharton Online 2022; McKinsey & Company 2023). Many analysts now view some variants of AI as a general-purpose technology — a technology that can affect not just specific sectors, but also the economy as a whole (Crafts 2021; Hötte et al. 2023). Because of the growing importance of AI to economic growth, government officials in many countries are determined to develop policies that advantage their AI firms over those of other countries. This phenomenon, called "AI nationalism," appears to be leading several countries to alter their data policies (Aaronson 2024b; *The Economist* 2024; Ian Hogarth 2018; Spence 2019).

Only two governments, China (Gamvros, Yau and Chong 2023) and the European Union,[4] have approved comprehensive AI regulation. Brazil, Canada and the United States, among others, are considering such regulation. But many of these efforts say very little about data. Some governments, such as Japan[5] and Singapore (Norton Rose Fulbright 2021), are so determined to encourage AI that they have declared that copyrighted articles could be scraped for generative AI.

Generative AI is created from two types of data: proprietary data that may include personal and copyrighted information from sources collected and controlled by the AI developer; and Web-scraped data. Developers do not have direct consent to utilize some of the Web-scraped personal and proprietary data (Argento 2023). Meanwhile, governments such as Canada, the United Kingdom and the United States are investigating the collection of such data for generative AI (Aaronson 2024a).

Policy makers have not yet figured out whether to encourage open-source versus closed or proprietary AI models. To be considered scientifically rigorous, all model developers

provide some information about their models, but open-source models provide greater detail about how they trained and filtered data and then developed their LLMs. Policy makers recognize that there are benefits and costs to open- versus closed-source models. Open-source models make it easier for outside researchers to utilize and improve a particular model and, consequently, may facilitate further research, while closed-source models are generally considered to be more reliable and stable (Davis 2023).[6] Some governments, including France (Robertson 2023) and the United Arab Emirates (Barrington 2023; *The National News* 2023), tout their support of an open-source approach to AI. The US government sought public comment and suggested that open-source models generally pose marginal risks; however, it should actively monitor any risks that could arise (National Telecommunications and Information Administration 2024, 2–3). China has done more than any other country to link data governance to its governance of generative AI (O'Shaughnessy and Sheehan 2023). The country requires AI service providers to do the following:

- use data and foundation models from lawful (legitimate) sources;
- not infringe others' legally owned IP;
- obtain personal data with consent or under situations prescribed by the law or administrative measures;
- take effective steps to increase the quality of training data, its truthfulness, accuracy, objectivity and diversity; and
- obtain consent from individuals whose personal information was processed.[7]

The European Union will soon finalize AI regulations that will require high-risk systems to provide more information about data provenance. In October 2023, the Biden administration issued an executive order on AI (The White House 2023b). Although the executive order mentioned data 76 times, it said very little about how data should be governed, except to say that personal data and IP should be protected.

In the name of national security, governments of countries such as China, the United Kingdom and the United States are making it harder to access large pools of personal or proprietary data (Sherman et al. 2023; Aaronson 2021). In Biden's executive order, the administration promised to consider the national security implications of the use of data and data sets on the training of generative AI models and makes recommendations on how to mitigate the risks related to the use of the data and data sets (The White House 2023b, section 4.4 B). If this proposal continues, AI developers will be less able to create accurate, complete and representative data sets (Aaronson 2024a).

The State of Global Data Governance and AI

The platform on which data services flow is a "commons," but policy makers in most nations have not focused on creating shared rules. Data generally flows freely among nations, but policy makers in a growing number of countries are erecting barriers to these flows. Internationally accepted rules would provide AI developers with certainty.

US policy makers first pushed for shared rules on cross-border data flows in 1997 with the Framework for Global Electronic Commerce.[8] Policy makers from the Organisation for Economic Co-operation and Development then established global principles (Thompson 2000; Organisation for Economic Co-operation and Development and the Inter-American Development Bank 2016, chapter 13), which were incorporated in various bilateral and regional trade agreements, such as the Digital Economy Partnership Agreement among Chile, New Zealand and Singapore,[9] and the Comprehensive and Progressive Agreement for Trans-Pacific Partnership[10] among 11 Pacific-facing nations.

These agreements delineated that nations should allow the free flow of data among signatories with long-standing exceptions to protect privacy, national security, public morals and other essential domestic policy goals.

In 2017, 71 nations began participating in the Joint Statement Initiative on e-commerce at the World Trade Organization (WTO). Today, some 90 members of the WTO are negotiating shared international provisions regarding cross-border data flows. These negotiations are being led by small open economies such as Australia and Singapore. Although the world's two largest economies and leading AI nations are participating, the United States and China are not key demandeurs of an agreement. The parties have made progress. In July 2024, participants agreed to what they called a "stabilized text." It includes language on personal data but no binding language regarding the free flow of data. The text says nothing about AI.[11]

As noted above, the United States led global efforts to encourage rules governing the free flow of data and exceptions to those rules since 1997. US policy makers argued that such rules would advance human rights, stimulate economic growth and clarify when nations could block such flows.[12] However, in November 2023, the United States announced that it would continue to negotiate such rules, but was seeking clarity and policy space to regulate the business practices of its data giants. Hence, the country could no longer support certain provisions on data flows, encryption and source code. With this new position, the United States seemed to be saying that the exceptions did not give it (and other nations) sufficient policy space for domestic regulation of data-driven technologies and business practices (Lawder 2023). Some argued that the United States was becoming more like China and India — nations that have long pushed for data sovereignty (Chander and Sun 2023; Mishra 2023). However, the Biden administration's first executive order did direct the US government to work internationally to set standards for the data underpinning AI (The White House 2023b, section 11).

Under international trade rules, a country cannot ban a product or service unless it can argue that such bans were necessary to protect public health, public morals, national security or other domestic policy objectives. In a rare move, Italy banned ChatGPT in 2023 for some three months, arguing that the AI application violated EU data protection laws. But in January 2024, the Italian data protection body, the Garante, announced it had finished its investigation and stated that OpenAI, the chatbot's parent company, had 30 days to defend its actions (Reuters 2024).

Meanwhile, policy makers are negotiating other agreements on AI, but these agreements are not focused on data. For example, in November 2023, some 18 countries, as well as the major AI firms, reached consensus on a non-binding plan for safety testing of frontier AI models (Satter and Bartz 2023). In November 2021, members of the United Nations Educational, Scientific and Cultural Organization (2021) agreed to a non-binding agreement on AI ethics.

Conclusion

Many of the world's people are simultaneously excited and scared by AI. They recognize that the technology could improve their quality and standard of living, but they also fear it could be misused (Kennedy 2023). Policy makers in many countries are responding to that ambivalence with policies to reduce risk, make AI safer, and ensure that AI is developed and deployed in an accountable, democratic and ethical manner. Yet policy makers do not seem to focus on data governance as a tool to accomplish these goals.

Why is data governance so disconnected from AI? This essay began by asserting several reasons: data is difficult to govern because it is multidimensional; data markets are

opaque; and data is simultaneously plentiful and scarce. The author noted that countries have different expertise and will to govern data, yet because data sets are global, policy makers must find common ground on rules. This sounds great on paper — but in the real world, the most influential AI powers are not leading efforts to govern data across borders. For example, China, India and the United States want policy space to govern data, data-driven technologies and data flows. In addition, many officials appear more concerned about their competitiveness in AI than about ensuring that the tedious process of negotiating internationally accepted rules on data is successful.

Hence, the author concludes this essay with a warning. Without such rules, it will be harder for AI developers to create accurate, complete and representative data sets. In turn, without accurate, complete and representative data sets, AI applications may continue to have significant flaws and inaccuracies. Users and policy makers may, over time, lose trust in the technology. And without trust, users and investors may turn to other methods for analyzing the world's data. If that were to happen, the world's people could fail to realize the full value of data.

Notes

1 This material is based on work supported, in part, by the National Institute of Standards and Technology-National Science Foundation (NIST-NSF) Institute for Trustworthy AI in Law and Society (TRAILS), which is supported by the NSF under award no. 2229885. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF, CIGI or the George Washington University (GWU).

2 Datafication refers to the transformation of subjects, objects and practices into digital data. Entities use this data to track individuals and make predictions about their behaviour. See Southerton (2020) and Shilova (2018); regarding the Spotify playlist, see Rojas (2024).

3 See www.canada.ca/en/security-intelligence-service/corporate/publications/csis-2021-public-report/national-security-threats.html.

4 The EU Parliament approved the AI Act in April 2024 (see Bracy and Andrews 2024) and the bill as of February 2, 2024 (see Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, 2021/0106(COD), online: <https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf>.

5 In Japan, copyrights are automatically generated when content is created, so not enforcing copyright made it easier to use older content (Nishino 2022; Wan 2023; Technomancers.ai 2023).

6 See https://allenai.org/olmo?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axioslogin&stream=top.

7 This regulation is the latest addition to AI regulations in China after the algorithm provisions in 2021 and the deep synthesis provisions in 2022. See Gamvros, Yau and Chong (2023); Cooley LLP (2023); Arcesati and Brussee (2023).

8 See https://clintonwhitehouse4.archives.gov/WH/New/Commerce/.

9 See www.mfat.govt.nz/en/trade/free-trade-agreements/free-trade-agreements-in-force/digital-economy-partnership-agreement-depa/depa-text-and-resources/.

10 See www.mfat.govt.nz/en/trade/free-trade-agreements/free-trade-agreements-in-force/cptpp/.

11 See www.wto.org/english/tratop_e/ecom_e/joint_statement_e.htm, which links to the stabilized text.

12 WTO, Work Programme on Electronic Commerce, *The Economic Benefits of Cross-Border Data Flows: Communication from the United States* (dated 14 June 2019), WTO Doc S/C/W/382.

Works Cited

Aaronson, Susan Ariel. 2018. *Data Is Different: Why the World Needs a New Approach to Governing Cross-border Data Flows*. CIGI Paper No. 197. Waterloo, ON: CIGI. www.cigionline.org/publications/data-different-why-world-needs-new-approach-governing-cross-border-data-flows/.

———. 2020. *Data Is Dangerous: Comparing the Risks That the United States, Canada and Germany See in Data Troves*. CIGI Paper No. 241. Waterloo, ON: CIGI. www.cigionline.org/publications/data-dangerous-comparing-risks-united-states-canada-and-germany-see-data-troves/.

———. 2022. *A Future Built on Data: Data Strategies, Competitive Advantage and Trust*. CIGI Paper No. 266. Waterloo, ON: CIGI. www.cigionline.org/publications/a-future-built-on-data-data-strategies-competitive-advantage-and-trust/.

———. 2024a. *Data Disquiet: Concerns about the Governance of Data for Generative AI*. CIGI Paper No. 290. Waterloo, ON: CIGI. www.cigionline.org/publications/data-disquiet-concerns-about-the-governance-of-data-for-generative-ai/.

———. 2024b. *The Age of AI Nationalism and Its Effects*. CIGI Paper No. 306. Waterloo, ON: CIGI. www.cigionline.org/publications/the-age-of-ai-nationalism-and-its-effects/.

Arcesati, Rebecca and Vincent Brussee. 2023. "China's Censors Back Down on Generative AI." *The Diplomat*, August 7. https://thediplomat.com/2023/08/chinas-censors-back-down-on-generative-ai/.

Argento, Zoe. 2023. "Data protection issues for employers to consider when using generative AI." IAPP, August 9. https://iapp.org/news/a/data-protection-issues-for-employers-to-consider-when-using-generative-ai.

Barrington, Lisa. 2023. "Abu Dhabi Makes Its Falcon 40B AI Model Open Source." *U.S. News & World Report*, May 25. www.usnews.com/news/technology/articles/2023-05-25/abu-dhabi-makes-its-falcon-40b-ai-model-open-source.

Bracy, Jedidiah and Caitlin Andrews. 2024. "EU countries vote unanimously to approve AI Act." IAPP, February 2. https://iapp.org/news/a/eu-countries-vote-unanimously-to-approve-ai-act.

Busch, Kristen E. 2023. "TikTok: Recent Data Privacy and National Security Concerns." Congressional Research Service Insight, March 29.

Cai, Vanessa. 2023. "China cracks down on theft of geographic data, warning of national security threat." *South China Morning Post*, December 11. www.scmp.com/news/china/politics/article/3244672/china-cracks-down-theft-geographic-data-warning-national-security-threat.

Chander, Anupam and Haochen Sun, eds. 2023. *Data Sovereignty: From the Digital Silk Road to the Return of the State*. New York, NY: Oxford University Press.

Cobbe, Jennifer, Michael Veale and Jatinder Singh. 2023. "Understanding accountability in algorithmic supply chains." In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1186–97. https://doi.org/10.1145/3593013.3594073.

Cooley LLP. 2023. "China Issues Measures on Generative Artificial Intelligence Services." Cooley LLP, August 7. www.jdsupra.com/legalnews/china-issues-measures-on-generative-3929442/.

Crafts, Nicholas. 2021. "Artificial intelligence as a general-purpose technology: an historical perspective." *Oxford Review of Economic Policy* 37 (3): 521–36. https://doi.org/10.1093/oxrep/grab012.

Davis, Gwen. 2023. "A developer's guide to open source LLMs and generative AI." *GitHub* (blog), October 5. https://github.blog/2023-10-05-a-developers-guide-to-open-source-llms-and-generative-ai/.

Gamvros, Anna, Edward Yau and Steven Chong. 2023. "China finalises its Generative AI Regulation." Norton Rose Fulbright, July 25. www.dataprotectionreport.com/2023/07/china-finalises-its-generative-ai-regulation/.

Geropoulos, Kostis. 2023. "Former MI5 chief warns of national security threats." NE Global, May 14. www.neglobal.eu/uk-former-counter-intelligence-chief-warns-of-new-cyber-security-concerns/.

Hacker, Philipp, Andreas Engel and Marco Mauer. 2023. "Regulating ChatGPT and other Large Generative AI Models." *arXiv*, May 12. https://arxiv.org/abs/2302.02337.

Hammond-Errey, Miah. 2022. "Big data and national security: A guide for Australian policymakers." Lowy Institute, February 1. www.lowyinstitute.org/publications/big-data-national-security-guide-australian-policymakers.

Hogarth, Ian. 2018. "AI Nationalism" *Ian Hogarth* (blog), June 13. www.ianhogarth.com/blog/2018/6/13/ai-nationalism.

Hötte, Kerstin, Taheya Tarannum, Vilhelm Verendel and Lauren Bennett. 2023. "AI Technological Trajectories in Patent Data: General Purpose Technology and Concentration of Actors." INET Oxford Working Paper No. 2023-09. www.inet.ox.ac.uk/publications/ai-technological-trajectories-in-patent-data/.

Hunton Andrews Kurth. 2024. "Final Draft of EU AI Act Leaked." *Privacy & Information Security Law Blog*, February 1. www.huntonak.com/privacy-and-information-security-law/final-draft-of-eu-ai-act-leaked.

Jakubowska, Ella and Sarah Chander. 2024. "Council to vote on EU AI Act: What's at stake?" *EDRi* (blog), January 31. https://edri.org/our-work/council-to-vote-on-eu-ai-act-whats-at-stake/.

LaCasse, Alex. 2024. "Report examines state of African nations' data protection laws, implementation efforts." IAPP, Februrary 1. https://iapp.org/news/a/evaluating-african-nations-comprehensive-privacy-laws-and-their-implementation/.

Lawder, David. 2023. "US drops digital trade demands at WTO to allow room for stronger tech regulation." Reuters, October 25. www.reuters.com/world/us/us-drops-digital-trade-demands-wto-allow-room-stronger-tech-regulation-2023-10-25/.

Kennedy, Alan. 2023. "World Risk Poll: Mapping Global Sentiment on AI." Visual Capitalist, July 12. www.visualcapitalist.com/sp/global-ai-opinion.

Khan, Mehtab and Alex Hanna. 2023. "The Subjects and Stages of AI Dataset Development: A Framework for Dataset Accountability." SSRN. October 11. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4217148.

Martineau, Kim and Rogerio Feris. 2023. "What is synthetic data?" *IBM* (blog), February 8. https://research.ibm.com/blog/what-is-synthetic-data.

McKinsey & Company. 2023. "The state of AI in 2023: Generative AI's breakout year." McKinsey & Company, August 1. www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year.

Mishra, Neha. 2023. "Data Governance and Digital Trade in India: Losing Sight of the Forest for the Trees?" In *Data Sovereignty: From the Digital Silk Road to the Return of the State*, edited by Anupam Chander and Haochen Sun, 240–63. New York, NY: Oxford University Press.

National Telecommunications and Information Administration. 2024. *Dual-Use Foundation Models with Widely Available Model Weights*. NTIA Report. July. Washington, DC: NTIA. www.ntia.gov/issues/artificial-intelligence/open-model-weights-report.

Nishino, Anna. 2022. "Japan to allow freer sharing of content with obscure copyrights." Nikkei Asia, June 3. https://asia.nikkei.com/Business/Media-Entertainment/Japan-to-allow-freer-sharing-of-content-with-obscure-copyrights.

Norton Rose Fulbright. 2021. "New Singapore Copyright Exception will propel AI revolution." *Inside Tech Law*, November 15. www.insidetechlaw.com/blog/2021/11/new-singapore-copyright-exception-will-propel-ai-revolution.

Office of the Privacy Commissioner of Canada. 2023. "Statement on Generative AI: Roundtable of G7 Data Protection and Privacy Authorities." OPC News, June 21. www.priv.gc.ca/en/opc-news/speeches/2023/s-d_20230621_g7/.

Organisation for Economic Co-operation and Development and the Inter-American Development Bank. 2016. *Broadband Policies for Latin America and the Caribbean: A Digital Economy Toolkit*. June. Paris, France: OECD. https://doi.org/10.1787/9789264251823-en.

O'Shaughnessy, Matt and Matt Sheehan. 2023. "Lessons From the World's Two Experiments in AI Governance." Carnegie Endowment for International Peace, February 14. https://carnegieendowment.org/posts/2023/02/lessons-from-the-worlds-two-experiments-in-ai-governance?lang=en.

Reuters. 2024. "OpenAI's ChatGPT breaches privacy rules, says Italian watchdog." Reuters, January 30. www.reuters.com/technology/cybersecurity/italy-regulator-notifies-openai-privacy-breaches-chatgpt-2024-01-29/.

Robertson, Derek. 2023. "France makes a big push on open-source AI." *Politico*, August 8. www.politico.com/newsletters/digital-future-daily/2023/08/08/france-makes-a-big-push-on-open-source-ai-00110341.

Rojas, Frank. 2024. "That Spotify Daylist That Really 'Gets' You? It Was Written by A.I." *The New York Times*, January 24. www.nytimes.com/2024/01/24/style/ai-spotify-music-playlist-algorithm.html.

Satter, Raphael and Diane Bartz. 2023. "US, Britain, other countries ink agreement to make AI 'secure by design.'" Reuters, November 27. www.reuters.com/technology/us-britain-other-countries-ink-agreement-make-ai-secure-by-design-2023-11-27/.

Sherman, Justin, Hayley Barton, Aden Klein, Brady Kruse and Anushka Srinivasan. 2023. *Data Brokers and the Sale of Data on U.S. Military Personnel: Risks to Privacy, Safety, and National Security*. November. https://techpolicy.sanford.duke.edu/data-brokers-and-the-sale-of-data-on-us-military-personnel/.

Shilova, Margarita. 2018. "The Concept of Datafication; Definition & Examples." Data Science Central, June 2. www.datasciencecentral.com/the-concept-of-datafication-definition-amp-examples/.

Southerton, Clare. 2020. "Datafication." In *Encyclopedia of Big Data*, edited by Laurie A. Schintler and Connie L. McNeely, 1–4. Cham, Switzerland: Springer. https://link.springer.com/referenceworkentry/10.1007/978-3-319-32001-4_332-1.

Spence, Sebastian. 2019. "The birth of AI nationalism." *The New Statesman*, April 10. www.newstatesman.com/science-tech/2019/04/the-birth-of-ai-nationalism-2.

Staff in the Bureau of Competition & Office of Technology. 2023. "Generative AI Raises Competition Concerns." *Office of Technology Blog*, June 29. www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns.

Struett, Thomas, Susan Ariel Aaronson and Adam Zable. 2023. *Data Governance Mapping Project Year 4*. Global Data Governance Mapping Project. Digital Trade & Data Governance Hub. https://globaldatagovernancemapping.org/.

Technomancers.ai. 2023. "Japan Goes All In: Copyright Doesn't Apply To AI Training." ACM News, June 1. https://cacmb4.acm.org/news/273479-japan-goes-all-in-copyright-doesnt-apply-to-ai-training/fulltext/.

The Economist. 2024. "Welcome to the era of AI nationalism." *The Economist*, January 1. www.economist.com/business/2024/01/01/welcome-to-the-era-of-ai-nationalism.

The National News. 2023. "The UAE has been quick to adopt AI. Now it must build and export it." *The National News*, September 4. www.thenationalnews.com/opinion/comment/2023/09/04/the-uae-has-been-quick-to-adopt-ai-now-it-must-build-and-export-it/.

The White House. 2023a. "Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence." Statements and releases, October 30. www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.

———. 2023b. "Executive Order on the Safe, Secure and Trustworthy Development and Use of Artificial Intelligence." October 30. www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

Thompson, Mozelle W. 2000. "U.S. Implementation of the OECD E-Commerce Guidelines." Federal Trade Commission, February 12. www.ftc.gov/news-events/news/speeches/us-implementation-oecd-e-commerce-guidelines.

United Nations Educational, Scientific and Cultural Organization. 2021. "UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence." Press release, November 25. www.unesco.org/en/articles/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence.

Wan, Audrey. 2023. "Why Japan is lagging behind in generative A.I. – and how it can create its own large language models." CNBC, July 6. www.cnbc.com/2023/07/07/why-japan-is-lagging-behind-in-generative-ai-and-creation-of-llms.html.

Wharton Online. 2022. "How Do Businesses Use Artificial Intelligence?" *Wharton Online* (blog), January 19. https://online.wharton.upenn.edu/blog/how-do-businesses-use-artificial-intelligence/.

World Bank. 2021. *World Development Report 2021: Data for Better Lives*. Washington, DC: World Bank. www.worldbank.org/en/publication/wdr2021.

WTO. 2023. "E-commerce co-convenors set out roadmap for concluding negotiations in early 2024." November 30. www.wto.org/english/news_e/news23_e/jsec_30nov23_e.htm.

Yao, Keping and Mi Kyoung Park. 2020. "Strengthening Data Governance for Effective Use of Open Data and Big Data Analytics for Combating COVID-19." Policy Brief No. 89. www.un.org/development/desa/dpad/publication/un-desa-policy-brief-89-strengthening-datagovernance-for-effective-use-of-open-data-and-big-data-analytics-for-combating.

About the Author

Susan Ariel Aaronson is a CIGI senior fellow, research professor of international affairs at George Washington University (GWU) and co-principal investigator with the NSF-NIST Institute for Trustworthy AI in Law & Society, where she leads research on data and AI governance. She is also a GWU Public Interest Technology Scholar. Her research interests relate to economic change, human rights and good governance.

Susan directs the Digital Trade and Data Governance Hub at GWU. The Hub was founded in 2019 and educates policy makers, the press and the public about data governance and data-driven change through conferences, webinars, study groups, primers and scholarly papers. It is the only organization in the world that maps the governance of public, proprietary and personal data at the domestic and international levels. The Hub's research has been funded by foundations such as Ford and Minderoo.

Susan directs projects on defining AI protectionism; how governments may incentivize more accurate, complete and representative data sets; and how open-source AI builds trust. She regularly writes op-eds for *Barron's* and has been a commentor on economics for NPR's *Marketplace*, *All Things Considered* and *Morning Edition*, and for NBC, CNN, the BBC and PBS.

A Mission-Driven Approach to Data for People and Planet

Lorrayne Porciuncula

In an increasingly interconnected and digital world, data has emerged as a powerful tool for addressing some of the most pressing challenges facing humanity. From combatting climate change to reducing inequality and improving public health, the potential of data to drive positive change is immense. However, realizing this potential requires a coordinated and concerted effort on a global scale. While various processes related to data governance exist — convened by international organizations, governments, businesses and civil society — they often operate in silos, are poorly resourced and lack effective coordination. This fragmented landscape inhibits progress toward a more inclusive and sustainable digital future.

In this essay, the case for adopting a mission-driven approach to data is explored, drawing on examples of how and why data governance is a wicked problem that is becoming increasingly important to address. The essay also explores why leveraging the successful multi-stakeholder and global cooperation models of the United Nations and the G20 could improve coordination between stakeholders on an issue of global concern and catalyze sustainable approaches to foster trust in data-driven innovation such as artificial intelligence (AI) technologies. This approach could be fostered either by creating a Data20 multi-stakeholder space (D20) in the Group of Twenty (G20) or an International Decade on Data for People and Planet (IDD) within the UN system.

The Wicked Problem

Despite the importance of data in achieving the Sustainable Development Goals (SDGs) and advancing AI and other technologies, it also poses significant challenges. These challenges are intertwined and relate to multiple human rights, as well as policy, economic and security dimensions. To name a few:

- **The volume, velocity and variety of data is growing, but access to it is asymmetrical:** Many critical data sets are locked behind proprietary systems, limiting access for researchers, policy makers and organizations that could leverage this data for social and environmental progress. Moreover, there are vast disparities in data literacy, availability and quality across regions and communities. Developed countries often

have better data infrastructure, while developing nations may lack essential data collection and analysis capabilities, leading to data divides and asymmetries that hinder global progress. There is also a need for the right environment of incentives to be put in place to encourage companies to proactively identify data sets that could create value for all (Porciuncula 2023). Under appropriate protections and value-sharing agreements (if needed), data from companies could be made accessible for reuse in the public interest to promote social and environmental goals.

- **Data flows raise various concerns, overlapping security, economic and human rights dimensions:** Although a valuable resource, data can also be a potential source of risk, ranging from privacy concerns to security threats and even national security. Efforts to limit the free flow of data across borders reflect concerns over sovereignty and control, as nations seek to protect their citizens' data and assert jurisdiction over digital domains. The Datasphere Initiative's report, *We Need To Talk About Data: Framing the Debate Around the Free Flow of Data and Data Sovereignty*, discusses "data sovereignty" and the "free flow of data," and offers key recommendations to foster a collaborative discussion on how to organize our common data sphere (de La Chapelle and Porciuncula 2021). Key to this approach is addressing the complex issue of the diversity of cultures, even those existing within a single country, and legal contexts to effectively build trust and enable the free flow of data.
- **Data is not addressed in a holistic agenda:** A space is needed to bring together diverse stakeholders with a shared interest in coordinating and shaping the future of our digital society (ibid.). For example, efforts to address trust in data flows should start by opening up more conversations about data and how to responsibly unlock its value for all, identifying what is that value and what its relevance is for different actors. These conversations should happen not only at the highly technical and policy levels, but also with all actors involved. Inclusion is a fundamental element in these discussions to support trust in outcomes and needs to be transversal to all policies, frameworks and solutions developed.

In a world facing wicked problems such as pandemics and climate change, Mariana Mazzucato (2021) calls for bold solutions. Data governance is one such challenge, requiring collaboration across public and private sectors. A mission-driven approach can bridge development gaps, leverage private sector expertise and incentivize data sharing. This will foster innovation to tackle global challenges and ensure that everyone benefits from the data revolution.

A Moonshot on Data for People and Planet

In February 2024, the Datasphere Initiative hosted the event "Digital Dialogues: Thinking Together about the G20 Digital Agenda," bringing together a diverse group of stakeholders to further the debate on digital economic priorities in the context of the Brazilian presidency of the G20. Through structured panels and discussion sessions on the priorities of the G20 Digital Economy Working Group (including meaningful connectivity, AI, digital government and information integrity) and intersectional topics (for example, data, climate justice, Indigenous rights, gender, youth, and so on), the more than 300 participants reached three main action points:

- **Multi-stakeholder cooperation and engagement:** The G20 needs to improve collaboration with other stakeholder groups (T20, C20, and so on) and consider how digital issues intersect with broader societal concerns (for example, climate change and gender equality) to create a more inclusive digital agenda.

- **Catalyzing a mission-oriented process:** A mission-driven approach to data governance was proposed to ensure that everyone benefits from the digital economy. This approach would involve investment, education, collaboration and leveraging data for the SDGs.
- **Innovating in multi-stakeholder cooperation:** In a rapidly changing digital world, overcoming complex challenges requires innovative collaboration. Agile frameworks such as sandboxes were seen as key for faster solutions, emphasizing the need for global cooperation with local partnerships.

In light of the agreement and action points reached as part of the digital dialogues, there is a need to catalyze a coordinated process that gives data "its own place in the multilateral agenda so that the different facets of data can be addressed holistically" (Diepeveen and Kapoor 2024) across the Global North and Global South. An intentional approach to data in its own right could thus take the form of a D20 group within the G20 or an IDD powered by the UN system.

A D20

The G20 is currently powered by 13 engagement groups, which provide recommendations to the G20 leaders and contribute toward the policy-making process. The creation of the D20 engagement group would provide high visibility to data governance discussions and a unique multi-stakeholder space to encourage a holistic approach to data to effectively unlock its value for people and the planet.

In fact, a D20 could bring together a variety of stakeholders and nations that engage in the G20 around a common mission for data. Existing digital inequalities are grounded in data issues, and the unequal distribution of the social and economic value of data is a systemic issue that requires systemic solutions. A D20 "could collate, quality-assure and share insights from complementary processes and anchor collaborative action from different stakeholders. It could also serve the G20 as a space to help encourage and feed into a more globally inclusive and multi-stakeholder knowledge base about data use, challenges and policy approaches" (Diepeveen and Kapoor 2024).

The objective of establishing a D20 is to allocate the time, space and resources to address the aforementioned challenges and to allow others to be collectively identified and tracked. A D20 will provide a unified platform for stakeholders to coordinate efforts, shape the global agenda on data governance and further the implementing of concrete solutions.

An IDD

In the face of our digital future, another mission-driven approach to data could include the establishment of an IDD as a powerful international tool to guide our path through the data-driven landscape. Throughout history, the United Nations has harnessed the power of International Decades to address global imperatives.[1] An IDD was first recommended in the report of the High-Level Advisory Board on Effective Multilateralism (HLAB 2023), which stipulated that effective multilateralism must support critical, multilateral and generational reflection on the benefits and risks of the digital age. This proposal has been further developed by David Passarelli, Muznah Siddiqui and Alona Savishchenko (2023), Lorrayne Porciuncula (2023) and Stefaan G. Verhulst (2023).

An IDD could represent a pivotal opportunity to leverage the power of data for transformative global impact. Drawing on principles of collaborative governance, inclusive development and sustainable innovation, an IDD could help address pressing global challenges through a mission-oriented approach to data governance. Through

collaborative governance models, long-term vision and planning and the redefinition of success metrics, this IDD could build resilient data ecosystems that can drive meaningful change and address the complex challenges facing humanity.

Aiming High: Mission-Driven Processes

Central to these arguments for the D20 and the IDD is the adoption of a mission-oriented approach to data governance. By setting ambitious goals and mobilizing resources around key priorities such as climate action, poverty reduction and public health, we can harness the power of data to drive meaningful impact. Mission-oriented initiatives provide a framework for aligning diverse stakeholders and catalyzing innovation in pursuit of common objectives.

By centring the discussion around people and the planet, a mission-driven space such as the D20 or the IDD would prioritize the interests and needs of individuals and communities, ensuring that data governance serves the broader goals of social justice, inclusion and sustainability. The following subsections present some of the key governance challenges and societal issues that these approaches to data would address.

Reimagining Global Governance

The rapid proliferation of digital technologies and the growing importance of data in shaping economies and societies call for a reimagining of global governance structures. Traditional models of governance are ill-equipped to address the complex and interconnected nature of contemporary challenges. A global concerted effort could aim at establishing new frameworks that prioritize inclusive, collaborative and digitally self-determined approaches to data governance. Collaborative models, such as public-private partnerships and multi-stakeholder initiatives, would also be instrumental in driving progress toward shared data governance objectives. The private sector is the key to designing innovations and leveraging collected and produced data through models for data-sharing and use, such as homomorphic encryption. Similarly, governments play a crucial role in funding research, infrastructure and capacity building to ensure the availability and accessibility of high-quality data. Collaborative partnerships between governments, businesses, academia and civil society are essential for leveraging diverse expertise and resources to address complex challenges.

Fostering Investment and Collaboration

Effective data governance requires substantial investment and collaboration across sectors. Less than 10 years away from the deadline for the achievement of the SDGs, the role of funders to catalyze change and foster a more equitable data economy is more important than ever. Yet donors, as well as funded agencies and initiatives, are facing a complex web and fragmented environment that is hard to navigate and tends to derive from people demanding accountability from funders and funders being unable to generate their full impact. The estimated SDG funding gap for developing countries is US$4.2 trillion and growing (United Nations Inter-agency Task Force on Financing for Development 2024). A new compact for investors and funders to foster interoperability and data sharing would facilitate the deployment of more capital toward more SDG solutions at scale. Data plays a vital role in this goal, and framing an intentional approach with data as a common thread could help connect and support a more collaborative community to unlock the value of data for all.

Redefining Value Creation

In the digital age, value creation goes beyond traditional economic metrics to encompass social, environmental and ethical considerations. Data-driven innovations have the potential to generate immense value for society, but this value must be distributed equitably and sustainably. Our societies and economies are deeply unequal and fragmented, and the value of data is unevenly distributed. Large corporations are more often benefitting from collecting and processing data, but less frequently sharing and distributing its value with communities and the public in general. Moreover, the urgent need to address environmental challenges such as climate change requires innovative solutions informed by data-driven insights. From monitoring deforestation and tracking wildlife populations to optimizing renewable energy systems, data plays a critical role in advancing environmental sustainability. A process that has a clear mission to put people and the planet at the centre of how the value of data is being created could provide an opportunity to redefine value creation in ways that prioritize well-being and sustainability over short-term profits.

Planning for the Long-Term and Monitoring Progress

Achieving meaningful impact with data requires long-term vision and planning. A mission-driven approach could provide a platform for strategic planning and goal setting, enabling countries and organizations to align their efforts and investments with long-term sustainability objectives. It can also provide the framework to measure the success of data initiatives in a coordinated manner, moving beyond traditional economic indicators to embrace a broader set of metrics that capture social and environmental well-being. Indicators such as data accessibility, quality and impact on human development can provide valuable insights into the effectiveness of data governance efforts. By adopting a forward-thinking approach anchored in key performance metrics, we can ensure that data initiatives contribute to lasting positive change and resilience in the face of future uncertainties.

Maintaining Adaptability

While setting ambitious goals and mobilizing resources to achieve them is important, the path to achieving these goals may require flexibility and adaptability. Therefore, a mission-driven process for data should also be iterative, allowing for experimentation and learning along the way. This requires stakeholders to be proactive, responsive and willing to adjust strategies based on feedback and new information.

One way a mission-driven process could foster agility is by establishing sandboxes, which are controlled environments in which stakeholders can experiment with new ideas and approaches. Sandboxes provide a safe space for testing innovative solutions and allow for rapid iteration based on feedback (Datasphere Initiative 2022). By creating sandboxes, the D20 or an IDD can encourage experimentation and innovation, enabling stakeholders to test new approaches to data governance and collaboration.

Seeing Through the Clout of Emerging Technologies

The clout surrounding emerging technologies, particularly AI, has led to a surge of policy initiatives, political attention and investments across various sectors, from health care to agriculture. This wave of interest is driven by both the excitement of potential advancements and the fear of unforeseen consequences. While the intense focus on AI has catalyzed essential policy and technical work, it often overshadows the fundamental element at AI's heart: data. More nuanced discussions are necessary to ensure that the

conversation includes how we collect, process and use data, which are critical factors influencing AI's effectiveness and ethical deployment.

Data governance plays a pivotal role in shaping AI governance, determining how fairly we distribute the benefits of AI and mitigate its associated risks. The quality and integrity of data directly impacts the outcomes that AI technologies can produce. For instance, without properly collected and cleaned data, AI algorithms cannot learn effectively or make accurate predictions. This is particularly critical in high-stakes applications such as predicting extreme weather events. Inaccurate data can lead to false alerts or missed warnings, potentially resulting in significant agricultural losses or even more severe natural disasters. Hence, robust data governance frameworks are essential to ensure that AI systems are reliable, fair and beneficial to society.

Moving from a patchwork of uncoordinated actions around AI to one that ensures that AI works for people and planet requires emphasizing the core role of data governance in responsible AI use. By prioritizing data governance, we can ensure that AI systems are built on a foundation of high-quality, unbiased and ethically sourced data. This approach not only enhances the accuracy and reliability of AI predictions, but also fosters trust among users and stakeholders. Moreover, it enables policy makers and technologists to address the broader implications of AI, including equitable access to its benefits and the minimization of risks. In essence, a mission-driven process for data governance could ensure that AI development aligns with societal values and ethical standards, promoting a future where AI serves the common good.

Conclusion

In conclusion, both the D20 and the IDD highlight the potential for a mission-driven approach to data governance, with each having their own strengths and challenges. A D20 could leverage the G20's established framework to provide high visibility and a dedicated platform for data governance among major economies, fostering targeted solutions to data inequalities. However, its focus might be limited to G20 members, potentially excluding valuable perspectives from non-G20 nations and particularly developing countries. Conversely, an IDD under the UN's inclusive framework would ensure global participation and long-term vision, fostering collaborative governance and sustainable innovation to address global challenges holistically. But despite these advantages, an IDD might struggle to secure support for its approval and sustained commitment and resources over a decade, risking fragmentation of efforts. Ultimately, both approaches underscore the critical need for a coordinated global agenda to responsibly unlock the value of data for societal and environmental benefits.

Data is shaping cultures, communities, economies and society at large. A mission-driven process, such as the creation of a D20 or an IDD, could help align investment, resources and long-term planning to tackle data governance collectively. Effective multi-stakeholder shaping of such an agenda could provide an opportunity to re-evaluate how we embrace and develop data-driven technologies and how they interact with people and our planet. In the face of AI, the time is now to re-imagine our relationship with data and collaborate across geographies, sectors, stakeholders and generations to redefine our digital world in a way that works for all.

Note

1 See www.un.org/en/observances/international-decades.

Works Cited

Datasphere Initiative. 2022. *Sandboxes for data: creating spaces for agile solutions across borders*. May 25. www.thedatasphere.org/datasphere-publish/sandboxes-for-data.

de La Chapelle, Bertrand and Lorrayne Porciuncula. 2021, *We Need to Talk About Data: Framing the Debate Around Free Flow of Data and Data Sovereignty*. Internet and Jurisdiction Policy Network, April 6. www.thedatasphere.org/datasphere-publish/we-need-to-talk-about-data.

Diepeveen, Stephanie and Astha Kapoor. 2024, "We need global coordination on data, not just AI. Here's why." World Economic Forum, June 14. www.weforum.org/agenda/2024/06/need-global-coordination-on-data-not-just-ai/.

HLAB. 2023. *A Breakthrough for People and Planet: Effective and Inclusive Global Governance for Today and the Future*. New York, NY: United Nations University. https://highleveladvisoryboard.org/breakthrough.

Mazzucato, Mariana. 2021. *Mission Economy: A Moonshot Guide to Changing Capitalism*. New York, NY: Harper Business.

Passarelli, David, Muznah Siddiqui and Alona Savishchenko. 2023. "An International Decade for Data: Multistakeholder Cooperation in a Data-Driven World." Working Paper. United Nations University Centre for Policy Research. November. https://unu.edu/publication/international-decade-data-multistakeholder-cooperation-data-driven-world.

Porciuncula, Lorrayne. 2023a. "Data: The Missing Piece in ESD Frameworks." *Aspen Digital* (blog), July 17. www.aspendigital.org/blog/missing-piece-in-esg.

———. 2023b. "Why We Need an International Decade for Data." Working Paper. United Nations University Centre for Policy Research. November. https://unu.edu/publication/why-we-need-international-decade-data.

United Nations Inter-agency Task Force on Financing for Development. 2024. *Financing for Sustainable Development Report 2024: Financing for Development at a Crossroads*. New York, NY: United Nations. https://desapublications.un.org/publications/financing-sustainable-development-report-2024.

Verhulst, Stefaan G. 2023. "Unlocking the Potential: The Call for an International Decade of Data." Working Paper. United Nations University Centre for Policy Research. October 27. https://unu.edu/publication/unlocking-potential-call-international-decade-data.

About the Author

Lorrayne Porciuncula is the executive director of the Datasphere Initiative. For the last 15 years, her professional and academic experiences have been focused on issues around data, internet governance, infrastructure regulation and communication policy. Prior to leading the Datasphere Initiative, she was the director for the data program at the Internet & Jurisdiction Policy Network (2020–2021), where the Datasphere Initiative was incubated. She worked at the Organisation for Economic Co-operation and Development (OECD) (2014–2020) as the strategic advisor and internet economist for the Digital Economy Policy Division, coordinating issues related to data governance, AI and blockchain, and leading the production of several reports and country studies related to connectivity, infrastructure regulation, technology convergence and inclusion.

Lorrayne has acted as the OECD focal point and speaker at high-level international meetings and fora such as the Internet Governance Forum, UN Broadband Commission for Sustainable Development (UN-BBCom), the Asia-Pacific Telecommunity, UN Economic and Social Commission for Asia and the Pacific, UN Economic Commission for Latin America and the Caribbean, and the EQUALS Global Partnership for Gender Equality in the Digital Age. Prior to the OECD, Lorrayne worked as an economist at the International Telecommunication Union, in the Secretariat of the UN-BBCom (2012–2014).

Lorrayne is an affiliate of the Berkman Klein Center for Internet & Society at Harvard University, conducting research on data for development. She holds a master's degree in development economics from the Graduate Institute of International and Development Studies, Switzerland, an executive MBA from the Quantic School of Business and Technology and a bachelor's degree in international relations from the University of Brasilia, Brazil. She speaks English, Portuguese, Spanish and French.

The Shifting Value of Personal Data

Teresa Scassa

Data is in high demand for research and innovation, and data about humans and their activities is particularly sought after. Such data undoubtedly has commercial value — but it also has a different kind of value for those to whom it pertains. That value has typically been articulated in non-monetary terms, focusing on the importance of personal data to an individual's autonomy and dignity. However, in an increasingly data-driven society, how data about humans and their activities is categorized and valued is changing. Arguments about the importance of both group privacy and collective privacy broaden the rights-based focus from individuals to larger groups. In addition, the kinds of data in which individuals have an interest — and the nature of those interests — are also evolving. This essay is about those changing interests, as well as emerging data rights that give individuals — and, perhaps, communities — more control over both the personal and non-personal data that they generate.

Personal and Anonymized Data

The social and economic importance of personal data is well understood. Personal data — typically defined as "information about an identifiable individual" — is central to a person's identity and can reveal intimate details about them. For businesses and governments that collect personal data, this information is often necessary to provide goods or services to specific individuals. However, personal data can also be used in the design and creation of goods or services. For example, such data can be used to create profiles or to drive targeted advertisements. Personal data is used in massive quantities to train AI systems; it is also important to a variety of research activities.

Data protection laws place limits on the use of personal data, including on its sharing with others. Organizations seeking to use data in new ways, or to monetize their stores of personal data, often run up against these laws, making the use of personal data legally complex. As a result, organizations have turned to anonymization as a means of freeing up data for reuse. Anonymization, which can be defined as "irreversibly and permanently modify[ing] personal information, in accordance with generally accepted best practices, to ensure that no individual can be identified from the information, whether directly or indirectly, by any means,"[1] is used where data about people is required, but where it is unnecessary for that data to be linked to identifiable individuals. By breaking the link to the individual, it can free the data from governance under data protection or privacy laws. The European Union's General Data Protection Regulation (GDPR), for example, states that data protection law "should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not

or no longer identifiable."[2] Similarly, both section 6(5) of Canada's proposed Consumer Privacy Protection Act (CPPA) and section 1798.140 (v)(3) of the California Consumer Privacy Act would place anonymized data out of scope of the legislation. This "out of scope" nature of anonymized data lasts only so long as the data remains anonymous. If re-identification takes place, the data falls once again within the definition of personal information. The growing risk of re-identification — due to vast quantities of data, growing compute power and sophisticated algorithms — means that many data protection laws now impose penalties for the deliberate re-identification of anonymized data.

From Individual to Group Privacy

Data protection laws privilege the protection of personal data because of its link to rights-bearing individuals. An individualist notion of privacy — tied to a person's autonomy and dignity — is the foundation for its protection. Data protection laws set rules for the collection and processing of personal data, for exceptions to those rules in the public interest, and for compromises that attempt to align the expectations of individuals with certain data practices seen as necessary or socially beneficial. Anonymization fits within these frameworks as a means of freeing data for reuse.

However, the growing power of analytics techniques has raised concerns about the use of the data of many to interpret and shape the lives of both groups and individuals. Anonymized data permits inferences and generalizations that can have as powerful an effect on both individuals and groups as personal data — regardless of accuracy. Much of this activity is known as "profiling" (Hildebrandt 2008).

These deterministic uses of data have spurred calls for greater attention to the concept of "group privacy" — a theory that challenges data protection law's individualistic approach to privacy rights. The concept of group privacy highlights how anonymized data can be used in the profiling of individuals and the creation of ad hoc groups (Floridi 2014; Mittelstadt 2017). An ad hoc group serves particular needs (such as targeting of advertising) and might be defined in terms of a cluster of presumed shared characteristics using profiles and inferences. Leveraging the group privacy concept, a broader human rights-based approach to data protection would be less exclusively concerned with how particular data points relate to a specific individual and what rights of control the individual has, and more concerned with how data — even if anonymized — impacts the lives and choices of people more generally. Group privacy is an uneasy fit with data protection law. The privacy dimensions of profiling (which can impact dignity and autonomy) are not properly addressed by individual consent and control. They require a greater focus on regulating activities or outcomes.

Collective Interests in Data

In addition to this "group privacy" approach to data, which shifts the narrative from the identifiable individual's control over their personal information to group and individual harms, there have been growing claims in various contexts for collective rights to some categories of data that are about both individuals and the communities to which they belong. One example is the growing Indigenous data sovereignty movement, where Indigenous communities assert sovereign rights over data about the members of the community (Kukutai and Taylor 2016; Walter and Russo Carroll 2021). Such claims are founded on principles of self-determination. There are also other collective rights claims to data that are based on empowerment and enfranchisement concerns. For example, Ontario's Black Health Equity Working Group (2021) makes a strong case for a form of collective governance of the health data of Black communities in Ontario. In

the context of the failed Sidewalk Labs smart cities proposal for the Quayside land in Toronto, there were proposals for community governance of data that would have been collected in and about the development (Scassa 2020). Taking an even broader definition of community data rights, the Government of Ontario (2022, section 3.3.1) has explored the principles that might inform decisions to provide access for research and innovation to the province's significant stores of health administrative data. These discussions have included considering how public benefit might be derived from the sharing (for example, in the form of intellectual property, royalties, access to medical treatments and so on), as well as social licence (Paprica, Nunes de Melo and Schull 2019).

Claims to collective rights in data emphasize collective interests in the data about a community and the right of the community to control and benefit from it. To the extent that such data is also personal data, it may be separately protected as personal data — the collective rights claim can be layered on top of rights to personal information — and it can also attach to anonymized data of the collective.

Human-Derived Data

Another category of human-derived data is linked to humans because it is derived from them or their activities, but it may never have been personal data and, as a result, it is also not anonymized data — at least in the sense of having been processed to achieve anonymity (Scassa 2023). An example is data extracted from wastewater — a practice that became much more widespread during the COVID-19 pandemic. Such data reveals the kinds and volumes of excreted virus found in wastewater and has proven to be extremely useful in understanding the presence and trajectory of the virus. Wastewater testing is also used to detect other diseases or substances (Scassa, Robinson and Mosoff 2022). When such data is collected from public wastewater systems, it is generally not linked to individuals (although it is not impossible for the location and methods of collection and correlation to create a risk of identification). Human-derived data can also be used to make decisions that impact groups; for example, the presence of certain banned substances in wastewater from particular communities could lead to decisions by public authorities to increase policing in that community — or to increase public health interventions. Although on one level, this will seem like basic data-driven decision making, it is arguable that the extraction of human-derived data from public infrastructure brings with it, at least, obligations of public notice and engagement (ibid.).

The Co-creation of Data

Increasingly, laws may recognize additional rights or interests of individuals in data generated by their activities. An early example is the introduction, in the European Union's GDPR, of a new right of control over personal data — a data portability right. A similar right is included in section 1798.130 of California's Consumer Privacy Act. In Canada, Bill C-27 recognizes a new data mobility right (section 72), and section 27 of Quebec's newly amended private sector data protection law provides for data portability. Data portability is in only its very early stages in Canada. If passed, the CPPA will enable the mobility of data on a carefully curated, sector-by-sector basis. The first experiment will be with open banking (or consumer-directed finance), which has been promised for early 2025 (FinTech Global 2024). In the case of both Canada's CPPA and Quebec's Loi 25, the right is limited to a subset of personal data. For example, under section 72 of the CPPA, it will be personal data that is collected from the individual. Under section 27 of Loi 25, it is "computerized personal information collected from the applicant, and not created or inferred using personal information concerning him." Data portability rights essentially make a subset of personal data portable in the hands of the data subject. This important new right is more closely linked to competition and consumer rights than it is to privacy per se.

The right of control over personal data that is reflected in data portability rights, has been extended in the European Union beyond personal data to include non-personal data generated through human interaction with digital products and services by the EU Data Act. Thus, for example, data about one's connected car (for example, about its road or engine performance) is generated through the driver's activities, but it is not personal data. Yet the Data Act would give the individual the right to obtain that data — or to port it to a service provider of their choice. Rights under the Data Act can also be exercised by organizations whose activities generate data. These rights — under both the GDPR and the Data Act — recognize the value of data and give greater control to those who are essentially its co-creators. According to recital 15 of the Data Act, "The data represent the digitisation of user actions and events and should accordingly be accessible to the user."[3]

Conclusion

Although the link between personal data and the individual is grounded in clearly recognized privacy rights, there may be rights in data that go beyond those of the individuals who are the original source. Interests may also extend beyond privacy rights, with the ability to exercise control over data offering a growing range of benefits for individuals or groups. These may include better or more competitive services (as where porting one's data permits access to different service providers) or downstream benefits (as where the exercise of collective interests in data gives a community the ability to insist on some form of give-back).

While data protection laws have typically also balanced privacy rights against competing interests in the use of personal data, new legislated approaches to data are beginning to recognize the interests of individuals not just as data subjects, but as co-creators of data in certain contexts. This adds a different type of weight in any balancing of interests — indeed, it alters the nature of the interests. None of this should come as a surprise. As the social and economic importance of data continues to grow, it is natural that how we understand and negotiate the different interests in that data will also change. The signs of this change are becoming evident in both emerging legal frameworks and in novel claims by individuals, groups and communities. As these changes begin to shape domestic approaches, they may also pose new challenges to international data governance frameworks.

Notes

1 Bill C-27, *An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts*, 1st Sess, 44th Parl, 2022, s 2(1) (first reading 16 June 2022), online: <www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading>.

2 EC, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, [2016] OJ, L 119/1, recital 26.

3 EC, *Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)*, [2023] OJ, L 2023/2854, recital 15.

Works Cited

Black Health Equity Working Group. 2021. *Engagement, Governance, Access, and Protection (EGAP): A Data Governance Framework for Health Data Collected from Black Communities*. https://blackhealthequity.ca/wp-content/uploads/2021/03/Report_EGAP_framework.pdf.

FinTech Global. 2024. “Canada to boost financial innovation with new open banking laws.” FinTech Global, April 18. https://fintech.global/2024/04/18/canada-to-boost-financial-innovation-with-new-open-banking-laws/.

Floridi, Luciano. 2014. “Open Data, Data Protection, and Group Privacy.” *Philosophy & Technology* 27: 1-3. https://doi.org/10.1007/s13347-014-0157-8.

Government of Ontario. 2022. *Ontario Health Data Council Report: A Vision for Ontario’s Health Data Ecosystem*. October 13. www.ontario.ca/page/ontario-health-data-council-report-vision-ontarios-health-data-ecosystem#section-4.

Hildebrandt, Mireille. 2008. “Defining Profiling: A New Type of Knowledge?” In *Profiling the European Citizen: Cross-Disciplinary Perspectives*, edited by Mireille Hildebrandt and Serge Gutwirth, 17-29. New York, NY: Springer. http://ndl.ethernet.edu.et/bitstream/123456789/72212/1/123.pdf.pdf#page=44.

Kukutai, Tahu and John Taylor. 2016. “Data sovereignty for indigenous peoples: current practice and future needs.” In *Indigenous Data Sovereignty: Toward an Agenda*, 1-22. Canberra, Australia: Australian National University Press.

Mittelstadt, Brent. 2017. “From Individual to Group Privacy in Big Data Analytics.” *Philosophy & Technology* 30: 475-94. https://doi.org/10.1007/s13347-017-0253-7.

Paprica, P. Alison, Magda Nunes de Melo and Michael J. Schull. 2019. “Social licence and the general public’s attitudes toward research based on linked administrative health data: a qualitative study.” *CMAJOpen* 7 (1): E40-E46. https://doi.org/10.9778/cmajo.20180099.

Scassa, Teresa. 2020. “Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto.” *Technology and Regulation* 2020: 44-56. https://doi.org/10.26116/techreg.2020.005.

——— 2023. “Governing Human-Derived Data”/ « La gouvernance des données d’origine humaine ». In *Platform Governance in Canada/La gouvernance des plateformes au Canada*, edited by Heidi Tworek and Taylor Owen. Centre for the Study of Democratic Institutions and Centre for Media, Technology and Democracy. www.mediatechdemocracy.com/all-work/governing-human-derived-data.

Scassa, Teresa, Pamela J. Robinson and Ryan Mosoff. 2022. “The Datafication of Wastewater: Legal, Ethical and Civic Considerations.” *Technology and Regulation* (2022): 23-35. https://doi.org/10.26116/techreg.2022.003.

About the Author

Teresa Scassa is a CIGI senior fellow. She is also the Canada Research Chair in Information Law and Policy and a full professor at the University of Ottawa’s Faculty of Law, where her groundbreaking research explores issues of data ownership and control.

Teresa is an award-winning scholar, and is the author and editor of seven books, and more than 90 peer-reviewed articles and book chapters. She has a track record of interdisciplinary collaboration to solve complex problems of law and data. Teresa is a founding member of the University of Ottawa’s Centre for Law, Technology and Society, is cross-appointed to the School of Information Studies at the University of Ottawa, and is a member of the Geomatics and Cartographic Research Centre at Carleton University.

At CIGI, Teresa contributes expertise on the legal challenges associated with data ownership and the need for a national data strategy in a data-driven economy. Teresa also examines the governance of smart cities data and its implications for innovation, transparency, accountability, sovereignty and privacy.

Teresa holds degrees in civil and common law from McGill University, as well as an LL.M. and a doctorate from the University of Michigan. She clerked for Mme. Justice Claire L’Heureux-Dubé at the Supreme Court of Canada from 1988 to 1989.

Data as Representation: Fiduciary Models as Relational Valuation Frameworks

Sean Martin McDonald

There are a lot of ways to understand the value of data, economic and otherwise, and they vary substantially in both method and purpose (see, for example, Beauvisage and Mellet 2020; Birch, Cochrane and Ward 2021; Parsons and Viljoen 2024). This essay does not try to start from the process of valuation. Instead, it reverse engineers the way we assign value to data in context and identifies two critical observations. First, that data's primary value comes from its use in a particular context, when it is used as a representation; and second, that the value of data as a representation is conditioned, if not determined, by the supply chain of relationships that connect its creation to its use — and, critically, by the appropriateness of the parties using data to act as representatives of its subjects in that context. In other words, the value of using data as a representation is significantly determined by whether the user is an appropriate representative.

This essay proposes an approach that focuses on: the articulation and limitation of animating purpose — essentially, the rubric for determining the legitimacy of use in a context; the relationships between the scope of representation, standards of expertise and care, and boundaries of available representative actions; and the responsibility to support, if not provide, means of independent oversight and accountability. These broad dynamics will not address a commodifying approach to data governance — rather, they will provide ways to contextually assess the value and risks of the different approaches to building data supply chains, based on the highest-integrity models implemented in relevant contexts.

Data as a Representation, by a Relationship, in a Context

This analysis starts from a few assumptions, worth articulating upfront — the first and most important of which is that the primary purpose of data use is as information to inform a decision. Data is never the product of immaculate conception; it is created, transformed, mobilized and reconfigured at specific sites, each featuring particular actors, using particular instruments and for specific purposes (see, for example, Baker and Millerand 2007; D'Ignazio and Klein 2020; Leonelli and Tempini 2020).

All data is not equal; what makes some data more valuable than other data in a particular context is the perceived quality of the supply chain by which the data is produced and through which it travels. The more sensitive the context, the more important it is for decision makers to have confidence in the data they use as representations, which, critically, includes confidence in the legitimacy, technical capacity and accountability of the supply chain that produces it.

However, the entire premise of "big data" as a novel form of knowledge production is predicated on the mobility and reuse of data beyond its context of origin (Bates, Lin and Goodale 2016; Borgman, Scharnhorst and Golshan 2019; Thylstrup et al. 2022). When data is then concatenated, munged or otherwise combined, knowledge of its embedded limitations and subjectiveness is frequently irreversibly lost (Benjamin 2019; Bowker 2005; Chun and Barnett 2021). This digital political economy has previously been referred to as "the supply chain shredder" (Gansky and McDonald 2022). When industries for whom representational integrity is paramount are examined, different structures and practices of accountability and data maintenance are observed. Rooting the valuation of data in use has a number of important, secondary effects — importantly, it moves from a universal abstract directly into focusing on the value of the underlying decision, the impact of the data representation on the decision and the contextualizing role of the interests of the representative.

This section takes fitness for high-impact and high-value use as a framing assumption for the analysis and uses existing models for high-integrity representation relationships to reverse engineer the core characteristics and operational requirements for valuing data supply chains.

Data Is Not Fungible

The fundamental difference between data and other units of exchange is that data is information that is, predominantly, used in a specific context to affect a specific decision. The reason that the value of data is non-fungible is because the decisions that data influences are not fungible. The value of that information cannot be separated from its relationship to the context and role of its use and, as a result, cannot be meaningfully abstracted to a unit that standardizes that value acontextually.

This basic fact has in recent years been somewhat confused by the popularization of discursive framings of data (for example, as oil, sand or plutonium; cf. Doctorow 2008; O'Reilly 2021), which metaphorize data as a fungible market commodity — and which is partly responsible for engendering a wave of failed consumer-to-business data projects (Beauvisage and Mellet 2020) — to say nothing of the misapprehension of data as "personal" rather than social (for a historical account of this discourse, see Igo [2018]; for a rigorous theoretical account, see Viljoen [2021] and Parsons and Viljoen [2024]). Critiques of this kind of data-as-commodity thinking have more closely examined the actual practices of data-entangled corporations (Birch et al. 2021) and non-commercial

projects alike (Vertesi and Dourish 2011) to demonstrate that sets of data are in fact highly differentiated and their exchanges conditioned by cultural and infrastructural, as well as economic, factors. Said more plainly, data is not fungible, regardless of total volume, because the decisions data is meant to influence are not fungible, nor are the supply chains that meet the requirements of high-integrity, use-based markets.

Data as a Representation

This analysis starts from the "endpoint" of data use in order to focus on the characteristics of the underlying relationships, the rights holders' relationship to the context of use and the ways in which the digital transformation of acts of representation can and does impact their value — as opposed to attempting to derive contextual significance from the characteristics of data. While data may be valuable as a representation because of its "truth-value," the contextual criteria for "truth" in one context is often different from another (Gressin 2023). Understanding the value of data use as a representation requires examining both the "technical" quality of data and its claims (toward usability and veracity) and the quality of the supply chain of relationships that produce it (Ferryman 2017; Viljoen 2021).

The integrity of a relationship is not just based on the two people involved, it is also based on the relationship of each party to the context in which it happens. You may have totally appropriate relationships with your doctor and your lawyer, for example, but that does not mean it is appropriate for your doctor to represent you in court or your lawyer to make decisions about your medical care. The appropriateness of any use of data, or exchange of data, is related to whether the people involved are fulfilling their expected and aligned role within the context of their relationship to the data subject. In most high-impact contexts, we limit who gets to act on behalf of another person — typically, we require representatives to be a certified expert in the subject matter, as well as to have a direct, individual accountability to the person being impacted.

The specific qualities of the assertions made by a given data set or data stream are conditioned by these supply chains. Over the past two decades, scholars have developed a body of evidence and theory converging on a number of shared propositions regarding the politics and epistemology of data practices and infrastructures. This is often a necessary step in achieving the purposes for which the data is intended (Edwards 2010). In other words, when attempting to value data, the question of whether the person or organization using the data is an appropriate representative of the data subjects in that situation is as, if not more, important than its technical or substantive characteristics.

Data Use as an Act of Representation

One of the primary differences between the existence of data and its use as a representation, and especially a digital representation, is that the latter acknowledges the context, the intended impact and the associated liabilities for that representation. Data has created an explosion in the kinds of behaviours that can be observed, as well as increasing the number of actors and contexts implicated by the creation and use of data. The exchange of data raises questions not only about the validity of the facts asserted, but also about how and why the parties exchanging those facts are the appropriate actors to be doing so. The role and interests of the representative are fundamental, and categorically undervalued, where not outright ignored, in mapping data supply chains and economies. Perhaps the most clarifying advantage of framing a data valuation through the context of representation is that it starts from the recognition that representations, especially those made on the behalf of others, require a legitimate basis. People are not entitled to represent you — especially in high-impact, rights-affecting contexts — simply because they purport to hold information about you.

But we do not always apply the same limitations to the use and sharing of data, even when the data in question comes from regulated relationship models. In particular, duty-bearing professions — those that are regulated by public and private institutions — provide models for the way that we might govern digital representation relationships; they are realized by institutionally regulated, tangible, operational infrastructures designed to ensure the integrity, equity and symmetry of power in inherently asymmetrical representation relationships (Balkin 2020; Richards and Hartzog 2015, 2021). At a basic level, for example, for any data used by a fiduciary representative to meet the standards created by their duties, the production supply chain needs to be both explicit and accessible (Gansky and McDonald 2022).

While there is a significant range of practice, both within and between duty-bearing professions, there are common governance design patterns that offer valuable guidance for those attempting to design integrity measures for data and digitally intermediated relationships. The role of a fiduciary representative is to represent another person's, or group's interests in a defined context. To be a fiduciary representative, a professional must be able to understand their client's interests, triangulate the data and resources available to advance those interests and be able to describe their representation to both the client and the decision-making context. In other words, in order for data to be suitable as a representation, by a representative, in a high-value context, the data itself not only needs to be fit for purpose, but it also has to come from a representative source with a legitimate basis to make that assertion.

Courts do not allow, for example, anyone to wander in and, with no relationship to the parties, the court or the subject matter, make argumentation or submit evidence. That is not because we assume that people will not try, it is because the physical, procedural and practical design of legal systems makes it difficult to do so. The integrity of the representatives and representations made in rights-affecting contexts is protected by the context of use, not by the supply chain of production. The model of relationship designed for high-impact situations — especially across power asymmetry — is called a fiduciary relationship. The value of a fiduciary relationship is almost exclusively predicated on how well the representative understands and pursues the best interests of the person they are representing. The core characteristics of fiduciary relationships are a blueprint of the relational requirements for data supply chains that lead to data use as an act of representation in high-value contexts (and thus markets); they are also useful as a foundational framework for identifying the characteristics of data's value as a representation.

Fiduciary Models as Valuation Frameworks

The term "fiduciary" can seem nebulous or abstract, but in very concrete terms, it is a legal term for relationships where one person represents another's interests. While there is a lot of contextual variance in application, the design, oversight and enforcement of fiduciary relationships highlight a number of core elements of a high-integrity representation relationship — as well as the appropriate and legitimate basis for using representations to make high-impact decisions. The highest-impact decisions are often the most valuable — consider how every major tech company has tried, and mostly failed, to enter medicine and medical informatics (Foley 2019; Garcia 2019; Lomas 2022). Participation in high-value decisions, especially those influenced by data- and computation-intensive processes, is valuable but also difficult to evaluate independent of a wide range of subjective and contextual factors.

Fiduciary models rely on three things: duties of care, duties of loyalty and independent oversight. Duties are different than standards — they require active, case-by-case consideration in ways that do not appeal to universalizable rules that abstract to the

technical level/layer. Being a fiduciary representative is more art than science, but the duties that fiduciaries fulfill offer a functional model and set of system requirements that can be used to evaluate the integrity, quality and, thus, perhaps, the value of a data supply chain as a representative relationship and data use as an act of representation.

Representative Purpose Limitation: Defining and Reverse Engineering Value from Context

Perhaps the single most important characteristic of fiduciary representation as a model for data is that it both recognizes the role of "interests" and compels those involved to clearly articulate, limit and be held accountable to achieving those interests on behalf of a specific person. That liability means that while fiduciaries must know the interests of the people they represent, they also need to have enough information about the context, and the tools they use in representing those interests, to be able to explain how and why they made the decisions they did. One measure of data's value, especially relative to its fitness in high-impact contexts, is the degree to which its technical and legal format supports understanding its relationship to the interests of those involved in its production.

Relationship Definitions, Agency and Limitation

Another critical, if counterintuitive, difference between data and representations is that data is often produced and designed in order to maximize reuse, whereas representations are specifically designed for a specific context — and conducted inside the bounds of a defined, limited relationship. Fiduciary representatives have explicitly defined responsibilities that are limited in a range of common ways — for example, for a fixed period of time or related to a specific subject matter. As a vehicle for establishing the value of data — especially in the context of making a contextual representation, use-based limitations are often an indicator of specialization, fitness for purpose and, as a result, value.

Direct and Independent Governance and Oversight

One of the greatest indicators of integrity in any system is that the producers do not ask you to take their word for it — they make it easy for you to hold them to their word. Most data and digital systems are architected, whether by virtue of dependence or as a proactive means of arbitrage, in ways that explicitly avoid liability (for example, through disclaiming warranties, using open licensing and publishing to avoid transactional liabilities, and/or working in ambiguously defined jurisdictions). And yet, in order for a fiduciary to be able to explain and justify their use of data in a particular context, they need to be able to actively resolve disputes arising from its use — meaning they need to be able to identify the relevant parties and the relevant dispute resolution system and be able to compel the parties involved to accept the decision of that body. In other words, in order for data to be fit for purpose as a fiduciary representation, that data also needs to be transparently governed by an explicitly articulated, relevant authority.

Ultimately, the value of data is conditioned by the integrity, accessibility and ongoing oversight of the supply chains that produce and mobilize data from their point of origin to their use as representations to influence decision making. The particularities of what makes such supply chains fit for purpose is dependent on the context in which data as representations are articulated. Fiduciary models are by no means a universal or perfect solution (Khan and Pozen 2019). They do, however, provide an additional vehicle and mechanism for rights holders to participate in the governance and oversight of those that represent their interests. The characteristics of that governance and oversight

vary, but they provide for explicit reporting structures, time- and context-bounded relationship definitions, authority definitions, continuous burdens of reporting and proof, awareness and conflicts of interest, and mechanisms for administering disputes and contests.

Conclusion

We cannot regulate or establish value for data as a fungible object because, as an assertion of fact informing a decision, data is differently significant and valuable depending on the context in which it is used. We can, however, use the functional requirements of fiduciary relations as both a framework to understand the value of data (as representations) in context and the characteristic requirements for data production systems to maximize the opportunity for contextual value. Understanding the value of data as assertions informing high-impact decisions requires aligning our data valuation frameworks toward the quality and integrity indicators that representatives (for example, lawyers, medical professionals and accountants) use to manage high-impact supply chains.

The mapping, definition and bounding of the constituent interests embedded in data supply chains — from the situated perspective of the representatives using data to inform decisions impacting individuals and populations — is a relatively novel economic, legal and operational project. For the fiduciary, the value of data is predicated on the extent to which it enables them to make representations that adhere to their duties of loyalty and care, under the oversight of independent forms of accountability. This valuation logic is instructive; it points away from standardizable, universalizable valuations, toward granular, situated and justifiable understandings of data's value, directly linked to the supply chains that produce them.

Works Cited

Baker, Karen S. and Florence Millerand. 2007. "Articulation Work Supporting Information Infrastructure Design: Coordination, Categorization, and Assessment in Practice." *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, 242a-242a. https://doi.org/10.1109/HICSS.2007.88.

Balkin, Jack M. 2020. "The Fiduciary Model of Privacy." SSRN, November 18. https://papers.ssrn.com/abstract=3700087.

Bates, Jo, Yu-Wie Lin and Paula Goodale. 2016. "Data journeys: Capturing the socio-material constitution of data objects and flows." *Big Data & Society* 3 (2). https://doi.org/10.1177/2053951716654502.

Beauvisage, Thomas and Kevin Mellet. 2020. "Datassets: Assetizing and Marketizing Personal Data." In *Assetization: Turning Things into Assets in Technoscientific Capitalism*, edited by Kean Birch and Fabian Muniesa. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/12075.001.0001.

Benjamin, Ruha. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK: Polity Press.

Birch, Kean, D. T. Cochrane and Callum Ward. 2021. "Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech." *Big Data & Society* 8 (1). https://doi.org/10.1177/20539517211017308.

Borgman, Christine L., Andrea Scharnhorst and Milena Golshan. 2019. "Digital data archives as knowledge infrastructures: Mediating data sharing and reuse." *Journal of the Association for Information Science and Technology* 70 (8): 888-904. https://doi.org/10.1002/asi.24172.

Bowker, Geoffrey C. 2005. *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

Chun, Wendy Hui Kyong and Alex Barnett. 2021. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge, MA: MIT Press.

D'Ignazio, Catherine and Laura F. Klein. 2020. *Data Feminism*. Cambridge, MA: MIT Press.

Doctorow, Cory. 2008. "Personal data is as hot as nuclear waste." *The Guardian*, January 15. www.theguardian.com/technology/2008/jan/15/data.security.

Edwards, Paul N. 2010. *A Vast Machine: Computer Models, Climate data, and the Politics of Global Warming*. Cambridge, MA: MIT Press.

Ferryman, Kadija. 2017. "Reframing Data as a Gift." SSRN, July 22. https://doi.org/10.2139/ssrn.3000631.

Foley, Mary Jo. 2019. "Microsoft is closing its HealthVault patient-records service on November 20." ZDNET, April 5. www.zdnet.com/article/microsoft-is-closing-its-healthvault-patient-records-service-on-november-20/.

Garcia, Ahiza. 2019. "Google's 'Project Nightingale' center of federal inquiry." CNN Business, November 15. www.cnn.com/2019/11/12/tech/google-project-nightingale-federal-inquiry/index.html.

Gansky, Ben L. and Sean M. McDonald. 2022. "CounterFAccTual: How FAccT Undermines Its Organizing Principles." In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, 1982–92. https://doi.org/10.1145/3531146.3533241.

Gressin, Seena. 2023. "FTC lawsuit insists on FCRA compliance and transparency from background report providers." *Federal Trade Commission Business Blog*, September 11. www.ftc.gov/business-guidance/blog/2023/09/ftc-lawsuit-insists-fcra-compliance-transparency-background-report-providers.

Igo, Sarah E. 2018. "Me and My Data." *Historical Studies in the Natural Sciences* 48 (5): 616–26. https://doi.org/10.1525/hsns.2018.48.5.616.

Khan, Lina M. and David E. Pozen. 2019. "A Skeptical View of Information Fiduciaries." *Harvard Law Review* 133 (2): 497–541. https://harvardlawreview.org/print/vol-133/a-skeptical-view-of-information-fiduciaries/.

Leonelli, Sabina and Niccolò Tempini, eds. 2020. *Data Journeys in the Sciences*. Cham, Switzerland: Springer International. https://doi.org/10.1007/978-3-030-37177-7.

Lomas, Natasha. 2022. "Google faces new suit over DeepMind NHS patient data scandal." TechCrunch, May 16. https://techcrunch.com/2022/05/16/google-deepmind-nhs-misuse-of-private-data-lawsuit/.

O'Reilly, Tim. 2021. "Data Is the New Sand." The Information, February 24. www.theinformation.com/articles/data-is-the-new-sand.

Parsons, Amanda and Salomé Viljoen. 2024. "Valuing Social Data." *Columbia Law Review* 124: 993–1079. https://scholar.law.colorado.edu/faculty-articles/1651/.

Richards, Neil M. and Woodrow Hartzog. 2015. "Taking Trust Seriously in Privacy Law." SSRN, September 5. https://doi.org/10.2139/ssrn.2655719.

———. 2021. "A Duty of Loyalty for Privacy Law." *Washington University Law Review* 99: 961–1021. https://doi.org/10.2139/ssrn.3642217.

Thylstrup, Nanna Bonde, Kristian Bondo Hansen, Mikkel Flyverbom and Louise Amoore. 2022. "Politics of data reuse in machine learning systems: Theorizing reuse entanglements." *Big Data & Society* 9 (2). https://doi.org/10.1177/20539517221139785.

Vertesi, Janet and Paul Dourish. 2011. "The value of data: Considering the context of production in data economies." *CSCW '11: Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, 533–42. https://doi.org/10.1145/1958824.1958906.

Viljoen, Salomé. 2021. "A Relational Theory of Data Governance." *Yale Law Journal* 131 (2): 573–654.

About the Author

Sean Martin McDonald is the co-founder of Digital Public, which builds legal trusts to protect and govern digital assets. He is a lawyer and the CEO of FrontlineSMS, an award-winning global technology social enterprise; a fellow at the Duke Center on Law & Technology; a visiting fellow at Stanford University's Digital Civil Society Lab; and a former affiliate at Harvard University's Berkman Klein Center.

Sean is an adviser to Digital Democracy and the Ethics and AI Committee of the Institute of Electrical and Electronics Engineers, and a researcher and writer whose work has been published by the *International Review of the Red Cross*, *Foreign Policy*, *Stanford Social Innovation Review*, Cornell University's Legal Informatics Institute, IRIN and *Innovations*, among others.

He holds a J.D./M.A. from American University, with specialization in international law and alternative dispute resolution, and is a member of the New York State Bar Association. Sean's research focuses on civic data trusts as vehicles that embed public interest governance into digital relationships and markets.

Building a Data Wealth Fund

Kean Birch

As digital data has become more important in our technoscientific economies, it is increasingly understood and treated as an economic asset that holds significant value for businesses, governments and citizens (Birch 2023). More than a decade ago, for example, the World Economic Forum (2011) specifically characterized personal data as a "new asset class." Since then, it is possible to trace how data has been reframed and transformed into an economic asset. Despite the ongoing assetization of data, however, it is still difficult to define and unpack the economic and social value of data, which raises a number of social and policy implications for data governance.

Despite these difficulties, there has been considerable interest over the last few years in standardizing the economic understanding, treatment and governance of digital data. Several international institutions, including the United Nations, the World Bank, the International Monetary Fund, the Organisation for Economic Co-operation and Development and the European Commission, are revising the System of National Accounts (SNA) to incorporate data and its value into national economic indicators. The SNA is a set of national accounting standards used by countries around the world to standardize and harmonize their economic statistics and forecasts. These standards configure how countries define and estimate indicators, such as GDP and capital investment. A major revision to the SNA will be released in 2025, and it will include a new standard for the treatment of digital data as an asset (Mitchell and Lesher 2021).

As a result of these international standardization efforts, it is an opportune time to consider new policy initiatives, mechanisms and instruments for governing digital data. Countries will increasingly treat data as an economic and strategic asset, and this will have knock-on effects across our economies and societies as policy makers, businesses and citizens rethink their understandings and treatments of data and the economic and social benefits from its collection and analysis. These changes will inevitably lead to growing demand for transparency around who controls and benefits from data, especially personal data; for example, businesses will need to work out how to record data on their balance sheets — something they have not had to do to date (Birch, Cochrane and Ward 2021). Consequently, it will become clearer where data sits, who controls access to it and who benefits from that control; at the same time, this increased transparency will lead to growing demands for changes to the way that data is governed.

Data as an Economic Object

Digital data is an economic object that has value. But data does not fit neatly into prevailing definitions of goods or services. Rather, data is increasingly conceptualized and framed as a resource, and specifically as an economic asset (Birch, Cochrane and Ward 2021; Birch, Marquis and Silva 2024). For example, the Government of Ontario has an ongoing strategy to build a "Digital Ontario" on this basis, stating: "Data has become one of the world's most valuable assets. In 2018 alone, it added over $150 billion in value to Canada's economy. Data powers Ontario's government, its businesses, and its communities."[1] The framing of data as an asset is evident across a range of strategies, discussions and tool kits developed by various social actors, including businesses, policy makers and civil society organizations (see Box 1, for example).

Box 1: First Nations Information Governance Centre

The First Nations Information Governance Centre published a data governance "strategy" in 2018 with the following aim: "Our Vision: A First Nations-led, national network of modern information and statistical service centres at national and regional levels, to serve the data capacity needs of communities and Nations and to advance the realization of data sovereignty that is in alignment with First Nations' distinct worldviews" (First Nations Information Governance Centre 2018, 6).

The strategy is based on several data stewardship principles, including empowering "evidence-based decision-making," closing "data gaps," improving "services to First Nations," transferring "government services back into the hands of rights holders," supporting "self-determination and self-governance," and increasing "fiscal capacities" through data (ibid.).

The strategy outlines how data is framed and can be treated as a strategic and valuable asset that must be managed by rights holders, where such data assets include "cultural and traditional information, administrative information, personal information, and information regarding the territory, resources, and environment" (ibid., 41).

Defining digital data as an economic asset entails thinking about data's particular characteristics (Coyle et al. 2020; Mitchell and Lesher 2021; Birch 2023). Data is characterized as a non-rivalrous good, in that it can be used by several people at once. However, data is still excludable, in that data's collection, storage and use can entail significant financial investment, which precludes its collection and use by everyone. Moreover, as facts, data is not subject to conventional intellectual property (IP) regimes, but businesses can still treat data as an asset by controlling access to it (Cohen 2019). Finally, data assets have emergent properties, in that their use and value are an effect of data's aggregation and the relationship between multiple data points; consequently, their usefulness and value are more than a simple sum of their parts (Esayas 2017; Viljoen 2021).

These characteristics mean that it is difficult to calculate and measure the value of data as an asset. International economic standards setters, such as the SNA, have been developing an approach based on a sum-of-costs method, which defines the data value as the cost of its production (for example, labour costs, equipment costs, and so forth). An important assumption underpinning the SNA's treatment of data is that data is a "produced asset," meaning that data will be treated as if it is the result of organizational decisions akin to a range of other intangible assets (for example, IP) (Mitchell and Lesher

2021). However, these conceptual and measurement discussions can miss important conceptual and policy implications of data's characteristics. As Salomé Viljoen (2021) notes, the value of data comes from its relational and collective configuration; that is, data has economic and social value precisely when it is brought together in collective data holdings, combining data from multiple people or organizations to generate new inferential and other insights. These insights depend upon data's emergent properties, in that the combination of data creates outcomes or outputs that cannot be foreseen by a summing of the parts.

As an asset, digital data is consequently difficult to value and to govern. It is increasingly evident that current data governance regimes entail a series of contradictions, even paradoxes, when it comes to the use and value of data. First, data is an important and valuable resource for the twenty-first century, but there is currently little agreement on how to measure or calculate its value. Second, data has considerable social value, especially because of its non-rivalrous qualities, which means diverse data sets can be combined to stimulate widespread innovation; however, data governance regimes do not stop organizations or individuals from making data excludable, thereby limiting innovation and ensuring that only a few entities (for example, big tech) benefit from these combinations. Last, the wider social benefits and value of data depend upon sharing it, but the economic value of data depends upon restricting access to it (Birch 2023).

We need a new data governance regime to address these paradoxes and ensure that the benefits of digital data are shared by everyone rather than primarily by a few multinational businesses.

Data Governance

Debates about digital data governance have been ongoing for a while and cover an array of possible approaches (Micheli et al. 2020). The reason for this is because data is increasingly identified as a valuable asset that a few large multinational businesses — usually defined as big tech — have claimed for free, despite data being the collective product or output of everyone's digital activities (for example, using smartphones, using social media, using search engines). Yanis Varoufakis (2024), the ex-finance minister of Greece and heterodox economics professor, calls this a system of "technofeudalism" in which most people labour away as "cloud serfs" generating the resources and assets that "cloud capitalists" take to generate their revenues. A less poetic way to think about this is to consider personal data as an economic rent that businesses extract from our lives, behaviours and decisions (Birch, Chiappetta and Artyushina 2020).

Data governance approaches have been developed to address these concerns. Such governance approaches can be split into three main philosophical camps. Viljoen (2020) defines two camps as "propertarian" and "dignitarian," while Barbara Prainsack (2019) defines a third camp as "solidaristic." Propertarians include thinkers who want to extend individual property rights to data, especially personal data. According to these thinkers, data could be treated as IP that can be claimed by the identifiable person it relates to (Lanier 2014), or data could be treated as the product of people's work (Posner and Weyl 2019), harking back to the seventeenth-century philosopher Thomas Hobbes. Propertarians think that individuals should gain direct personal benefit from their data. Dignitarians include thinkers who consider personal data to be a human right and therefore want to limit its commodification and assetization in order to preserve privacy and democracy (Zuboff 2019). While the propertarian and dignitarian camps are certainly noteworthy, providing helpful insights into current data governance regimes, they both tend to naturalize individualistic framings of data governance.

The third, solidaristic, camp offers a more optimistic and compatible way to rethink and restructure data governance in light of data's increasing treatment as an asset, especially as a collectively generated resource. As Prainsack (2019) explains, a good starting point for understanding data is to consider it like other "common" resources; commons can be easily undermined when we rely upon individualistic governance mechanisms (for example, markets, individual rights) in which there are serious and significant power asymmetries between social actors (for example, between an individual and a multinational). Instead, thinking of data as a common resource highlights the need to create an individual and collective governance mechanism, in order to protect individual and collective rights and the social benefits that accrue from the collective sharing of data. Solidaristic politics can support both the sharing of data and the economic and social benefits that sharing brings, as well as ensuring that those benefits are distributed equitably.

It is vital to find the right organizational structures and mechanisms to support the solidaristic understanding of data governance. There are several options proposed by a range of different social actors. However, the proposal in this essay is for the establishment of a data wealth fund (DWF) that combines the treatment of data as an asset with the solidaristic politics of understanding data as a collectively generated and commonly held resource.

DWF

There is a range of digital data governance models that fit with this solidaristic position, many of which have been developed over the last few years in response to growing concerns about the collection, use and exploitation of data by multinational businesses (i.e., big tech). Examples of these data governance models are outlined in Table 1, which also includes some of the limitations of each model.

Table 1: Data Governance Models

Model	Outline	Limitations
Data-sharing pools	Partnership and sharing agreements between social actors in pursuit of economic ends.	Treats data as a commodity that is shared for private profit, and data subjects are excluded from arrangement.
Data trusts	Entity that oversees access to data it manages, providing access to a range of social actors.	Does not stop powerful social actors (for example, big tech) from accessing data and depends on trustees to provide oversight.
Data cooperatives	Entity and data-sharing arrangement in which participating social actors and data subjects retain control over data.	Hard to scale and manage privacy and other concerns of data subjects.

Source: Micheli et al. (2020); Ali, Munnelly and Wolf (2023); Birch (2023).

It is important to consider alternative models of data governance in light of these limitations. One alternative, which reflects the solidaristic idea of data assets as collectively generated and commonly held resources (Prainsack 2019), is the concept of a DWF. A DWF would copy some of the organizational operations and structure of other resource funds, such as oil and gas funds run by national governments. An example is Norway's "Oil Fund," which was established in 1990 as part of a strategy to diversify Norway's economy and make it more resilient to changing oil and gas prices, as well as

to other global economic shocks. The Oil Fund currently represents more than $2 trillion in a portfolio of assets from around the world. The fund receives billions of dollars each year from taxation, fees and government ownership shares in the oil and gas fields.

The concept of a DWF is based on an understanding of digital data as a collectively generated and commonly held resource or asset, in that data represents aggregated information about a population that the population produces through its actions, and the data has value precisely because of its relational and emergent qualities directly derived from that population. A DWF is a governance model that ensures a country's data is used in ways that its residents want and derives a share of the value that the data generates for its users, especially for businesses. While copying other wealth funds, the DWF would require new organizational operations and structures to support the underlying principles for any data governance framework. These underlying principles could be:

- to protect the privacy of a country's residents by coordinating and managing access to and the use of the country's data, including government, health and personal data;
- to encourage the sharing of a country's data in support of digital innovation across business, government and civil society;
- to ensure that a country's residents benefit from the sharing and use of a country's data through taxation, fees and ownership shares; and
- to hold users of a country's data to account for their use of data, especially through public engagement initiatives to establish what a country's residents would like their data to be used for.

Eventually, a DWF could become the data steward for a country's data (Cameron 2024), being responsible and accountable for the data collected about a country's population. It would be a national and publicly managed entity, independent of government but accountable to a country's residents through public engagement processes. It would collect, store and hold data, which could be accessed by different social actors on a differential fee basis, depending upon criteria such as the direct social benefits expected and the direct private benefits received. It would necessitate regulation requiring that any data-collecting entity must deposit said data in the DWF, with fines imposed if an entity fails to comply. Operationally, this requirement to deposit could follow a terms and conditions agreement requesting data, HTTP cookies or a similar declaration of data collection — that is, if an entity makes a request for data through these means (or others), then it would be required to deposit that data in the DWF. Through this requirement to deposit, any entity could be held accountable for data deposit.

Unlike other wealth funds, there are a range of distinct issues that any DWF would have to address in its operations and structures to be functional. A few issues that any policy maker would need to address would include:

- **Data heterogeneity:** Most wealth funds generate revenues from one or two well-known and often fungible non-produced assets (for example, oil and gas); however, digital data is often heterogeneous in origin and function. A DWF would need to ensure that it has a clear definition of digital data and that it has a specific set of requirements for diverse entities (for example, businesses, hospitals, governments, civil society organizations, and so forth) on what data they are required to deposit. This might mean that internal operational data is not covered by the requirement to deposit unless the data includes personal information.
- **Policy intervention:** Digital data is different in that it is increasingly framed as a produced asset, resulting from the action of specific entities generating it. A DWF would need to ensure that the digital observing, recording and storing of information (as data) is the focus of policy intervention, rather than the produced asset itself.

- **Administrative entity:** Digital data is collected and aggregated by a range of public administrative entities such as national statistical offices (NSOs). A DWF could be built out from an existing NSO since these entities have existing operational capacities to collect data, permit access and disseminate according to established principles.
- **Tensions with open data:** As strategic assets, governments and policy makers increasingly seek to open up public data holdings. A DWF would have to balance the societal benefits of open data versus charging for access to its data holdings.

From 2025, data will be increasingly treated as an economic asset, and this shifting attitude will ripple across our economies and societies. Although it is difficult to predict the effects of this change, it is still important to consider its implications for data governance. In particular, it is clear that the importance of data will only grow, in light of the ongoing boom in interest in artificial intelligence technologies. How we handle data, then, is critical for societies. Returning to outdated and downright problematic governance frameworks based on notions of naturalized markets is not viable anymore; we need to take a more concerted, collective approach to data governance that can address the paradoxes emerging in the data economy. For the author, solidaristic principles underpinning a DWF are one option that tries to deal with these paradoxes and attempts to spread the benefits of the data economy beyond a few big tech firms and their investors.

Acknowledgements

Thanks to Bob Fay, Jenny Thiel and Andy Best for their feedback and advice. The research underpinning this essay was funded by the Social Sciences and Humanities Research Council of Canada (Ref. 435-2023-0704).

Note

1 See www.ontario.ca/page/building-digital-ontario.

Works Cited

Ali, Vinous, Carly Munnelly and Rachel Wolf. 2023. *Changing the Conversation: Canada's Model for 21st Century Data Governance*. Long Eaton, UK: Public First & Innovate Cities.

Birch, Kean. 2023. *Data Enclaves*. Cham, Switzerland: Palgrave Macmillan.

Birch, Kean, Margaret Chiappetta and Anna Artyushina. 2020. "The problem of innovation in technoscientific capitalism: data rentiership and the policy implications of turning personal digital data into a private asset." *Policy Studies* 41 (5): 468–87. https://doi.org/10.1080/01442872.2020.1748264.

Birch, Kean, D. Troy Cochrane and Callum Ward. 2021. "Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech." *Big Data & Society* 8 (1). https://doi.org/10.1177/20539517211017308.

Birch, Kean, Sarah Marquis and Guilherme Cavalcante Silva. 2024. "Understanding Data Valuation: Valuing Google's Data Assets." *IEEE Transactions on Technology and Society* 5 (2): 183–90. https://doi.org/10.1109/TTS.2024.3398400.

Cameron, Courtney. 2024. "Data Stewardship: The Way Forward in the New Digital Data Landscape." Opinion, Centre for International Governance Innovation, May 13. www.cigionline.org/articles/data-stewardship-the-way-forward-in-the-new-digital-data-landscape/.

Cohen, Julie E. 2019. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. Oxford, UK: Oxford University Press.

Coyle, Diane, Stephanie Diepeveen, Julia Wdowin, Lawrence Kay and Jeni Tennison. 2020. *The Value of Data: Policy Implications*. February. Cambridge, UK: Bennett Institute for Public Policy.

Coyle, Diane and Annabel Manley. 2021. "Potential social value from data: an application of discrete choice analysis." Working Paper. August 5. Cambridge, UK: Bennett Institute for Public Policy.

Esayas, Samson. 2017. "The idea of 'emergent properties' in data privacy: towards a holistic approach." *International Journal of Law and Information Technology* 25 (2): 139–78.

First Nations Information Governance Centre. 2018 (revised 2020). *A First Nations Data Governance Strategy*. Akwesasne, ON: First Nations Information Governance Centre. https://fnigc.ca/what-we-do/first-nations-data-governance-strategy/.

Lanier, Jaron. 2014. *Who Owns the Future*? New York, NY: Simon & Schuster.

Micheli, Marina, Marisa Ponti, Max Craglia and Anna Berti Suman. 2020. "Emerging models of data governance in the age of datafication." *Big Data & Society* 7 (2): 1-15. https://doi.org/10.1177/2053951720948087.

Mitchell, J. D. Ker and M. Lesher. 2021. "Measuring the economic value of data." *OECD Going Digital Toolkit Notes*, No. 20. Paris, France: OECD. www.oecd-ilibrary.org/science-and-technology/measuring-the-economic-value-of-data_f46b3691-en.

Posner, Eric A. and E. Glen Weyl. 2019. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton, NJ: Princeton University Press.

Prainsack, Barbara. 2019. "Logged out: Ownership, exclusion and public value in the digital data and information commons." *Big Data & Society* 6 (1). https://doi.org/10.1177/2053951719829773.

Sadowski, Jathan. 2019. "When data is capital: Datafication, accumulation, and extraction." *Big Data & Society* 6 (1). https://doi.org/10.1177/2053951718820549.

Varoufakis, Yanis. 2024. *Technofeudalism: What Killed Capitalism*. London, UK: Bodley Head.

Viljoen, Salomé. 2020. "Data as Property?" *Phenomenal World*, October 16. www.phenomenalworld.org/analysis/data-as-property/.

———. 2021. "A Relational Theory of Data Governance." *Yale Law Journal* 131 (2): 573–654.

World Economic Forum. 2011. *Personal Data: The Emergence of a New Asset Class*. Geneva, Switzerland: World Economic Forum.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism*. New York, NY: Public Affairs.

About the Author

Kean Birch is the director of the Institute for Technoscience & Society, Ontario Research Chair in Science Policy, and professor in the Science & Technology Studies (STS) Graduate Program and Department of Science, Technology & Society at York University, Toronto, Canada. He has been a visiting scholar at the Copenhagen Business School in Denmark and the Munich Center for Technology in Society, Technical University of Munich, in Germany.

Kean is the co-editor of the leading international STS journal *Science as Culture*, published by Taylor & Francis, and co-founder and series editor of the Technoscience and Society book series at University of Toronto Press. His most recent books are *Data Enclaves* published by Palgrave Macmillan and *Assetization: Turning Things into Assets in Technoscientific Capitalism*, co-edited with Fabian Muniesa and open-access published by MIT Press.

Competition and the Valuation of Data

Keldon Bester

Countries around the world are experiencing a competition moment. After decades of policy supporting consolidation and a lax approach to concentrated corporate power, the assumptions underlying this approach are being revisited (Bester 2022). In the United States, competition law authorities have stepped up enforcement of the country's anti-monopoly laws. In the European Union, the Digital Markets Act seeks to rein in the power of the giants of digital markets. In Canada, a series of reforms in 2023 and 2024 have dramatically strengthened the Competition Act, the country's competition policy law (Department of Finance Canada 2023, 2024).

But largely absent from the policy discussion preceding these responses has been the intersecting roles of competition and data governance and, in particular, the role that competition can and should play in the valuation of data (Iacobucci 2021). Already critical amid the rise of digital markets, the growing use of large language models and their voracious appetite for data makes the task of understanding the true value of data even more urgent.

Competition as a Diviner of Value

As a public policy goal, competition is understood as a means rather than an end. Sidestepping an important debate about the kinds of competition that are beneficial and detrimental to the economy, the results of competition are well understood: lower prices, higher quality (Bester 2023). In addition, the fuel for the innovative process that delivers new goods and services bears the fruits of an economic system that rewards challenges to the status quo in markets across the economy.

Less advertised and more relevant to the ongoing work to better understand the value of data is the role of competition in constructing the value of the interlocking components of the economy.

The concept of value is a core component of competition. The value of a good or service, its many dimensions, the interpretation of those dimensions by individual users and societies, and the variety of those interpretations make up the heart of the competitive process. Often framed in terms of price discovery, competitive markets are the primary route for the messy process of determining what something is worth. Constraints on competition, therefore, can undermine this process and thus obscure the value of data.

This can be understood narrowly in the metaphor of the market exchange, with many buyers and sellers coming together and comparing bids and asks, but the reality of value discovery through competition is a much richer process. One example is the evolving debate over the role of privacy in competition and accordingly the enforcement of competition laws. Once considered outside the scope of traditional competition law analysis, privacy is increasingly valued by market participants as a dimension on which commercial decisions could be based. As a result, competition law is forced to reckon with privacy as a dimension of competition and understand its relevance in the competitive process. The contrary position argues that privacy is a goal best handled by discrete policies, and that competition law should maintain its focus on just that, competition. But this position attempts to skirt competition's role as a method for divining the value of the attributes of products or services, and a law protecting competition must be flexible enough to incorporate these disparate dimensions of value.

Going a level deeper, a core motivation for competitive access to the inputs that drive the economy is the belief that a diversity of participants is the surest way to unlock not only the most efficient use of a given input, but also the full range of its possible uses. The idea that the competitive process will uncover answers in the aggregate that would have never been considered by even the commanding heights of the economy is a driver of the skepticism of monopoly power. Allowing monopolies to form bottlenecks at key points in the economic system has the potential not only to increase a wide definition of costs but also to blunt competition's power of discovery.

Together, the unimpeded process of refining and diversifying uses of the building blocks of the economy driven by competition arrives at the most accurate approximation of its true value. The use of the word "true" does not necessarily connote a single value, given the multiple dimensions of value possible when viewed through the prism of economic actors and their forever fluid nature. For example, the value of the underlying components of mobile phones, and of mobile phones themselves, has undergone multiple transformations. Their value was seemingly well understood until it was uprooted with the introduction of the path-breaking BlackBerry, only to be transformed again with the introduction and evolution of the iPhone. For each iteration, had that combination been frustrated, markets would have continued to labour under a current but incomplete view of the potential value of the product and its constituent parts.

This is the case for data, as it is for physical goods in a more traditional view of the economy. Looking at the companies that have come to dominate digital markets over the past 20 years, there is no question that data has tremendous value. At the time of writing, nearly all of the public companies valued at more than a trillion dollars have data central to their business models or provide the inputs for others to make use of data. But an input to economic activity can be extremely valuable while at the same time having the full extent of its value misunderstood. There exists a real risk that the monopolies that have grown amid the explosion in the use of data are frustrating that competitive process and are overdue for exposure to the enlightening power of competition.

The largest companies on the planet take up the lion's share of headlines related to competition and data, but this phenomenon extends to some of the most traditional markets of the economy. Consider the case of farmers, whose commercial activity generates not only food that keeps society fed but also constant streams of data about weather, soil quality, crop yield and the effectiveness of agricultural inputs. Today that data is often captured by major equipment manufacturers such as John Deere rather than the farmers themselves (if it is captured at all). This has implications for the more familiar boundaries of competition law, with access to data conditioned on locking farmers into a given equipment manufacturer's platform, possibly choking off competition in the market for farming implements. But the issue extends beyond this scope to the goal of understanding the value of data in a market core to human

flourishing. While a company such as John Deere has turned those data flows into a valuable revenue stream for itself, the true value of this data is obscured and likely discounted by pulling control away from individual farmers and locking it within a walled garden. While maintaining scarcity over this data may improve its financial value for a single actor, its true potential value and contribution to the economy and society is obscured.

Competition Law's Oblique Contributions to Data Valuation

The monopolization of the flows of data is not a new issue. In 2011, the Canadian Competition Bureau challenged the Toronto Real Estate Board's (TREB's) monopoly over real estate data in the city of Toronto. This was one of the most consequential cases of abuse of dominance under the provisions of the Competition Act and focused on anti-competitive activity by dominant firms. Hinging their case on an innovation-based theory of harm, the bureau successfully argued that TREB was foreclosing access to valuable real estate data, thereby suppressing the ability of other companies to enter the market.

TREB is an example of both the intersection and the gap between more traditional competition policy and the valuation of data. Though not an explicit component of the bureau's case, TREB is a prime example of the suppression of the true value of data in the hands of a monopolist. Even if TREB were, in theory, the optimal user of the data in question, the range of that data's uses would be necessarily bounded by the realities of the organization, the path dependency of its capacity and its limited and potentially conflicting business incentives. From a competition law perspective, the goals of the organization were misaligned with the more optimal outcome of the use of its underlying resources in a more open and competitive market. From a data valuation perspective, narrow control of the data was blunting the ability to understand the true potential value of the data to the economy and to society.

The US Department of Justice's (DOJ's) 2023 complaint against Google's dominance in the advertising technology (ad tech) market is another illustration of the importance of competition, not just as a distributor of value generated, but also as a tool for understanding the true value of data (US DOJ 2023). Much of the modern digital advertising infrastructure functions similar to a stock exchange. Demand and supply are worked out in an infinitely recurring series of automated auctions matching advertisers and publishers. Google is accused of establishing positions across both sides of the exchange market, as well as the exchange itself, and using those positions to tilt the outcomes to its own benefit. This conduct, which the DOJ suggests leads Google to claim 30 percent of every ad dollar spent through its platforms, comes at the cost of advertisers and publishers who are harmed by the monopolization of the advertising market. Following the lead of the DOJ, Canada's Competition Bureau has expanded its own investigation into Google's ad tech practices, including a predatory pricing theory of harm unavailable under American antitrust law (Competition Bureau Canada 2024).

Without focusing explicitly on the value of data, both of these cases speak to the power of competition to unearth a truer picture of value, as well as the power of monopoly to distort that picture. A company such as Google brings in myriad data flows through its services and funnels them toward a variety of purposes, many of which have provided value to a global base of users (Birch 2023). But the centralization of these flows within a single corporate entity represents a barrier to unlocking not only the potential of that data in a more open market, but also an understanding of the value of the data beyond the noisy headline figure of the corporation's market capitalization. The use of these

financial metrics to understand the value of data is frustrated, both because market capitalization is too crude a figure to extract any meaningful view and because the value itself is distorted when the control of data is monopolized. With the competitive process frustrated, the result is a crude, inflated and incomplete picture.

The sky-high valuations of these companies test the belief in the competitive process to ascertain value over centralized economic decision making. Setting aside competition law's inherent wariness of economic dominance, it is reasonable to assume that Google is best placed to decipher and deploy a deep understanding of the value of data as one of the world's largest companies. But, as in the case of TREB, a corporation such as Google is bound by its own capacity and incentives. For all its sprawling services and business lines, Google, as well as its parent company Alphabet, is first and foremost a digital advertising company. Accordingly, the use of data within the company ultimately serves the goal of success in that market and its value is distorted by the competitive moats encircling its business. As the company's mythology of risky bets outside its core competency fades into memory and the quality of that core competency itself appears to degrade, the consequences of monopoly in understanding and unlocking the true value of data become harder to ignore, making the task of opening these bottlenecks more urgent (Bevendorff et al. 2024).

Privacy and Control as the Foundation for Competition

With all the talk of openness, it is fair to worry that the use of competition as a tool to derive the true value of data implies a free-for-all with the often sensitive data that individuals and organizations generate. In the case of companies dominant in digital markets, the agreement frequently offered has been that privacy will be preserved as long as the data is the sole remit of the company in question. Scandals such as Facebook-Cambridge Analytica poked holes in this narrative, but alongside the Wild West of the data broker industry, these scandals have ultimately reinforced the idea that data should be entrusted to the chosen few. Returning to the TREB case, arguments for privacy were unsuccessfully marshalled in defence of the data monopoly, casting the organization as a responsible steward amid reckless potential competitors.

This view is incorrect. Rather than encouraging the careless use of data, the potential for competition as a method for sussing out the value of data reinforces the need for reliable privacy protection and control of the data generated by individuals and organizations. Though the nature of data and its generation, collection and use force new ways of thinking about control and ownership just as it does with the concepts of value and valuation, this does not mean the task can be forfeited. Nods to data portability and interoperability driven by the actors sitting on data chokepoints are unable to escape the incentives at the heart of their own business models. As a result, work to this end has often been little more than window dressing by monopolists trying to protect their own turf as opposed to efforts to truly spur broader access and competition. To realize the true potential of competition for the task of valuing data, policy makers and organizations seeking broader access to data have a mutual interest in creating reliable systems to support the responsible use of data.

To be successful, these approaches to privacy regulation must appreciate and reflect the value of competition. Echoing the motivation for skepticism of monopoly, a narrow reading of the potential uses of data will lead to narrow results. Compare the approaches taken by the United Kingdom and Australia in providing their citizens with systems for control of their data. The United Kingdom's open banking regime has given the British public greater control over the use of the data that makes up their financial lives, as well as increasing transparency and competition in an important market — the provision of high-quality financial services.

But the adoption and success of that system is limited by the boundaries of its ambition to improve competition in the financial sector. In contrast, the Australian approach to enacting a consumer data right was always anchored in a broad and evolving understanding of the potential uses of data in many sectors, even if the initial scope was limited. This has allowed organizations from a wider range of sectors to create services that provide value to consumers on the basis of the data they now control. Framed in the language of privacy and consumer choice, this wider conception of the potential uses of data implicitly adopts competition as a tool for better approximating the value of the data in question.

Conclusion

After decades of dormancy, competition policy has returned to the foreground amid the rise and persistence of digital market giants that are driven by business models based on the collection and utilization of vast flows of data. As competition policy turns to the more familiar question of addressing the economic dominance of these giants, the role that competition can and should play in better understanding the value of the data that has supported their business models has lacked discussion. While these firms have recognized fantastic economic value from the data they hold, the centralization of the control of that data runs counter to the promise of competition as a true diviner of value beyond the narrow performance and financial metrics of a given firm.

A pillar of the faith in the competitive process is that working out approaches through competition and contestation, as opposed to the genius or capability of a single actor, is more likely to reveal the true and evolving value of a given commodity. Attempts to discern the value of data emerging from a diverse range of actors — located in both the private and public spheres, as well as in academic and civil society — mimic this competitive process and its potential for enlightenment. But to fully realize that potential, the walled gardens frustrating that process must be brought down and replaced with contestation, both for the sake of markets and in order to understand the true value of their underlying components.

Works Cited

Bester, Keldon. 2022. *Merger Policy for a Dynamic and Digital Canadian Economy*. CIGI Paper No. 268. September. Waterloo, ON: CIGI. www.cigionline.org/publications/merger-policy-for-a-dynamic-and-digital-canadian-economy/.

– – –. 2023. *Fair Competition for an Evolving Economy*. CIGI Paper No. 284. September. Waterloo, ON: CIGI. www.cigionline.org/publications/fair-competition-for-an-evolving-economy/.

Bevendorff, Janek, Matti Wiegmann, Martin Potthast and Benno Stein. 2024. "Is Google Getting Worse? A Longitudinal Investigation of SEO Spam in Search Engines." In *Advances in Information Retrieval: Proceedings of the 46th European Conference on Information Retrieval, Part III*, edited by Nazli Goharian, Nicola Tonellotto, Yulan He, Aldo Lipani, Graham McDonald, Craig Macdonald and Iadh Ounis, 56–71. Cham, Switzerland: Springer Nature. https://downloads.webis.de/publications/papers/bevendorff_2024a.pdf.

Birch, Kean. 2023. *Data Enclaves*. Cham, Switzerland: Palgrave Macmillan.

Competition Bureau Canada. 2024. "Competition Bureau expands its investigation into Google's advertising practices." News release, February 29. www.canada.ca/en/competition-bureau/news/2024/02/competition-bureau-expands-its-investigation-into-googles-advertising-practices.html.

Department of Finance Canada. 2023. "*Affordable Housing and Groceries Act* receives Royal Assent to build more rental homes and help stabilize grocery prices." News release, December 15. www.canada.ca/en/department-finance/news/2023/12/affordable-housing-and-groceries-act-receives-royal-assent-to-build-more-rental-homes-and-help-stabilize-grocery-prices.html.

– – –. 2024. "Legislation to make life more affordable, build more homes, and strengthen economy for everyone receives Royal Assent." News release, June 20. www.canada.ca/en/department-finance/news/2024/06/legislation-to-make-life-more-affordable-build-more-homes-and-strengthen-economy-for-everyone-receives-royal-assent.html.

Iacobucci, Edward M. 2021, "Examining the Canadian *Competition Act* in the Digital Era." December 27. https://sencanada.ca/media/368377/examining-the-canadian-competition-act-in-the-digital-era-en-pdf.pdf.

US DOJ. 2023. "Justice Department Sues Google for Monopolizing Digital Advertising Technologies." Press release, January 24. www.justice.gov/opa/pr/justice-department-sues-google-monopolizing-digital-advertising-technologies.

About the Author

Keldon Bester is a CIGI fellow and the executive director of the Canadian Anti-Monopoly Project, a think tank dedicated to addressing the harms of monopoly and building a more democratic economy. He is a leading voice in Canada's competition policy conversation, and his writing and thinking have been featured in *The Globe and Mail*, *National Post* and *Toronto Star*.

Keldon has worked as a special adviser at Canada's Competition Bureau, as a fellow at the Open Markets Institute and as a consultant for organizations across the Canadian economy. He holds a master of public policy degree from the Harvard Kennedy School.

Digital Public Infrastructure: Orientation Matters

Soujanya Sridharan, Vinay Narayan and Jack Hardinges

Over the past two decades, messaging platforms, online marketplaces, app-based transport and digital payment systems have become intertwined with our daily lives. This rapid digitalization is characterized by private corporate ownership and control. Ride-hailing platforms have fundamentally altered public transportation in many parts of the globe. Uber alone has 3.5 million drivers operating across 10,000 cities (Duncan 2022). Of the six social media platforms that claim one billion or more monthly active users, three are owned by Meta.[1] The implications of this control surface from time to time. WhatsApp's outage in 2021 affected millions of people around the world, including citizens in Argentina and Lebanon whose governments were relying on the service to provide public health updates on the COVID-19 outbreak (Cheng 2021). The mass reliance on Google Pay, Visa and Mastercard to make payments was exposed when these companies suspended services in Russia in the wake of the country's war with Ukraine (Stognei and Fraser 2022).

Beyond platforms, the spectre of private ownership in digital infrastructure is also evident when we look to the cloud. Three private companies — Amazon, Microsoft and Google — control nearly two-thirds of the market share for cloud infrastructure (Statista 2024). Control over these systems often provides private entities with effective control over the data they generate. Insights derived from this data have significant potential for public good, but private companies often restrict — or outright deny — access to this data, citing various concerns such as privacy and business interests. Uber and Lyft, for example, have refused to share data with city authorities in the past (Austermuhle 2018). Given this influence, private ownership and control over digital infrastructures and the data within them are of significant concern.

In this context, the concept of digital public infrastructure (DPI) provides an alternative. India's Unified Payments Interface is an example of DPI. It is a government-backed payment system that facilitates instant interbank, peer-to-peer and person-to-merchant transactions. While private firms and their apps can plug into the system, they must use open protocols for exchanging information and enabling payments, and adhere

to the rules of the system set out by the Reserve Bank of India. The Unified Payments Interface has significantly broadened access to digital banking and has become the preferred mode of payment in India (*Business Standard* 2023), with more than 12 billion transactions recorded each month (TOI Tech Desk 2024). Its success has also seen India share the technology with other countries, including Australia, France, Saudi Arabia and Singapore.

DPI borrows lessons from nation-states' experience of building physical infrastructure (Eaves 2023), such as roads and railways. Just as physical infrastructure was considered crucial to the movement of people and goods in the twentieth century, so too is digital infrastructure, such as identity and payment systems, for contemporary society. And just as governments have had a critical role to play in building and maintaining physical infrastructure, proponents of DPI argue that an active state is needed to ensure that a nation runs smoothly and citizens' needs are met. Ethan Zuckerman (2020) described DPI as a crucial lever for civic engagement, with the potential to transform the state-citizen relationship by enabling more effective citizen-government interactions, promoting transparency and fostering citizen participation.

The exact nature of the state's role in DPI is, however, a matter of debate. On one hand, the state's position as a builder, procurer or facilitator of DPI impacts the reach and authority of institutional mechanisms and has implications around the accountability of DPI systems. If the state is the sole builder and provider of DPI solutions, there is greater scope for intervention on the part of citizens and institutions to mandate the utmost commitment to fairness. On the other hand, as a procurer or facilitator for DPI, the state might rely more on private entities, raising concerns about privacy, surveillance and the commodification of essential services. In certain circumstances, the actions of private entities may not be open to the scrutiny of public institutions, raising concerns about oversight and transparency. This debate around the role of the state takes on added significance with the evolution toward a "digital welfare state" (Gupta 2023), which has seen a shift in how welfare services are delivered, with an increasing reliance on digital platforms and tools.

Breaking private strongholds and revitalizing the role of the state in the digital realm are not the only drivers of DPI. Unlike traditional digital government services, which have often been built in silos that mirror the department or ministry responsible for them, DPI seeks to create cross-cutting components and linkages between systems. Estonia's X-Road is an example. Built using open-source software, X-Road is at its heart a data exchange mechanism that allows various public and private services to integrate and work together more effectively and create innovative new services for citizens.

Effective DPI therefore has a role to play in enabling governments to derive new value from data that may otherwise remain siloed within service-, department- or topic-oriented silos. As with X-Road, this value may take the form of more efficient service delivery by — and across — government departments through better data exchange, or in the enabling of new services to be built using new combinations of data. While measuring the value of providing new, or enhanced, services through better data exchange is difficult, it is an important intended impact of the DPI agenda.

Thus far, most discourse surrounding DPI has laid overwhelming emphasis on its constituent parts and technical specifications. A preoccupation with the components of DPI tends to belabour its technical features, making adoption a mere function of technological feasibility. As a result, critical questions around the non-technical attributes of DPI (Aapti Institute 2023), such as capacity, governance and sustainability, tend to be deprioritized by stakeholders within the DPI ecosystem. This is reflected in the emergence of efforts to "transfer" DPI knowledge and technology (United Nations Development Programme [UNDP] 2023b), while attempts to contend with aspects

such as governance and sustainability have been lagging (Seth et al. 2023). A tech-first approach can also lean toward the commodification of essential services with efficiency having primacy over equity.

Excess focus on attributes also threatens to exclude previous or ongoing efforts at digitalization that might not be explicitly labelled DPI, nor conform to its latest specification, but have been successful or represent progress. This is particularly salient in the context of developing countries where attempts toward digitalization already exist, however nascent; proposing DPI as an "alternative" would seem blinkered, especially given prevailing financial and political investments.

Rather than focus overly on components of DPI, we must think more critically about its orientation. How and why is an infrastructure placed in the space it occupies?

Aapti Institute is becoming more attuned to orientation as we encounter different examples of DPI from around the world. In response to Russia's military aggression and war in Ukraine, ensuring continuity in connectivity, communication and access to state services in Ukraine has been a vital national priority. Supported by the European Union and designed in response to these priorities, the EU4Digital initiative spans high-speed broadband improvements, new digital services, cybersecurity and building citizens' digital skills. Understandably, the focus of the DPI built by EU4Digital has been *resilience*.

Gaia-X is a European association of governments, technology firms, academics, public bodies and not-for-profits, brought together to build common cloud infrastructure and standards. Ursula von der Leyen, president of the European Union, has situated Gaia-X as a key part of the European Union's effort to protect the rights of Europeans, including "the right to privacy and connectivity, [and] freedom of speech" (European Commission 2020). Gaia-X has been described as seeking to help the European Union regain the *sovereignty* it has lost to the United States and China in recent years through those nations' exporting of technology to the region.

Aadhaar is India's biometric identity system built using similar open standards to its Unified Payments Interface. As of March 2024, more than 1.3 billion residents of India have been enrolled on Aadhaar and more than 100 million transactions can be authenticated using the system each day.[2] As the UNDP (2023a) has reported, Aadhaar is playing a key role in tackling poverty by improving the economic resilience of marginalized groups and increasing access to private and public services. With this platform, as well as with others such as Togo's Novissi payment system, the DPI approach has proven instrumental in achieving greater economic *inclusion*.

There are silent orientations of DPI, too, and the potential for disconnect between their stated and actual orientations. While Aadhaar has clearly helped to promote inclusive economic development, it has also been criticized as giving the Indian government "unjust powers to surveil its citizens and deny them their fundamental rights" (Jain 2019). The rollout of a new digital identity system in Kenya has also been criticized for its absence of transparency, public engagement and legal safeguards (Burt 2023). In launching the "State of DPI" study, Anjum Dhamija et al. (2023) have expressed the need for development actors to support DPI adoption and safeguarding.

Those designing and building DPI should be cognizant of orientation. Sarah Drummond[3] has written about intent as a core aspect of "full stack service design," emphasizing the effects of underlying missions, policies and values to the delivery of digital services. Reflecting on the United Kingdom's Government Digital Service experience, Richard Pope (2019) has described how DPI is not just about technological advancement but also about how it is integrated into the wider public policy realm. Keyzom Ngodup Massally, Rahul Matthan and Rudra Chaudhuri (2023) have argued that in order to create

transparent, accountable and fair digital ecosystems, DPI must embed governance into its architecture, including through privacy by design, user autonomy, protocol-based supervision and obligations in code.

Aapti Institute is also exploring how orientations of physical infrastructures could be more intentionally applied to DPI. For example, the concept of *critical* infrastructure is a long-standing one, where assets that serve as the backbone of a nation's economy and society are designated and afforded special protection.[4] The Organisation for Economic Co-operation and Development has advocated for *quality* infrastructure,[5] using it as a concept to promote economic development, effective regulatory frameworks and enabling environments for investment.

As the DPI agenda develops at pace, it is vital that we proactively consider its orientation and introduce new governance frameworks to ensure that digitalization does not exacerbate extant inequities and that it bridges the "digital divide" that often confronts developing nations.

Notes

1 See https://datareportal.com/social-media-users.

2 See https://uidai.gov.in/aadhaar_dashboard/index.php.

3 See https://sarah-drummond.com/full-stack-service-design/.

4 See www.cisa.gov/topics/critical-infrastructure-security-and-resilience.

5 See www.oecd.org/en/topics/sub-issues/infrastructure-and-development.html.

Works Cited

Aapti Institute. 2023. "Resilience for digital infrastructure: Developing assets and impact around the non-technical layers." The Digital Public Lab, February 12.

Austermuhle, Martin. 2018. "Uber And Lyft Push Back Against D.C. Council Demand For Data, Citing Privacy Concerns." WAMU 88.5 American University Radio, May 24. https://wamu.org/story/18/05/24/uber-lyft-push-back-d-c-council-demand-data-citing-privacy-concerns/.

Burt, Chris. 2023. "Kenyan rights groups warn digital ID program repeating past mistakes." BiometricUpdate.com, September 15. www.biometricupdate.com/202309/kenyan-rights-groups-warn-digital-id-program-repeating-past-mistakes.

Business Standard. 2023. "UPI has emerged as most popular and preferred payment mode in India." *Business Standard*, March 8. www.business-standard.com/article/news-cm/upi-has-emerged-as-most-popular-and-preferred-payment-mode-in-india-123030800203_1.html.

Cheng, Amy. 2021. "Much of the world relies on WhatsApp. Its outage ground their virtual lives to a halt." *The Washington Post*, October 5. www.washingtonpost.com/world/2021/10/05/whatsapp-global-outage-blackout/.

Dhamija, Anjum, David Eaves, Krisstina Rao and Jordan Sandman. 2023. "Launching the 'State of DPI' Project." Co-Develop, December 22. www.codevelop.fund/insights-1/launching-the-state-of-dpi-project.

Duncan, Pamela. 2022. "The worldwide scale of the Uber files – in numbers." *The Guardian*, July 15. www.theguardian.com/news/2022/jul/15/the-worldwide-scale-of-the-uber-files-in-numbers.

Eaves, David. 2023. "Unpacking Digital Public Infrastructure and its Importance." Keynote Presentation at the Global Workshop on Digital Public Infrastructure 2023. World Bank, September 12. https://thedocs.worldbank.org/en/doc/98949920afb52c54cd4fc4dd15a02dbd-0050112023/original/1-1-TBS-Sept-11-DPI-presentation-edition.pdf.

European Commission. 2020. "State of the Union Address by President von der Leyen at the European Parliament Plenary." September 15. https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_20_1655.

Gupta, Asha. 2023. "Digitalisation of the Welfare State: Lessons for the Emerging Economies." *Indian Journal of Public Administration* 69 (2): 453–67. https://doi.org/10.1177/00195561231153903.

Jain, Mardav. 2019. "The Aadhaar Card: Cybersecurity Issues with India's Biometric Experiment." The Henry M. Jackson School of International Studies, University of Washington, May 9. https://jsis.washington.edu/news/the-aadhaar-card-cybersecurity-issues-with-indias-biometric-experiment/.

Massally, Keyzom Ngodup, Rahul Matthan and Rudra Chaudhuri. 2023. "What is the DPI Approach?" Carnegie Endowment for International Peace, May 15. https://carnegieendowment.org/research/2023/05/what-is-the-dpi-approach?lang=en.

Pope, Richard. 2019. "Playbook: Government as a Platform." Harvard Kennedy School Ash Center for Democratic Governance and Innovation. November. https://ash.harvard.edu/wp-content/uploads/2024/02/293091_hvd_ash_gvmnt_as_platform_v2.pdf.

Seth, Aaditeshwar, Luís Fernando Vitagliano, Nachiket Udupa, Parminder Jeet Singh, Rakshita Swamy, Subrata Singh and Vineetha Venugopal. 2023. "A Governance Framework for Digital Public Infrastructure: Learning from the Indian Experience." T20 Policy Brief. July. www.defindia.org/wp-content/uploads/2023/07/T20_PB_TF2_205_DPI-Indian-Experience.pdf.

Vailshery, Lionel Sujay. 2024. "Vendor market share in cloud infrastructure services market worldwide 2017-2024." Statista, May 21. www.statista.com/statistics/967365/worldwide-cloud-infrastructure-services-market-share-vendor/.

Stognei, Anastasia and Simon Fraser. 2022. "Ukraine invasion: Russians feel the pain of international sanctions." BBC News, March 1. www.bbc.com/news/world-europe-60558731.

TOI Tech Desk. 2024. "Nearly 8 out of 10 digital payments are now done through UPI, claims RBI." *Times of India*, March 6. https://bfsi.economictimes.indiatimes.com/news/financial-services/nearly-8-out-of-10-digital-payments-are-now-done-through-upi-claims-rbi/108256394.

UNDP. 2023a. *Accelerating the SDGs through Digital Public Infrastructure: A Compendium of the Potential of Digital Public Infrastructure*. New York, NY: UNDP. www.undp.org/sites/g/files/zskgke326/files/2023-08/undp-g20-accelerating-the_sdgs-through-digital-public-infrastructure.pdf.

———. 2023b. "11 'First-Mover' Countries Launch 50-in-5 Campaign to Accelerate Digital Public Infrastructure Adoption around the World." November 9. www.undp.org/news/11-first-mover-countries-launch-50-5-campaign-accelerate-digital-public-infrastructure-adoption-around-world.

Zuckerman, Ethan. 2020. "What is Digital Public Infrastructure Zuckerman, Ethan. 2020. Center for Journalism & Liberty, November 17. www.journalismliberty.org/publications/what-is-digital-public-infrastructure.

About the Authors

Soujanya Sridharan is a senior manager at Aapti Institute, where her research spans a wide range of themes that speak to participatory governance efforts in the data, infrastructure and platform domains. Specifically, Soujanya has been instrumental in leading the development of Aapti's platform economy initiative that advances fair working conditions for millions of gig workers. She holds an M.A. in development from Azim Premji University. Her interest in tech policy stems from her thesis research on labour unions within India's information technology (IT) industry. This was followed by a stint with IT for Change, where she was involved in gender and technology initiatives.

Vinay Narayan is a senior manager at Aapti Institute. He leads Aapti's work on data stewardship and structures for bottom-up data sovereignty, where the current focus is on building networks of support for bottom-up data initiatives in the Global South. He is also the co-lead of Aapti's artificial intelligence (AI) research practice that examines the impact of AI technologies on livelihoods and rights frameworks that can address ongoing and potential impacts. Vinay's current research interest centres around investigating the critical digital infrastructures that support the ecosystem with a view to crafting more effective governance approaches. He is also the lead in Aapti's engagement as a member of the steering committee of the Global South Alliance, a network of 15 civil society organizations dedicated to the promotion of mutual learning and the advancement of digital rights with a Global South perspective.

Jack Hardinges is an independent adviser to organizations that research, build or fund new data products, systems and infrastructure, and serves as fund lead for the Data Empowerment Fund and board member of Open Supply Hub. Previously, Jack was head of programs at the Open Data Institute, where his work advanced the understanding and practice of data stewardship, privacy-enhancing technologies and participatory approaches to data. Over the past 10 years, Jack has published on topics including data rights, the economic impact of data, smart cities and the transparency of foundation models. He has also explored the cultural use and impacts of data, including as an artistic material and its publication to engender trust in sport. Jack holds a B.Sc. in economics from the University of Southampton.

Data Marketplaces and Governance: Lessons from China

Alex He and Rebecca Arcesati

China has lots of data. By one estimate, the country produced 7.6 zettabytes of data in 2018 and will account for 27.8 percent of the global total by 2025, surpassing the United States (Reinsel et al. 2019). The world's largest population of internet users generates vast troves of data as citizens go online to access information, buy and sell products, make payments, chat, order taxis, learn, and consume and produce entertainment. Meanwhile, the world's largest network of surveillance cameras watches their every movement and public services are digitalizing.

Unlocking the value of all this data is a major theme in the Chinese government's digital strategy — one linked to important security, public policy and economic objectives. China's government considers data not only as a tool to cement its authoritarian rule (Hoffman 2019; Mozur, Xiao and Liu 2022), but also as an economic "factor of production" on par with land, labour, capital and technology — a foundation for national power and competitiveness (CCP Central Committee 2019; CCP Central Committee and State Council 2020). It wants to harness data's potential to drive digital transformation, innovation and the upgrading of China's "real economy" (Creemers, Costigan and Webster 2022). By the end of 2025, Beijing wants an efficient market where companies and government bureaucracies share and trade more data (State Council 2021).

Numbers are not everything, however. China faces challenges in getting good data where needed. For example, tech firms are struggling to find enough artificial intelligence (AI) training data, an increasingly pressing issue amid fierce competition with the United States to develop the most powerful large language models (CAICT 2023a). Additionally, as of 2016, more than 80 percent of China's information and data resources were said to be jealously kept by government bureaucracies (*Beijing Daily* 2016), hindering economic development and efficient governance. To tear down these "data islands" and match supply with demand, policy makers are stepping in.

Concerns around national security and socio-economic stability in recent years have led to a massive regulatory overhaul of China's digital economy. A data governance regime is now in place, and regulators have cracked down on big tech's data monopolies and abuses of citizens' personal information (Reuters 2021; He 2023; Zhang 2024). The focus

has therefore shifted to a key missing piece of the puzzle: creating a data trading market for the "orderly sharing of data" (He 2020), which is intended to boost productivity and public welfare while safeguarding security and protecting personal data — under the close watch of the party-state (Arcesati and Groenewegen-Lau 2023).

This essay illuminates this ongoing project by zooming in on the development of local data exchanges — essentially, marketplaces for data. The authors first trace the policy, regulatory and institutional context, explaining how and why China's newly established data exchanges differ from similar experiments of the past. The authors then present key findings from a review of 17 Chinese data exchanges, including their business models, ownership structures, regulatory arrangements, product lists and track record of brokering deals, based on information available on their websites and other public records.

The authors find that these exchanges, especially the more institutionalized ones in Beijing, Shanghai, Shenzhen, Guiyang and Guangzhou, are piloting solutions to some challenges in data economics and governance that are common to other jurisdictions, such as data ownership, data valuation and trust building between data providers and buyers. They are also emerging as innovative testing grounds in new areas: trading of AI training sets, cross-border data transfer and the marketization of public data. At the same time, the authors' observation suggests that the state-centric feature of China's data market may constrain its further development.

Data Trading in China: Progress, Setbacks, Institutionalization

For all the hype around China's data advantage (Lee 2018), Chinese policy makers and leading experts worry that this potential has yet to be realized. To multiply other factors of production, they believe data must flow to those economic actors that can generate value from it. In other words, China needs to "activate the factor value of data" through an efficient data trading market (Yu 2021) and "data resource system" — a foundation toward digital development on equal footing as infrastructure (CCP Central Committee and State Council 2023).

Data exchanges are not new. Following the release of the national strategy for big data development in 2015 (State Council 2015), dozens of pilot data trading platforms mushroomed across the country (Shen and Zhang 2022). They function as intermediary institutions where organizations can buy and sell data products, query some data sets or access related services, such as cleaning, visualization and desensitization. Products run the gamut from training data for autonomous vehicles to corporate credit information. Until recently, however, those pilots were empty shells, accounting for an underwhelming two percent of China's total data trading activity in 2021 (China Mobile 2023), most of which is carried out over the counter.

The problem was straightforward: Without laws, regulations and standards in the areas of data security, personal information protection and data trading, nobody trusted the system — especially not tech firms, who typically like to freely profit off personal data but not undersell the products they develop through processing it. How to properly value and price data assets, define rights of ownership and use, and find trustworthy providers, sellers as well as third-party providers for key services (such as security audits and dispute arbitration), were big questions. Meanwhile, China's data black market has thrived, reaching a scale of CNY 150 billion in 2021 (Shen and Zhang 2022).

This chaotic situation did not sit well with the government's resolve to crack down on monopolies, leaks, theft and misuse of citizens' data by private actors (Shen 2021). In

2021, a new wave of data exchanges began to emerge that, unlike their predecessors, are supposed to operate under tighter government control and within clearer legal boundaries (CAICT 2023b). Authorities hope this "data trading system 2.0" will fix the regional and bureaucratic turf wars of the past (Duan 2022). Besides foundational legislation to protect data security and personal information,[1] a dedicated National Data Administration was created to oversee China's data resources and transactions (Reuters 2023).

Since 2021, there has been progress and setbacks. On the one hand, data exchanges are seeing more activity: The data exchanges of Shenzhen, Guiyang and Guangzhou reached a trading volume of more than CNY 1 billion each as of mid-2023 (see Table 1). By comparison, the Guiyang Big Data Exchange, dubbed "the big data valley of China," had an annual trading volume of less than CNY 5 million before its restructuring (Mu 2021). On the other hand, this volume is still limited compared to the huge size of China's big data industry and total data trading. Issues around data ownership and pricing, as well as a persistent lack of trust, continue to disincentivize companies from trading their data in the market, leading to a supply bottleneck.

To solve this, central and local authorities are striving to bring greater legal and regulatory clarity around data transactions.[2] The Data Security Law called for a "data exchange market" where intermediaries verify traders and have them explain the origin of the data they are selling, to avoid compromising any personal information or other sensitive data.[3]

One key policy, the Opinions on Building a Basic Data System to More Effectively Maximize the Role of Data Elements (also known as the Twenty Data Measures) from December 2022 (CCP Central Committee and State Council 2022), encouraged experimentation around issues such as pricing models and data property rights. Following this impetus, the data exchanges of Shenzhen, Guiyang, Shanghai, Beijing and, to a lesser extent, Guangzhou, began trialling and testing new solutions.

Table 1: The Five Major Data Exchanges in China

Name and Year Founded	Organization Type	Business Model	Main Products	Trading Volume
Guiyang Global Big Data Exchange, 2015	State-owned	Data value-added services	Data products and services, algorithmic tools and resources	CNY 1.4 billion as of July 2023
Shenzhen Data Exchange, 2022	State-owned	Data value-added services	Data products, services and tools	CNY 1.8 billion as of March 2023
Shanghai Data Exchange, 2021	State-owned assets holding	Quasi-public service institution charging data service fee	Data sets, data services	CNY 0.1 billion as of December 2022
Beijing International Big Data Exchange, 2021	State-owned assets holding	Data value-added services	Data products, including data sets, API, reports and data services	Not available
Guangzhou Data Exchange, 2022	State-owned assets holding	Data value-added services	Data products and services, data resources, data assets	CNY 1 billion as of May 2023

Source: Authors' compilation.
Note: Except for the Guangzhou Data Exchange, all the exchanges listed above offer an online platform for data trading.
API = application programming interface.

Local Data Exchanges as Supervisory Bodies, Matchmakers and Testing Grounds

The creation of regulated exchanges with strict oversight over the whole trading process should put every link of the chain under proper supervision. Most of China's newer data exchanges are tightly controlled by the state through various ownership arrangements (see Tables 1 and 2). The Guiyang Global Big Data Exchange underwent restructuring from private to 100 percent state control. Government backing, coupled with the rapid commercialization of new technologies and a more mature regulatory environment, have turned these exchanges from simple intermediaries into full-fledged service providers with a supervisor's hat.

Not only do the exchanges introduce and certify new buyers and sellers, but they also take charge of compliance verifications, security and personal information protection assessments and technical support. Moreover, several exchanges have developed their own rules and guidelines, covering issues ranging from catalogues of data prohibited from trading to specific transaction standards.[4] In Guangdong, the local government tasked "chief data officers" to coordinate the use of public data across government departments (Xiao and Zeng 2022).

The first challenge is to determine ownership, which is tricky because data is a semi-public good and the allocation of related property rights among consumers and firms is ambiguous (El-Dardiry, Dinkova and Overvest 2021). Due to such ambiguity, until recently, data transactions in China were left in a legal limbo. The Twenty Data Measures marked an important step forward by dividing the legal rights of participants in the data market into three categories: ownership of data resources, rights to process and use, and rights to commercialize data. This policy is slowly paving the way for clearer data ownership rules and systems upon which data can be legally traded in China.

Some data exchanges introduced data ownership registration systems to certify the different rights associated with the data being traded on their platform, as well as market entity registration for providers and buyers (Zhejiang Lab et al. 2022). The certificates can be used as a legal basis for data trading, as well as for other purposes, such as financing and debt repayment, incorporating data assets into balance sheets, accounting and dispute resolution (Shenzhen Development and Reform Commission 2023). This approach could incentivize more companies to buy or sell data via institutional exchanges; for example, by guaranteeing the protection of the property rights and interests of data processors such as digital platform companies (Zhang and Xia 2023a).

Digital technologies, such as blockchain, privacy-enhancing technology (PET) and federated learning, are helping with traceability by certifying different data rights throughout the whole data trading process. Data owners and processers can be granted different levels of control, and the latter can only access the information required for processing and using the data, ensuring that "data being traded can be used but not seen" (Du 2022; Zhang 2023). Digital technologies also allow for the tracing of the source, transfer history and final use of the traded data.

A second challenge is to put a price tag on data, which, until recently, was left up for negotiation between providers and buyers. Combined with the scenario-based, highly customized features of data transactions, this approach easily leads to chaos and extortion. Large digital platforms, for example, can charge higher prices thanks to their sheer data power (Zhang and Xia 2023b).

Both government officials and professional associations have offered recommendations for data valuation and pricing. Wang Jiandong, deputy director of the National Development and Reform Commission's (NDRC's) Price Monitoring Center in 2023,

advocated using cost pricing for data resources and an income-based approach for data assets. The former considers all types of investment, such as labour, time and equipment

Table 2: Other Active Data Exchanges

Name and Year Founded	Organization Type	Business Model	Main Products	Trading Volume
North Big Data Exchange Center, 2021	Mixed ownership with state-owned assets' shares	Data value-added services	Data products and services	CNY 0.15 billion (aspirational target as of May 2023)
East China Jiangsu Big Data Exchange Center, 2015	Joint stock company	Annual membership fee	Data products and services	Not available
Zhengzhou Data Exchange Center, 2022	State-owned assets holding	Quasi-public service institution providing data value-added services	Data products and services	CNY 0.1 billion as of June 2023
Western China Data Exchange, 2021	100 percent state-owned assets	Data value-added services	Data products	CNY 0.1 billion as of January 2023
Changjiang Data Exchange, 2015	State-owned assets holding	Data trading and renting services; membership fee	Data products	Not available
Zhejiang Big Data Exchange Center, 2016	State-owned assets holding	Commissions, membership and data service fees	Data sets, API, reports, AI models, data services	Not available
De Yang Data Exchange, 2022	State-owned assets holding	Data trading services	Data products and services	CNY 23.8 million as of July 2023
Shanxi Data Exchange, 2020	Public-private partnership between the Shanxi government and Baidu	Data trading services and data value-added services	AI data sets, API, index	CNY 50 million as of March 2021
Shandong Data Exchange, 2019	State-owned provincial data service platform	Data trading services and data value-added services	Data sets, reports, applications, API, privacy-enhancing computing, data services	Not available; 2022 revenues were CNY 14.4 million; 2022 net profits were CNY 1.2 million
Beibu Gulf Big Data Trading Center, 2020	State-owned assets holding	Data value-added services; authorized use or direct purchase of data	Data sets, API, solutions	CNY 15 million as of 2020
Hefei Data Factor Circulation Platform, 2021	State-owned via Hefei Big Data Asset Operation Co., Ltd.	Data value-added services	Data products (data sets, API, reports), services tools	CNY 41 million as of June 2023
Hainan Supermarket for Data Products, 2021	Run by the Hainan provincial government	Platform for public data products	Data sets, API, reports, models, data services	CNY 0.4 billion as of July 2023

Source: Authors' compilation.
Note: Except for the East China Jiangsu Big Data Exchange Center, all the data exchanges listed above offer an online platform for data trading.

in data collection and standardization, plus data quality and privacy, to estimate the value of data resources. The latter, which Wang recommended for data assets, sets a price based on the expected income from future value (Yu 2023). By contrast, the China Appraisal Society (2020) did not differentiate between data resources and data assets and suggested considering a combination of costs, expected income and historic prices.

Some exchanges introduced recommendations and guidelines in this regard. Once a data provider has made an initial offer, the data exchanges or third-party agencies set a reference price, considering both the embedded costs in the data as well as the benefits that buyers could derive from it, plus other factors such as consumers' expectations, supply and demand, historic prices and customer segments. Importantly, the central government seems to favour the exploration of data-pricing formation mechanisms through close cooperation with the data exchanges. The Price Monitoring Center began working with exchanges in 2023 to valuate data assets and test pricing mechanisms.

A third challenge is to establish trust in the market, absent which providers and buyers will continue to prefer over-the-counter trading over institutional channels — with all the privacy, security and legal risks that come with it. To overcome this, several exchanges are trying to create trusted ecosystems. Providers in data-rich sectors, such as utilities and internet platforms, as well as buyers, such as commercial banks, government agencies and AI companies, are incorporated into the ecosystem alongside third-party service providers. The exchanges handle basic services and supervise trading, while third parties deal with value-added services, such as data-quality certification, security and compliance verifications, and dispute resolution.

For example, the Shenzhen Data Exchange brokered a loan agreement between China Everbright Bank's Shenzhen branch and an AI infrastructure company, Shenzhen Weiyan Technology, based on the latter's data products listed on the exchange (Tang 2023; Zhu 2023). The ecosystem around the exchange, which includes law firms and other third-party service providers that assist with determining data rights and valuation, assessing data quality, and verifying compliance, played a key role in facilitating the deal.

Data exchanges can also build trust by introducing high-quality products, acting as matchmakers for transactions that otherwise may not materialize. For example, one enterprise's electricity usage data product on sale on the Shenzhen Data Exchange was used by a local government bureau to evaluate whether to grant companies the high and new technology enterprise status, one of China's main tax incentives for innovation. Based on the same information, the Bank of Ningbo approved a loan to an electronic device manufacturer (Pan 2023).

Trust building is also a precondition to personal data trading, whose scope is extremely restricted under the Personal Information Protection Law of 2021.[5] The Guiyang Global Big Data Exchange became the first to carry out personal data trading. Based on PET and other digital technologies, the recruiting platform Haohuo desensitized the resumés of job seekers as data products, such that any personally identifiable information would be hidden from users. The resumé data product was then listed on the exchange, which assigned it a reference price, while a law firm provided a legal assessment. Individuals whose resumés were traded would, at least on paper, receive a share of the revenues from Haohuo (Fang 2023).

It is important to note that, so far, the successful cases of deals brokered by China's main data exchanges are largely due to government coordination among state-owned or state-linked participants. This makes it challenging to determine the extent to which participation in the ecosystem is even voluntary. For example, the data asset-based credit line to Shenzhen Weiyan Technology was instructed by the Shenzhen Municipal Government and the city's financial supervision agencies. Many deals brokered in Guiyang, Beijing and Shanghai seem to have followed the same model.

Emerging Trends: Trading Data for AI Training, for Public Services and Across Borders

Chinese data exchanges are also tackling emerging challenges that are relevant for other data valuation and trading efforts around the world. These include the rapid growth of AI training data trading, cross-border data trading and the trading of public (government) data.

Amid booming demand and strict legal and regulatory requirements, China's labour-intensive AI training data collection and annotation market is changing (Matsakis 2023). Initially reliant on their own teams and crowdsourcing, AI firms have set up dedicated bases for data collection and annotation (labelling) — key steps in preparing data for model training. Baidu, one of China's leading tech giants, jointly built one such base with the Shanxi Data Exchange platform. The base employs 2,000 data annotators,[6] with popular use cases spanning autonomous driving and biometric recognition.

The Shanxi Data Exchange platform aims to become China's biggest marketplace for AI data products in China and a one-stop "data factory" for collection and annotation. As of this writing, 381 data products were listed on the exchange, 261 being AI-related.[7] Looking ahead, it is possible that more AI training data will be collected, annotated and offered to tech companies through data exchanges. Other data exchanges, such as the Beijing International Big Data Exchange, are already catching up and listing their own AI training data products to ride the wave of AI development.

By contrast, only a few exchanges have started offering cross-border data-trading services. Among them, only the Shenzhen Data Exchange has conducted trials, whereas the Shanghai Data Exchange appears to only provide data import services on its international data board (*Shanghai Observer* 2023).[8] The Beijing International Big Data Exchange, meanwhile, offers data hosting and desensitization services to multinational corporations operating in China (Chaoyang District People's Government of Beijing Municipality 2022). That Shenzhen is an isolated case is not surprising, given China's extremely stringent localization requirements and security review process for data exports.

As of March 2023, 16 cross-border deals had been closed through the Shenzhen Data Exchange (Shenzhen Municipal People's Government 2023), for a total value of more than CNY 11 million (Gong 2022). The first deal, worth CNY 5 million, involved a foreign hedge fund purchasing ChinaScope's flagship data product, the SmarTag news analysis engine, which uses a natural language processing algorithm to convert unstructured Chinese language news text into machine-readable metadata. The product compiles sentiment indicators linked to Chinese companies, supporting market analysis (Yuan 2022).

Here, again, the impetus came from the central government. The Ministry of Commerce has been trying to encourage free trade zones in China to pilot "safe and orderly" cross-border trading since 2020 (Sino-German Cooperation on Industrie 4.0 2020), with few results. This move probably prompted the government to bet on Shenzhen (NDRC and State Council 2022). The city was also the first locality in China to define data ownership rights and is trialling AI data trading with Hong Kong (Yuan 2022).

The sustainability of these trials is far from clear. The powerful and security-focused Cyberspace Administration of China (CAC) has been dragging its feet over most applications for data export security review, which is required when sensitive data such

as personal information and “important” or “core national” data is involved (Arcesati and Groenewegen-Lau 2023). The CAC recently announced a major policy relaxation that would let business, not regulators, decide when cross-border data flows are necessary for their global operations (CAC 2024). However, implementation will need to follow.

The government would like data exchanges to work with the CAC on these security assessments, but it has not given them the policy space and regulatory certainty to do so. The party-state’s ongoing push to obscure from foreign eyes more and more data about China’s economy, science and technology, and industries (Brussee and von Carnap 2024) casts doubt on the future of cross-border data trading.

Another space to watch is the trading of public data, a prerequisite for economic and public policy innovation around the world. Despite a nearly decade-long effort, China’s open-data government platforms are still marred by quality problems, with as much as 85 percent of data collected and made available for public inquiry said to be incomplete (CAICT 2023c). There is hope that data exchanges could solve the problem by creating an effective pathway to the safe circulation of standardized and high-quality public data, considering the sensitivities around national security, personal information protection and business secrets.

The Beijing International Big Data Exchange and the Hainan Supermarket for Data Products are both front-runners, although their models notably differ. After some initial success in the finance sector, the local government of Beijing entrusted the data exchange with managing its entire public data resources, turning government open-data platforms into marketplaces (Li 2023). The platform remains a work in progress, with most products featuring unstructured statistical data and the government’s own open-data platform still offering a greater variety for free. Tellingly, half of the listed products are credit inquiry services offered by the same state-owned financial big data company that runs the exchange on behalf of the government.[9]

The government-run model in Hainan is based on a larger ecosystem of actors and data developers and seems more promising. The Hainan Big Data Administration invites companies to develop products and services based on data resources made available by the local government. These data products and services are then listed and traded at the Hainan Supermarket for Data Products (Dong 2021). As a technology partner, Tianyi Cloud, a state-owned cloud service provider, desensitizes public data using PET, secure multi-party computation and federated learning.[10] As of this writing, 1,070 data products and services had been developed and commercialized.[11]

Toward the World’s First State-Led Data Market

China is a clear first mover in elevating data to a national strategic priority and designating it as a factor of production. This reflects a uniquely state-driven approach to digital governance, which can also be seen in the high degree of government control and coordination behind the data exchanges discussed in this essay. Beijing’s push to have the state direct data circulation provides the momentum behind the ongoing reform of China’s data market, yet it may also pose the biggest obstacle to its further development. Although more private firms are joining in, China’s data exchanges are still a playground for state-owned enterprises and companies with strong government connections.

Moreover, as most exchanges are abandoning commission fees in favour of membership-based business models where users are charged a fee to access value-added services, profitability remains a question. With access to abundant capital thanks to government involvement, data exchanges can presumably afford not to be profitable for some

time. However, this could become a challenge long term if the most competitive (private) tech companies continued to snub these institutional channels. China's newly established National Data Administration will have some convincing to do, following an unprecedented regulatory crackdown led by the CAC and other agencies that burned more than US$1 trillion in market value from China's leading tech firms, strengthened the party-state's grip over privately held data and damaged the economy.

These developments will carry implications beyond offering lessons for other jurisdictions, considering that Chinese authorities are also exploring how some of these marketplaces could serve as gateways for cross-border data transmissions (von Carnap 2022). The Shanghai Data Exchange, for example, has pledged to align with standards outlined in the Digital Economy Partnership Agreement, which China officially applied to join. The extent to which the country will seek to integrate its data market with those of its trading partners or prioritize domestic circulation in the name of national security remains to be seen.

Authors' Note

This essay is based on a longer unpublished paper for which research was completed in the fall of 2023. It does not consider developments after China's National Data Administration kicked off its work and issued its first policy at the end of 2023.

Notes

1 中华人民共和国数据安全法 [*Data Security Law of the People's Republic of China*], 10 June 2021 (entered into force 1 September 2021) [*Data Security Law*], online: 中国人大网 [npc.gov.cn] <www.npc.gov.cn/npc/c30834/202106/7c9af12f51334a73b56d7938f99a788a.shtml>; 中华人民共和国个人信息保护法 [*Personal Information Protection Law of the People's Republic of China*], 20 August 2021 (entered into force 1 November 2021) [*Personal Information Protection Law*], online: 中国人大网 [npc.gov.cn] <www.npc.gov.cn/npc/c30834/202108/a8c4e3672c74491a80b53a172bb753fe.shtml>.

2 深圳经济特区数据条例 [*Data Regulations of Shenzhen Special Economic Zone*], 7 July 2021 (entered into force 1 January 2022), online: 深圳政府在线 [Shenzhen Government Online] <www.sz.gov.cn/attachment/0/980/980196/9835431.pdf>; 上海市数据条例 [*Data Regulations of Shanghai Municipality*], 25 November 2021 (entered into force 1 January 2022), online: 一网通办 [Government Online/Offline Shanghai] <www.shanghai.gov.cn/nw12344/20211129/a1a38c3dfe8b4f8f8fcba5e79fbe9251.html>; see also (National Information Security Standardisation Technical Committee 2022).

3 *Data Security Law*, *supra* note 1.

4 For example, the Shanghai Data Exchange released guidelines on data trading safety in October 2023.

5 See *Personal Information Protection Law*, *supra* note 1; see also Zhang and Xia (2023b).

6 Data from the Baidu (Shanxi) Artificial Intelligence Data Annotation Center; see https://zhongbao.baidu.com/mark/home/shanxi.

7 See the Shanxi Data Exchange platform's website at http://106.13.54.96/datahub/tradepage/mall/list?publicMethod=0.

8 See the website of the Shanghai Data Exchange's international board at https://nidts.chinadep.com/ep-hall.

9 Calculated based on the online data trading market at the Beijing International Big Data Exchange; see https://webs.bjidex.com/sys-bsc-home/#/bscConsole/tradingMarket.

10 See the introduction to Tianyi Cloud's services for the Hainan Supermarket for Data Products at www.ctyun.cn/cases/596200642071100416.

11 Calculated based on the online data trading market at the Hainan Supermarket for Data Products; see www.datadex.cn/app/dataMarket.

Works Cited

Arcesati, Rebecca and Jeroen Groenewegen-Lau. 2023. *China's data management: Putting the party-state in charge*. Hinrich Foundation Report. December. https://merics.org/en/report/chinas-data-management-putting-party-state-charge.

Beijing Daily. 2016. "李克强: 信息数据 "深藏闺中" 是极大浪费" [Li Keqiang: Information and data being hidden away by the government is such a huge waste]. *Beijing Daily*, May 13. www.gov.cn/xinwen/2016-05/13/content_5073036.htm.

Brussee, Vincent and Kai von Carnap. 2024. *The Increasing Challenge of Obtaining Information from Xi's China*. MERICS report. February 15. https://merics.org/en/report/increasing-challenge-obtaining-information-xis-china.

CAC. 2024. "促进和规范数据跨境流动规定" [Provisions on Facilitating and Regulating Cross-Border Data Flows]. Order of the CAC No. 16, March 22. https://web.archive.org/web/20240406020429/https://www.cac.gov.cn/2024-03/22/c_1712776611775634.htm.

CAICT. 2023a. "数据要素白皮书 (2023年)" [White Paper on Data Factors 2023]. September. www.caict.ac.cn/english/research/whitepapers/202311/P020231103487266783845.pdf.

———. 2023b. "数据要素白皮书 (2022年)" [White Paper on Data Factors 2022]. January. www.szlawyers.com/file/upload/20230117/file/20230117144644_9fee48fd8ac845bbb11989f5b197ede4.pdf.

———. 2023c. "*中国数字经济发展研究报告 (2023年)*" [Research Report on the Development of China's Digital Economy (2023)]. September. www.caict.ac.cn/kxyj/qwfb/bps/202304/P020240326636461423455.pdf.

CCP Central Committee. 2019. *Decision of the Central Committee of the Chinese Communist Party on Some Major Issues Concerning Adhering to and Refining the System of Socialism with Chinese Characteristics and Advancing the Modernization of China's National Governance System and Governance Capacity*. Adopted at the Fourth Plenum of the 19th Central Committee of the Chinese Communist Party, October 31, 2019. Translated by CSIS Interpret: China, November 5. https://interpret.csis.org/translations/decision-of-the-fourth-plenum-of-the-19th-central-committee-of-the-chinese-communist-party/.

CCP Central Committee and State Council. 2020. "中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见" [Opinions of the CCP Central Committee and the State Council on Improving the Systems and Mechanisms for Market-oriented Allocation of Factors of Production]. March 30. www.gov.cn/zhengce/2020-04/09/content_5500622.htm.

———. 2022. "中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见" [Opinions of the CCP Central Committee and the State Council on Establishing Data Basic Systems to Better Bring into Play the Role of Data Elements]. December 19. www.gov.cn/zhengce/2022-12/19/content_5732695.htm.

———. 2023. "数字中国建设整体布局规划" [Plan for the Overall Layout of Building a Digital China]. February 27. www.gov.cn/zhengce/2023-02/27/content_5743484.htm.

Chaoyang District People's Government of Beijing Municipality. 2022. "北京国际大数据交易所打造全国首个服务跨境场景的数据托管服务平台" [Beijing International Data Exchange establishes the first nationwide data trust service platform servicing cross-border use cases]. April 18. www.bjchy.gov.cn/lqjs/lqdt/4028805a811a47da01811f14f64e046b.html.

China Appraisal Society. 2020. "资产评估专家指引第9号----数据资产评估" [Expert Guidance on Asset Appraisal No. 9----Data Asset Appraisal]. 中国资产评估协会 [China Appraisal Society], January 9. https://data.scsio.ac.cn/api/web/v1/file/load/event/1565513788624912384.pdf.

China Mobile. 2023. "数联网（DSSN）白皮书（2023)" [Data Switching Service Network White Paper 2023]. April. www.txrjy.com/thread-1309523-1-1.html.

Creemers, Rogier, Johanna Costigan and Graham Webster. 2022. "Translation: Xi Jinping's Speech to the Politburo Study Session on the Digital Economy – Oct. 2021." DigiChina, January 28. https://digichina.stanford.edu/work/translation-xi-jinpings-speech-to-the-politburo-study-session-on-the-digital-economy-oct-2021/.

Dong, Xuegeng. 2021. "加快数据资源开发利用与数据产品超市平台构建" [Accelerate the exploration and utilization of data resources and the construction of market platforms for data products]. 国脉互联 (govmade.cn), November 24. www.govmade.cn/viewpoint/20211124/648121624679153664.html.

Du, Chuan. 2022. "数据交易2.0时代来临，隐私计算让数据"可用不可见" [Data trading 2.0 coming has come, and privacy-enhancing computing enables data "being used but not seen"]. *第一财经* [Yicai], June 7. https://m.yicai.com/news/101436079.html.

Duan, Siyu. 2022. "国家信息中心王建冬：数据要素市场化配置水平仍待提升" [Wang Jiandong from the State Information Center: The level of market-based data allocation needs improvement]. *第一财经* [Yicai], September 4. https://m.yicai.com/news/101527037.html.

El-Dardiry, Ramy, Milena Dinkova and Bastiaan Overvest. 2021. "Policy Options for the Data Economy – a Literature Review." CPB Netherlands Bureau for Economic Policy Analysis. CPB Background Document. May. www.cpb.nl/sites/default/files/omnidownload/CPB-Background-Document-Policy-Options-Data-Economy-Literature-Review.pdf.

Fang, Yali. 2023. "贵阳大数据交易所完成首笔个人数据合规流转场内交易" [Guiyang Big Data Exchange completes first transaction of compliant personal data circulation]. *贵州日报* [Guizhou Daily], May 11. https://finance.sina.cn/2023-05-11/detail-imytitxi0087843.d.html?from=wap.

Gong, Wenjun. 2022. "先行先试! 深圳又一交易所诞生, 如何打造?" [Go and try first. Shenzhen established another exchange. How should it develop?]. *中国基金报* [ChinaFund], December 3. https://finance.sina.com.cn/china/gncj/2022-12-03/doc-imqmmthc6944653.shtml?cre=tianyi&mod=pcfinf&loc=4&r=0&rfunc=18&tj=cxvertical_pc_finf&tr=12.

He, Alex. 2023. *State-Centric Data Governance in China*. CIGI Paper No. 282. Waterloo, ON: CIGI. www.cigionline.org/publications/state-centric-data-governance-in-china/.

He, Lisheng. 2020. "推动数据由资源向要素转化" [Promote data to transition from resource to factor of production]." *解放日报* [Jiefang/Liberation Daily], August 25. https://ex.chinadaily.com.cn/exchange/partners/82/rss/channel/cn/columns/snl9a7/stories/WS5f446834a31030182d641111.html.

Hoffman, Samantha. 2019. "Engineering global consent: The Chinese Communist Party's data-driven power expansion." Policy Brief Report No. 21/2019. Australian Strategic Policy Institute. www.aspi.org.au/report/engineering-global-consent-chinese-communist-partys-data-driven-power-expansion.

Lee, Kai-Fu. 2018. *AI Superpowers: China, Silicon Valley, and the New World Order*. New York, NY: Harper Business.

Li, Zhiyong. 2023. "北京金融公共数据专区助力金融 '活水' 精准 '滴灌'" [Beijing Financial Public Data Zone facilitates accurate financial flows]. *经济参考报* [Economic Information Daily], January 11. www.jjckb.cn/2023-01/11/c_1310689641.htm.

Matsakis, Louise. 2023. "The hidden army of workers in China influencing how AI is built." *Semafor*, March 3. www.semafor.com/article/03/02/2023/the-hidden-workers-in-china-influencing-ai-like-chatgpt.

Mozur, Paul, Muyi Xiao and John Liu. 2022. "How China Is Policing the Future." *The New York Times*, June 25. www.nytimes.com/2022/06/25/technology/china-surveillance-police.html.

Mu, Luo Mantian. 2021. "理想很丰满现实很骨感 贵阳大数据交易所这六年" [Plentiful in ideal, skinny in reality: Six years of Guiyang Big Data Exchange]. *证券时报网* [STCN], July 12. https://stock.stcn.com/djjd/202107/t20210712_3426536.html.

National Data Administration. 2023. "数据要素×"三年行动计划 (2024–2026年) (征求意见稿)" ["Data Elements ×" Three-Year Action Plan (2024-2026) (draft for comments)]. NDRC, December 15. www.ndrc.gov.cn/hdjl/yjzq/202312/t20231215_1362671.html.

National Information Security Standardisation Technical Committee. 2022. "信息安全技术重要数据识别指南(草案)" [Information security technology – Guideline for identification of critical data (draft)]. January 7. www.tc260.org.cn/file/2022-01-13/bce09e6b-1216-4248-859b-ec3915010f5a.pdf.

NDRC and State Council. 2022. "国家发展改革委、商务部关于深圳建设中国特色社会主义先行示范区放宽市场准入若干特别措施的意见" [Opinions of the National Development and Reform Commission and the Ministry of Commerce on Several Special Measures for Relaxing Market Access in Shenzhen's Building of a Pioneering Demonstration Zone for Socialism with Chinese Characteristics]. No. 35. *北大法宝* [PKU Law], January 24. www.lawinfochina.com/display.aspx?id=37775&lib=law&SearchKeyword=&SearchCKeyword=&EncodingName=big5.

Pan, Xutao. 2023. "数据二十条'促进数据合规高效流通使用" [Data Twenty Measures promote compliant and efficient data circulation and use]. *人民日报海外版* [People's Daily Overseas Edition], February 7. www.gov.cn/xinwen/2023-02/07/content_5740426.htm.

Reinsel, David, Lianfeng Wu, John F. Gantz and John Rydning. 2019. "The China Datasphere: Primed to Be the Largest Datasphere by 2025." IDC White Paper. January. www.seagate.com/files/www-content/our-story/trends/files/data-age-china-idc.pdf.

Reuters. 2021. "Ant Group's micro loan service Huabei begins to share data with China's central bank." Reuters, September 22. www.reuters.com/world/china/ant-groups-micro-loan-service-huabei-begins-share-data-with-chinas-central-bank-2021-09-22/.

———. 2023. "China to form a national bureau to manage its troves of data." Reuters, March 7. www.reuters.com/world/china/china-form-national-data-bureau-2023-03-07/.

Shanghai Data Exchange. 2023. "关于发布实施上海数据交易所交易安全合规指引的通知" [Notice on Issuing and Implementing the Data Transaction Security Compliance Guidelines of Shanghai Data Exchange]. October 19. www.chinadep.com/bulletin/notice/CTC_20231019151825828677.

Shanghai Observer. 2023. "数据跨境两步走？上海数交所首设国际板：先探进口，出口很小心但有新进展" ["A two-step approach for cross-border data trading? Shanghai Data Exchange initiates international data board: Data import first and cautious on data export but with new progress."] May 23. https://web.shobserver.com/staticsg/res/html/web/newsDetail.html?id=615395.

Shen, Xinmei. 2021. "To build a 'Digital China', the country must first deal with its rampant black market for personal information." *South China Morning Post*, March 30. www.scmp.com/tech/big-tech/article/3127485/build-digital-china-country-must-first-deal-its-rampant-black-market.

Shen, Yan and Junni Zhang. 2022. "数据流通的挑战与应对" [Data circulation: Challenges and responses]. *中新经纬* [Economic View], September 28. www.jwview.com/jingwei/html/09-28/505822.shtml.

Shenzhen Development and Reform Commission. 2023. "深圳市数据产权登记管理暂行办法" [Interim Measures of the Shenzhen Municipality on the Administration of Data Ownership Registration]. July 4. www.sz.gov.cn/cn/xxgk/zfxxgj/zcfg/content/post_10692613.html.

Shenzhen Municipal People's Government. 2023. "深圳数据交易所: 建设具有国际影响力的全国性数据交易平台" [Shenzhen Data Exchange: Building a national data trading platform with international influence]. May 12. www.sz.gov.cn/cn/xxgk/zfxxgj/bmdt/content/post_10586630.html.

Sino-German Cooperation on Industrie 4.0. 2020. "Cross-Border Data Transfer Piloting – Hainan Free Trade Port." Policy Briefing. December. www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/China/Policy-Briefing-Cross-BorderDataTransfer.pdf?__blob=publicationFile&v=1.

State Council. 2015. "促进大数据发展行动纲要" [Action Outline for Promoting the Development of Big Data]. Xinhua News Agency, September 5. www.gov.cn/xinwen/2015-09/05/content_2925284.htm.

———. 2021. "国务院关于印发"十四五"数字经济发展规划的通知" ["Notice of the State Council on Issuing the 14th Five-Year Plan on Digital Economy Development"]. State Council Announcement No. 29. December 12. www.gov.cn/zhengce/content/2022-01/12/content_5667817.htm.

Tang, Wei. 2023. "全国首笔！无质押数据资产增信贷款深圳落地]" [The country's first! Data asset-based unsecured credit enhancement loan in the country issued in Shenzhen.] *证券时报网* [STCN], April 4. www.stcn.com/article/detail/833337.html.

von Carnap, Kai. 2022. "Beijing's watchful eye on all data flowing in and out of China." MERICS, July 8. https://merics.org/de/kommentar/beijings-watchful-eye-all-data-flowing-and-out-china.

Xiao, Wengei and Peiqian Zeng. 2022. "广东进一步加快数据流通和交易" [Guangdong further accelerates data circulation and trading]. *南方日报* [Southern Daily], September 23. www.gov.cn/xinwen/2022-09/23/content_5711256.htm.

Yu, Lin. 2021. "发展数字经济应抓住数据要素市场化这个关键" [Development of digital economy should focus on the key issue of data factor marketization]. *光明日报* [Guangming Daily], July 20. http://theory.people.com.cn/n1/2021/0720/c40531-32162737.html.

Yu, Xiangming. 2023. "建立数据价格机制 护航数字经济发展——专访国家发展改革委价格监测中心副主任王建冬" [Establishing data pricing mechanisms to guide the development of the digital economy – Exclusive interview with Wang Jiandong, Deputy Director of the Price Monitoring Center, NDRC]. 上海证券报 [Shanghai Securities News], April 26. https://news.cnstock.com/industry,rdjj-202304-5052479.htm.

Yuan, Jiongxian. 2022. "数据交易问路: 全国首批跨境数据交易是这样'跨境'的" [Data trading exploration: This is how the first nationwide batch of cross-border data transactions 'crossed borders']. 南都大数据研究院 [Nandu Big Data Institute], May 17. https://m.mp.oeeee.com/a/BAAFRD000020220515683739.html.

Zhang, Angela Huyue. 2024. *High Wire: How China Regulates Big Tech and Governs Its Economy*. Oxford, UK: Oxford University Press.

Zhang, Dan and Xingrui Xia. 2023a. "数据交易合规系列(下) -数据交易中的合规要点及应对" [Series on Compliance in Data Trading: Major Issues and Solutions in Data Trading Compliance]. 锦天城律师事务所 [Allbright Law Offices], April 11. www.allbrightlaw.com/SH/CN/10475/b0be235dd460442.aspx.

———. 2023b. "数据交易合规系列(上) -数据交易制度的基本介绍" [Series on Compliance in Data Trading: Basic Introduction to Data Trading Systems]. 锦天城律师事务所 [Allbright Law Offices], April 4. www.allbrightlaw.com/CN/10475/7b4fe58ed4a1e455.aspx.

Zhang, Ye. 2023. "隐私计算: 让数据 "可用不可见" [Privacy-enhancing computing enables data "being used but not seen"]. 科技日报 [Science and Technology Daily], April 10. http://finance.people.com.cn/n1/2023/0410/c1004-32660651.html.

Zhejiang Lab et al. 2022. "数据产品交易标准化白皮书" [White Paper on the Standardization of Data Trading Products]. www.shujiaowang.cn/uploads/20230514/6d99e00f8b11f0177c8235edcc76cca3.pdf.

Zhu, Lin. 2023. "光大银行深圳分行携手深圳数据交易所成功落地首笔小微企业数据资产融资业务" [China Everbright Bank's Shenzhen branch joins hands with Shenzhen Data Exchange to successfully complete the first instance of data asset financing for small and micro businesses]. 深圳新闻网 [Sznews.com], April 6. www.sznews.com/news/content/2023-04/06/content_30160729.htm.

About the Authors

Xingqiang (Alex) He is a CIGI senior fellow. He is an expert on digital governance in China, the Group of Twenty (G20), China and global economic governance, domestic politics in China and their role in China's foreign economic policy making, and Canada-China economic relations.

Prior to joining CIGI in 2014, Alex was a senior fellow and associate professor at the Institute of American Studies at the Chinese Academy of Social Sciences (CASS) and a visiting scholar at the Paul H. Nitze School of Advanced International Studies, Johns Hopkins University, in Washington, DC (2009–2010).

Alex is the author of *The Dragon's Footprints: China in the Global Economic Governance System under the G20 Framework*, published in English (CIGI Press, 2016) and Chinese editions, and co-author of *A History of China-U.S. Relations* (Chinese Social Sciences Press, 2009).

Alex has a Ph.D. in international politics from the Graduate School of CASS and previously taught at Yuxi Normal University in Yunnan Province, China.

Rebecca Arcesati is lead analyst in the Science, Technology and Innovation Program at the Mercator Institute for China Studies (MERICS). Her research focuses on China's technology and digital policy as well as Europe-China innovation relations. She covers the global footprint of Chinese tech firms, digital infrastructure and surveillance tools, governance of data and AI, technology transfer and research collaboration. Prior to joining MERICS, Rebecca gained experience helping Italian tech start-ups scale in China and as a research assistant in the China office of UN Women.

She holds an LL.M. in China studies (politics and international relations) from Peking University, where she was a Yenching Scholar. Rebecca received a master's degree in international studies from the University of Turin and a bachelor's degree in language mediation and cross-cultural communication from the University of Milan. She has studied and worked in Beijing, Shanghai and Dalian.

Why We Need Inclusive Data Governance in the Age of AI

Jeni Tennison

The release of ChatGPT to the world in November 2022 resulted in multiple calls for urgent action on artificial intelligence (AI) governance. ChatGPT was a shock to the system in three ways.

First, it demonstrated — in consumer-friendly form — the expanding capabilities of AI, with clear implications that AI would soon be able to carry out a much wider range of tasks than had previously been thought. Concerns about the risks of AI went mainstream and started to hit the creative and professional services sectors that had previously seemed immune to automation.

Second, it highlighted the importance of a different kind of data: data we have an interest in because we produced it — our writing, art and code. Where previously data governance had a strong focus on privacy and personal data — data about us — generative AI has challenged previous notions of "fair use" and the implications of text and data mining exceptions to intellectual property rights.

Finally, it showed that big tech companies such as OpenAI, Microsoft and Google are able and willing to launch socially disruptive, and potentially dangerous, services into the world, placing governments in a position of catch-up. For politicians, it felt like history repeating itself. During a US Congressional hearing on AI, Senator Richard Blumenthal said, "Congress failed to meet the moment on social media. Now we have the obligation to do it on AI before the threats and the risks become real" (Zorthian 2023).

In truth, the impacts of AI and data processing, more generally, have been with us for some time (for example, see Emily Bender's [2023] blog post "Talking about a 'schism' is ahistorical"). Much of this existing research concluded that the public should be part of data (and, by extension, AI) governance. For example, Jathan Sadowski, Salomé Viljoen and Meredith Whittaker (2021) wrote, "Everyone should decide how their digital data are used — not just tech companies." The Ada Lovelace Institute (2021) has similarly called for public participation in the stewardship of data. The nuances of how to get participation and democratization right have been explored in critical papers from Abeba Birhane et al. (2022) and Johannes Himmelreich (2022).

However, despite this scholarship, much of the generative-AI-inspired attention on AI governance has been carried out through cordial agreements between governments

and the big AI companies, rather than inclusive multi-stakeholder processes (Chatterjee 2023). For example, few civil society organizations were invited to attend the United Kingdom's AI Safety Summit in November 2023; those that did released a statement saying:

> Because only a small subset of civil society actors working on Artificial Intelligence issues were invited to the summit, these are the perspectives of a limited few and cannot adequately capture the viewpoints of the diverse communities impacted by the rapid rollout of AI systems into public use. Here, too, governments must do better than today's discussions suggest. *It is critical that AI policy conversations bring a wider range of voices and perspectives into the room*, particularly from regions outside of the Global North. Framing a narrow section of the AI industry as the primary experts on AI risks further concentrating power in the tech industry, introducing regulatory mechanisms not fit for purpose, and excluding perspectives that will ensure AI systems work for all of us (AI Now Institute et al. 2023; author's emphasis).

This lack of inclusion is mirrored at a more operational level. With some important exceptions, such as public and patient involvement and engagement around health data,[1] organizations developing and deploying AI rarely involve the affected communities, the wider public or civil society in their data governance processes.

In this essay, the author explores the case for ensuring that governance of data, particularly in the context of AI, is inclusive. The author will first define what is meant by inclusion in data governance and illustrate how it manifests at different governance levels. Then the author will go on to explore a range of arguments for greater inclusion of both civil society and affected communities in data governance to support AI.

Defining Inclusive Data Governance

Data is sometimes described as the lifeblood of AI (Shubladze 2023). Machine learning draws patterns out from large quantities of training data, to make predictions, provide recommendations and, with the latest generative AI systems, create new data and content. The content, quantity and quality of training data have a significant impact on what AI systems can do and the biases they replicate.

Data governance is all about making and monitoring decisions about how data is collected, used and shared, with knock-on impacts on what AI systems can be built and how they work. Data governance decisions may be made at global, regional, national and local levels, including within individual organizations. For example, we might see regional discussions about limiting the use of facial recognition data, national decisions about which administrative data to use to replicate a census, and individual workplaces determining which documents to use to fine-tune customer service chatbots.

Data governance *frameworks* — spelling out *how* these specific decisions about data should be made and challenged, and any requirements around them, such as consultation and transparency — are similarly set at multiple levels, through multinational agreements, national regulation and organizational policies.

These decisions often end up being taken by a narrow community of actors, from closed-door intergovernmental agreements on global policy, to headteachers deciding on which education apps to share pupil data without reference to anyone else. *Inclusive* data governance processes involve multiple stakeholders, giving equal space in this decision making to diverse groups from civil society, as well as space for direct representation of affected communities as active stakeholders.

This links to, but is an idea broader than, the concept of multi-stakeholder governance for technology, which first came to prominence at the international level, in institutions such as the Internet Corporation for Assigned Names and Numbers and the Internet Governance Forum (Hofmann 2016). Some data sources and AI systems do operate at this international level: foundation models, and the data that is required to train them, are the obvious examples. There are a broad range of efforts at this scale: the AI Safety Summit series;[2] the Hiroshima AI Process[3] under the G7; the Global Partnership on Artificial Intelligence;[4] and the UN AI Advisory Board,[5] to name a few. An inclusive approach would mean both that the organization and membership of these groups, events and processes involve a diverse set of civil society organizations, and that they also hear directly from people affected by AI.

However, many more data governance decisions have to be made outside this international context, such as the design of guidance by sector regulators; the selection of training data for fine-tuning by chatbot implementers; and the deployment of data-based systems in schools and workplaces. The same principle of including multiple stakeholders in discussion and decision making can and should still apply in these contexts.

The Democratic Principle

There is a strong democratic case for inclusive governance of data and AI, and specifically to expand the current set of people and organizations making decisions about data and AI to include those affected by these technologies.

Many people and organizations working on social justice or specifically on data and digital rights have adopted the foundational principle of "nothing about us, without us," popularized by the disability rights community.[6] The global Indigenous data rights movement has similarly made the case for Indigenous peoples and nations to not only have the right to govern data, but also to be able to access and use data for their own governance, as a matter of justice.[7] These same arguments hold true for other communities.

Arguments for democratizing AI governance (Seger 2023) point to the power, largely unchecked, wielded by those developing AI technologies in the lab and deploying them in the field. This power manifests itself in the way AI labs have trained foundation models on writings, drawings and photographs without giving creators an opportunity for negotiation. It manifests in how little say we have in what data gets collected by digital services that are almost unavoidable — such as search engines, social media, online shopping and digital public services — and how that data is used and exchanged. And it manifests in the way that technology is rolled out, usually without consultation, in our schools, hospitals, workplaces and local communities.

Within democratic countries, some argue that governments — who are, after all, elected by the people — are able to act as representatives and advance public good in data governance decision making. However, these governments do not only act as democratic representatives and regulators, but they also wield power over their citizens and residents through data and AI, just as they do in authoritarian regimes. Governments collect and steward substantial amounts of the most sensitive data there is, about people's health, education, income, benefits, citizenship and much more. Thanks to the unique power of the state, they also use this data to make life-altering decisions about both individuals and communities. Cases such as Robodebt in Australia[8] or the Dutch child benefit scandal (Heikkilä 2022) are evidence of the damage this can cause.

For this reason, scholars and practitioners have been exploring other mechanisms for the exercise of democratic power — whether realized through direct democracy or through institutions including civil society, independent academia and journalism — to counter that of the companies and governments who control data and AI.

More Specific Arguments

While the general democratic case for including civil society and the public in data governance may appear self-evident to some — particularly those who do not hold power — it is not commonplace in practice, in part because the case for it has not been won.

Involving the public and civil society in decisions about data is not cost-free. Taking the steps that are needed to surmount the practical challenges, and skepticism about the utility of public involvement in a technical and technocratic field, frequently requires arguments that go beyond it being the right thing to do.

Policy makers are particularly concerned that costs and delays from enacting such measures might slow innovation down, diminishing the economic benefits of data and AI, particularly in an internationally competitive context. In this section, the author will therefore explore three specific arguments for different ways of democratizing data and AI that speak to the kinds of outcomes governments and companies care about.

Co-design Reduces Risk

The first argument the author will dig into is for stakeholder involvement in the design of data and AI systems to reduce risks and strengthen the marketplace.

Fitting uses of data and AI to what is expected by and acceptable to the public at the design stage — operating within the social licence for data use (Verhulst, Sandor and Stamm 2023) — reduces risks in many of the same ways as good user needs analysis[9] or human rights impact assessment (Mantelero and Esposito 2021). At an organizational level, getting a wide range of stakeholders involved early:

- reduces the risk of products and services functioning in ways that cause a backlash that may damage user retention or access to public services;
- decreases the risk of wasting time and money developing products and services that are not fit for purpose or need to be withdrawn; and
- lowers reputational risks arising from such backlashes, which may have adverse knock-on effects in a commercial context on share price, future investment or advertising revenue, and in a public sector context on trust in democratic institutions.

At a societal level, having a data ecosystem operating within the social licence also:

- reduces the actual risk of harm to people and communities, and to public goods such as equality; and
- creates a more trustworthy marketplace, enabling organizations to act with confidence and to rely on each other.

To give a couple of public sector examples of operating outside the social licence, in the United Kingdom, public backlashes to data sharing schemes such as the General Practice Data for Planning and Research scheme has led to both costly policy reversals and to patients opting out from sharing health data that is important for medical research.[10] In the Netherlands, the child benefit scandal, where thousands of parents were falsely accused of fraud, brought down the government (Pascoe 2021).

There are also examples from the private sector of problems that could arguably have been caught sooner through broader consultation. In the development of generative AI, where different approaches to fine-tuning — training generative AI about what outputs are most acceptable — have led to the Tay chatbot becoming a "racist asshole" (Vincent 2016) and Google's Gemini depicting multi-ethnic Nazis (Roth 2024). The Facebook-Cambridge Analytica scandal[11] led to Facebook altering and withdrawing a number of its application programming interfaces, illustrating the problem of platform transience (Barrett and Kreiss 2019), where data and AI services can change rapidly in response to external pressure such as scandals, with knock-on effects for those who are using them directly, and those who are dependent on them.

The problem of how to fit the way a service behaves to what is expected and allowed by diverse publics is a challenge social media companies face when automating content moderation worldwide. For those seeking to develop general AI, the same problem is framed as "alignment" — ensuring that future autonomous AI systems protect, rather than destroy, humanity.

Some companies are experimenting with involving a more diverse set of people in making these value-laden ethical decisions. Meta instituted their Oversight Board[12] in May 2020, bringing together experts from a range of countries and backgrounds to make judgments on specific content moderation cases and recommendations for future implementation. In May 2023, OpenAI (2023) announced a US$1 million "Democratic Inputs to AI" grant program, looking for scalable approaches to involving the public in aligning their models to humanity's values, and resulting in 10 pilot projects (OpenAI 2024). Anthropic[13] has partnered with The Collective Intelligence Project (2023)[14] to create Collective Constitutional AI with a similar goal.

While these efforts can be seen as steps in the right direction, in all these cases the company involved retains the locus of power. As Mona Sloane (2024) puts it, "This is a thin form of participation, because participation is limited to existing designs with pre-existing purposes." As a consequence, these initiatives can be seen cynically, as mechanisms to reduce the risk of being held accountable — either in popular opinion or by regulators — for things deemed unacceptable: an arms-length decision-making body, whether a board or a constituted public, can be blamed instead. And they can be seen as ways to reduce the risk of future regulation, which may be more challenging or costly for the organization to enforce.

The risks for people, communities and society, but also for organizations operating within the data and AI marketplace and supply chain, can be reduced through greater inclusion earlier in the design process. But organizational self-interest will not motivate the scope or depth that is required. Reducing the reality and perception of "participation-washing" means requirements for consultation in the design of data and AI systems need to be robust and enforceable. Giving genuine power to those voices helps to ensure that those risks are taken seriously and can help ensure organizational efforts in this space both are, and are seen to be, legitimate.

Civil Society Empowerment Speeds Up Innovation

The second argument the author will examine is to speed up innovation by deferring and focusing regulation through an enhanced and empowered role for civil society.

Regulation of emerging technologies often falls on the horns of the Collingridge dilemma: "Attempting to control a technology is difficult…because during its early stages, when it can be controlled, not enough can be known about its harmful

social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow" (Collingridge, quoted in Genus and Stirling 2018, 63).

This dilemma is particularly apparent for data and AI as they are general-purpose technologies, applied in a vast range of different sectors and contexts. While we might be able to point to some cross-cutting consequences of their adoption, such as on equality or the environment, many impacts are specific to particular types of data such as biometric data; technologies such as generative AI; or the specifics of a given application such as predictive policing. These technologies are also evolving rapidly, and continuing innovation requires regulatory responses that keep a similar pace.

Many policy makers are reluctant to adopt the precautionary principle[15] around the development of data technologies — not allowing them to be made available until they are proven safe — because they fear this will hold back innovation and leave much of the value of data unrealized. Equally, most are now wary of the consequences of entirely permissionless innovation (West 2020) — allowing anything to be built unless and until it is proven harmful — and are cognizant of the need to respond early to emerging harms.

Collingridge's prescription in this case is a middle ground of continuous monitoring and adaptation: an iterative approach to regulation that responds to emerging understanding as well as changing technological capabilities and societal norms.

In the areas where policy makers are shying away from regulating too early, avoiding a de facto permissionless innovation environment requires an approach to data governance that makes good use of civil society. Civil society is uniquely positioned to detect problems with technology early. When workers find themselves unfairly treated, they turn to labour unions; when benefit claimants struggle with new digital public services, they turn to organizations such as Citizens Advice;[16] when consumers are worried about how their data is being used, consumer rights organizations step in. Civil society organizations are the first to hear about the frontline impacts of technology on people, and to start to gather evidence about emerging patterns.

This is one reason why it is so important to include diverse civil society organizations, including those directly in touch with people affected by data and AI, in the high-level multi-stakeholder governance of AI. Civil society is much more directly exposed to the here-and-now effects of data than governments or companies and can bring this experience to the discussion.

But civil society organizations are not just useful for monitoring and understanding the impacts of technology. Civil society *action* can also happen at speed and in a way that prevents overregulation. Organizations may simply self-regulate their use of data and AI in response to being named and shamed, but private and collective legal action is also essential. The degree to which existing regulation covers emerging impacts can be tested in court through strategic litigation. An empowered civil society can thus provide clarity around existing law and identify gaps that require changes to regulation.

This is not to diminish the role of regulators, professional bodies, industry consortia and the legislature. These organizations should be empowered and equipped to respond more quickly and keep up with the pace of change of technology. However, there are natural limits on the ability of these institutions to both be exposed to the impacts of technology on people and communities, and to respond in timely ways. Equipping civil society to act as the canary in the coalmine and alert the wider system to the need for and shape of further regulation, would benefit everyone.

Public Participation Drives Sustainable Adoption

The final argument the author will look at is direct public participation as a mechanism for driving digital, data and AI literacy, and to smooth the path for adoption of technologies with public benefit.

We all have a vested interest in realizing the value of data and reaping the potential societal benefits of AI. While some claims of these benefits may be overblown, it is certainly the case that data could be used for beneficial purposes, such as medical breakthroughs, improving energy efficiency, personalizing learning experiences and so on, as well as bringing economic benefits, such as increasing productivity, spurring innovation and driving economic growth.

Broad and equitable use of these technologies is essential for realizing these positive impacts, including an active market, and that requires people to adopt them. People cannot get the benefit of technologies — either in their day-to-day lives or at work — if they do not know how to use them appropriately, or if they are put off because of concerns about the dangers they might pose. Hence, many governmental data and AI strategies include a focus on improving public understanding, literacy and skills, and building trust.

Getting those affected by data and AI involved in its governance — at all levels — could be an important mechanism for addressing adoption challenges. Being actively involved in shaping the purpose and implementation of technology from the start helps to ensure that it meets the needs of those who will eventually use it, which helps to smooth the way to its adoption. The process of co-design creates a shared understanding of what technology is for, reducing unwarranted distrust about potential ulterior motives.

Being involved in deliberations about data and AI can build literacy and enthusiasm. Like a training course, deliberations provide a supportive peer environment for learning and exploration of a topic. But when charged with making decisions about data and AI, participants take an active role: asking experts questions and digging into areas of uncertainty to help them feel confident about their decision making.

When Connected by Data ran the People's Panel on AI,[17] for example, members of the panel were rapidly exposed to and engaged with various aspects of the impact of AI. They gained confidence, and many became both enthusiastic about its potential and realistic about its downsides. In addition, they took their learning back into their communities, discussing their experience with their family, friends and colleagues.

There are multiple routes for building literacy, trust and adoption. Governments have tended to focus on ones in which the public is a passive recipient: public information campaigns, training programs, audit schemes and kitemarks. A more active role in data and AI governance would be a complementary mechanism that would develop public understanding and adoption through active decision-making power.

Conclusion

The first step in achieving inclusive governance of data and AI is to make the case that it is necessary. While many see the involvement of those affected by technology in its design as a matter of justice, it is also helpful to be equipped with arguments that highlight the advantages of particular modes of engagement with data governance, particularly for stakeholders who may be otherwise unconvinced or unconcerned: the reduction of risk, support for adaptive regulation, and building public understanding and sustainable adoption.

Once these motivating cases are made, and won, the next step is to move on to the harder questions of how. As the examples above have illustrated, there are multiple forms of data governance decisions that need to be made, at multiple governance levels. We need to identify methods of involving multiple diverse stakeholders in these decisions that are practical, cost-effective and provide real power to everyone involved.

Notes

1 See https://understandingpatientdata.org.uk/public-and-patient-engagement-activities.

2 See https://en.wikipedia.org/wiki/AI_Safety_Summit.

3 See www.soumu.go.jp/hiroshimaaiprocess/en/index.html.

4 See https://gpai.ai/.

5 See www.un.org/en/ai-advisory-body.

6 See https://en.wikipedia.org/wiki/Nothing_About_Us_Without_Us.

7 See www.gida-global.org/data-rights.

8 See https://en.wikipedia.org/wiki/Robodebt_scheme.

9 See www.gov.uk/service-manual/user-research/start-by-learning-user-needs.

10 See https://en.wikipedia.org/wiki/General_Practice_Data_for_Planning_and_Research.

11 See https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal.

12 See https://transparency.meta.com/en-gb/oversight.

13 See www.anthropic.com/.

14 See https://cip.org/.

15 See https://en.wikipedia.org/wiki/Precautionary_principle.

16 See www.citizensadvice.org.uk/.

17 See https://connectedbydata.org/projects/2023-peoples-panel-on-ai.

Works Cited

Ada Lovelace Institute. 2021. *Participatory data stewardship*. September. London, UK: Nuffield Foundation. www.adalovelaceinstitute.org/report/participatory-data-stewardship/.

AI Now Institute, Ada Lovelace Institute, Algorithmic Justice League, Alondra Nelson, Camille Francois, Center for Democracy & Technology, Centre for Long-Term Resilience et al. 2023. "AI Now Joins Civil Society Groups in Statement Calling For Regulation To Protect the Public." AI Now Institute, November 1. https://ainowinstitute.org/general/ai-now-joins-civil-society-groups-in-statement-calling-for-regulation-to-protect-the-public.

Barrett, Bridget and Daniel Kreiss. 2019. "Platform transience: changes in Facebook's policies, procedures, and affordances in global electoral politics." *Internet Policy Review* 8 (4): 1–22. https://doi.org/10.14763/2019.4.1446.

Bender, Emily M. 2023. "Talking about a 'schism' is ahistorical." *Medium* (blog), July 5. https://medium.com/@emilymenonbender/talking-about-a-schism-is-ahistorical-3c454a77220f.

Birhane, Abeba, William Isaac, Vinodkumar Prabhakaran, Mark Díaz, Madeleine Clare Elish, Iason Gabriel and Shakir Mohamed. 2022. "Power to the People? Opportunities and Challenges for Participatory AI." In *EAAMO '22: Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, Article No. 6, 1–8. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3551624.3555290.

Chatterjee, Mohar. 2023. "White House notches AI agreement with top tech firms." *Politico*, July 21. www.politico.com/news/2023/07/21/biden-notches-voluntary-deal-with-7-ai-developers-00107509.

Genus, Audley and Andy Stirling. 2018. "Collingridge and the dilemma of control: Towards responsible and accountable innovation." *Research Policy* 47 (1): 61–9. https://doi.org/10.1016/j.respol.2017.09.012.

Heikkilä, Melissa. 2022. "Dutch scandal serves as a warning for Europe over risks of using algorithms." *Politico*, March 29. www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/.

Himmelreich, Johannes. 2022. "Against 'Democratizing AI.'" *AI & Society* 38 (1): 1333–46. https://doi.org/10.1007/s00146-021-01357-z.

Hofmann, Jeanette. 2016. "Multi-stakeholderism in Internet governance: putting a fiction into practice." *Journal of Cyber Policy* 1 (1): 29–49. https://doi.org/10.1080/23738871.2016.1158303.

Mantelero, Alessandro and Maria Samantha Esposito. 2021. "An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems." *Computer Law & Security Review* 41 (1): 105561. https://doi.org/10.1016/j.clsr.2021.105561.

OpenAI. 2023. "Democratic inputs to AI." OpenAI, May 25. https://openai.com/index/democratic-inputs-to-ai/.

———. 2024. "Democratic inputs to AI grant program: lessons learned and implementation plans." OpenAI, January 16. https://openai.com/index/democratic-inputs-to-ai-grant-program-update/.

Pascoe, Robin. 2021. "Dutch government collapses in fall out from child benefit scandal." DutchNews, January 15. www.dutchnews.nl/2021/01/dutch-government-collapses-in-fall-out-from-child-benefit-scandal/.

Roth, Emma. 2024. "Google explains Gemini's 'embarrassing' AI pictures of diverse Nazis." *The Verge*, February 23. www.theverge.com/2024/2/23/24081309/google-gemini-embarrassing-ai-pictures-diverse-nazi.

Sadowski, Jathan, Salomé Viljoen and Meredith Whittaker. 2021. "Everyone should decide how their digital data are used — not just tech companies." *Nature* 595 (7866): 169–71. https://doi.org/10.1038/d41586-021-01812-3.

Seger, Elizabeth. 2023. "What Do We Mean When We Talk About 'AI Democratisation'?" *Centre for the Governance of AI* (blog), February 7. www.governance.ai/post/what-do-we-mean-when-we-talk-about-ai-democratisation.

Shubladze, Sandro. 2023. "Unlocking The Data Revolution: AI's Future Shaped by Public Data." *Forbes*, July 5. www.forbes.com/sites/forbestechcouncil/2023/07/05/unlocking-the-data-revolution-ais-future-shaped-by-public-data/.

Sloane, Mona. 2024. "Controversies, contradiction, and 'participation' in AI." *Big Data & Society* 11 (1): 1–5. https://doi.org/10.1177/20539517241235862.

The Collective Intelligence Project. 2023. "CIP and Anthropic launch Collective Constitutional AI." *The Collective Intelligence Project* (blog), October 17. https://cip.org/blog/ccai.

Verhulst, Stefaan G., Laura Sandor and Julia Stamm. 2023. "The Urgent Need to Reimagine Data Consent." *Stanford Social Innovation Review*, July 26. https://ssir.org/articles/entry/the_urgent_need_to_reimagine_data_consent.

Vincent, James. 2016. "Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day." *The Verge*, March 24. www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

West, Darrell M. 2020. "The end of permissionless innovation." Brookings, October 7. www.brookings.edu/articles/the-end-of-permissionless-innovation/.

Zorthian, Julia. 2023. "OpenAI CEO Sam Altman Asks Congress to Regulate AI." *Time*, May 16. https://time.com/6280372/sam-altman-chatgpt-regulate-ai/.

About the Author

Jeni Tennison is a CIGI senior fellow and the founder of Connected by Data, a campaign that aims to put community at the centre of data narratives, practices and policies. She is an adjunct professor at the University of Southampton's Web Science Institute and a Shuttleworth Foundation Fellow.

Jeni was CEO at the Open Data Institute, where she held leadership roles for nine years and worked with companies and governments to build an open, trustworthy data ecosystem; a co-chair of the Global Partnership on Artificial Intelligence's Working Group on Data Governance; and an associate researcher at the University of Cambridge's Bennett Institute for Public Policy. She sits on the boards of Creative Commons and the Information Law and Policy Centre.

She has a Ph.D. in artificial intelligence and an Order of the British Empire for services to technology and open data. She loves Lego and board games and is the proud co-creator of the open data board game, Datopolis.

A Systems Approach to Data Governance: The Global Platform Governance Model

Chris Beall

Aristotle is often (mis)quoted as writing, "the whole is greater than the sum of its parts" in his work *Metaphysics*. Just as nineteenth-century Gestalt psychology focused on understanding perception as a whole and not just its individual components, this quotation has direct implications for governing our global information environment,[1] of which data is a critical element. We need to stop looking only at the system's components in isolation and instead address the interplay between the various elements in order to understand how to enable the fair and open information system on which our democracies rely. But notably missing from many of these discussions is the need for a fair, inclusive approach to data governance that supports and strengthens the whole of society's engagement in a healthy information space.

The ground has shifted significantly since the Centre for International Governance Innovation (CIGI) released its Data Governance in the Digital Age series in 2018. The last six years have witnessed a slow convergence around aspects of a platform governance agenda, encompassing a wide range of technology, social and economic policies, legislation and regulation that touch on aspects of the global information environment. Governments, regulators, civil society organizations (CSOs) and even some industry partners have meaningfully worked to address the impacts of data, systems and online content on society, notably recognizing the role that very large online platforms and search engines[2] continue to play in this regard.

Sadly, however, little has practically changed. Despite substantial investments of time and money, the global outlook for an open and inclusive information environment, grounded in fair, equitable data access and usage, has worsened. As UN Secretary-General António Guterres recently expressed, in a statement that could have been uttered in 2017, "misinformation, disinformation and hate speech are fueling prejudice and violence; exacerbating divisions and conflicts; demonizing minorities; and compromising the integrity of elections" (Guterres 2024). Indeed, in many ways, bridging privacy, competition, security and safety, and data governance spaces feels less attainable now than it did six years ago. We urgently need to recognize that the information

environment is an interconnected, complex system that requires a comprehensive governance approach at both the international and domestic levels. We need to work together collaboratively to address current challenges within that system, while acknowledging the varied strengths, knowledge and experiences each partner brings to the table. Experiences and lessons learned from CIGI's Global Platform Governance Network (GPGN) offer some useful insights into how we can better achieve this.

While responsibility for the current situation sits with democratic partners' inaction, recent technological and societal developments have intensified the problem. Our increased reliance on online spaces during and since the global pandemic; growing authoritarian influence and on- and offline threats to global stability emerging from regional conflicts around the globe; and the meteoric rise of generative artificial intelligence (AI), built, in part, on massive levels of data capture and exploitation, have further fuelled the problems faced in 2018.[3] As a result, coherent, effective governance is in retreat on a myriad of fronts, from defending democratic institutions and elections, to protecting individual privacy and security, to ensuring fair and equitable trade and market access in the digital sphere.

Project CONNIE (Collaboration & Open Networking Needs for the Information Environment) is working to address this challenge. With support from the Omidyar Network, CONNIE is developing a new way to connect the myriad global networks that are working to support the integrity of the open, inclusive information space, including those focused primarily on data governance. Rather than create another network, CONNIE takes a "federated" approach, enabling partners to collaborate while retaining their autonomy and working from where they are. Taking a systems approach, CONNIE will connect current and planned activities to better enable meaningful collaboration across the information system. This approach deliberately encompasses and aligns with emerging efforts to create a data governance regime that spans and respects different jurisdictions, values and approaches, while building common ground to work together toward shared goals.

Indeed, like its subset data governance, the broad information environment cuts across a wide swathe of policy, operational and geographic divisions, each with its own unique history, community and shared experience. Public pressure to address high-profile crises across the information environment, such as coordinated, live-streamed terror attacks, foreign influence operations and disinformation, and the global collapse of independent media, have resulted in myopic and siloed solutions. Instead of considering each of these issues as part of a broader, interconnected system, democracies have tackled only specific aspects of the problem, focusing almost exclusively on the content layer of the online space, tackling challenges within domestic or regional borders and targeting symptoms such as disinformation. The absence of data governance perspectives in these largely responsive exercises is striking. When partners connect across policy and geographic silos to tackle emerging challenges in the information space, issues related to data capture, sharing, access and ownership are often excluded or addressed separately in parallel conversations; crises resulting from unfettered data extraction and exploitation rarely feature. As a result, governments, regulators, CSOs and donors continue to miss opportunities to work together, learn from one another and find new ways to approach their long-standing, seemingly intractable problems.

Trust, capacity, taxonomy and language, and urgency have all played key roles in leading us to this place. Many of the network initiatives that were created to link allies have focused on connecting across existing spaces and on aligning with known partners, largely within their own policy domains, rather than recognizing the gaps and working across the divides. This approach has led the fissures between these communities to become gaping chasms. As a unique subset of the problem, data governance poses a special challenge: despite being the fuel for the current AI boom, as well as the

foundation on which the information system runs, data is often the missing element of these conversations.

There continues to be a flurry of activity, as politicians, regulators, academics, civil society, donors and industry have raced to "do something" to address seemingly endless crises. Since 2018, we have seen the launch of numerous solutions designed to tackle various aspects of these governance challenges. Coupled with the continued emergence of new avenues of scholarship and the launch of new networks to support that work, however, the pace and volume of this activity is overwhelming.

In fact, what is striking is the sheer *number* of activities under way globally, both across related networks and communities and in disconnected policy areas working in parallel on related issues. There are far too many new initiatives, reports, research projects, networks and partnerships for any one person to track who is doing what, where they are doing it and which approaches they are using, let alone trying to distill what the global community could be learning as a result. This lack of visibility and coordination between project teams means that each fails to take advantage of possible economies of scale and scope across current efforts. This has led to significant risk of duplication or initiatives operating at cross purposes. It has meant that bright spots are missed and partners with promising initiatives, especially outside the Global North, can struggle to find support. Without a process to bring these ideas together and to build from them, we are missing opportunities to accelerate emerging successes and identify cul-de-sacs. Considering this critical challenge that democracies are facing, our current structure is insufficient to handle the task.

A Potential Way Forward

Drawing upon the knowledge and experiences gained through the two years of GPGN operations (2020–2022) offers a potential way forward on how global governance for data may be created.

The GPGN's starting premise was that individual government, civil society and multilateral partners cannot solve these problems on their own; a global problem requires a global solution. Partners need to work together to explore ways in which they can ensure effective governance for the information space. Traditional governance tools — legislation and regulation — operate in domestic spheres and reflect domestic interests. Furthermore, for many governments, media regulation, counterterrorism, technology policy, digital economy and data governance sit in different subject matter and operational areas and are overseen by different policy and legislative committees that ordinarily have little reason to interact. In fact, GPGN members recognized that the only people who had a complete picture of their governments' operations were the tech industry's public policy teams who worked with them.

With support from Reset and the Balsillie Family Foundation, the GPGN brought together a community of individuals working on a broad range of issues under the platform governance umbrella from countries whose values and experiences were broadly aligned and who were willing to share their successes and failures to build toward a better future.

In early 2021, the GPGN created three working groups to address questions of measuring the impact of interventions, aligning transparency reporting and collaborating on government research efforts. Each group was led by a global practitioner and worked with subject matter experts and government specialists to drive collaborative action on legislation, regulation and policies that were currently under development in home jurisdictions. The transparency working group's report, *Transparency Recommendations*

for Regulatory Regimes of Digital Platforms (MacCarthy 2022), provided practical advice on building and aligning regulatory frameworks, with at least one government regulator described using the report's key questions on a regular basis in her organization's working sessions.

Lessons Learned

Creating "networks of networks" to achieve a whole-of-society solution continues to resonate globally, as democratic partners think through how to lean on the kind of multi-stakeholder network collaboration that the GPGN fostered.[4] It is worth noting that meaningful collaboration that aligns governance efforts without relying on large-scale formal agreements and cumbersome infrastructure is difficult to achieve. Without careful, dedicated attention, promising horizontal initiatives can falter.

Key to successfully aligning efforts across silos is building trust, finding genuine connection, respecting and listening to differences, and being willing to share and learn from failure as well as success. Through its work, the GPGN identified several key issues for future governance efforts that need to be emulated in future attempts to align global governance in this space.

Platform/Data Governance and the Information Space Are Broader than We Think They Are

The impetus behind the GPGN's work came from the recognition that many of the government actors working on aspects of digital platform governance were unaware of or disconnected from work under way in other jurisdictions or in other policy areas within their own governments. The GPGN recruited members from a broad range of digital policy areas (including countering violent extremism and online harms, digital trade and data governance, and protecting democracy) and from a wide range of 25+ countries, as well as a selection of multilateral organizations (for example, the Organisation for Economic Co-operation and Development [OECD]; the UN Educational, Scientific and Cultural Organization; UN Trade and Development; and the Council of Europe). In practical terms, collaboration across sectors and geographies often provided members with tested, usable advice that they could put into practice in their own jurisdictions.

Key Lesson Learned

A greater variety of perspectives yields alternative approaches and useful workarounds to address apparently intractable challenges.

People and Relationships Matter More Than Position or Level

The GPGN benefited initially from 2020's global lockdowns. It meant that people were available; even those who usually faced significant travel barriers were grounded. Virtual meetings allowed more regular conversations that effectively built community: it is far easier to schedule a one-hour meeting than to carve out the time, expenses and logistics of in-person conversations. Operating virtually meant that the network could include a greater number of Global Majority voices, as participation did not require the costs of in-person travel.

Recognizing that the right people matter regardless of their current positions, the GPGN used a snowball sample approach to build its community, leveraging trusted relationships. The team started by reaching out to colleagues from governments and regulators in North and South America, Europe and the European Union, Asia and Asia-Pacific, and Africa. They also contacted existing networks operating in parallel that were already active in this space, including the Group of Seven Rapid Response Mechanism's country leads; the OECD's community working to advance transparency reporting on terrorist and violent extremist content online; the Institute for Strategic Dialogue's Digital Policy Lab; the Christchurch Call network, the Global Internet Forum to Counter Terrorism; the Carnegie Endowment for International Peace's Partnership for Countering Influence Operations; and the Internet & Jurisdiction Policy Network.

Rather than target high-profile heads of organizations, the GPGN approached officials at the junior or mid-range executive levels, onboarding those who were senior enough to be able to effect meaningful change, but close enough to the work to not need briefing ahead of each meeting or to require talking points. The unique makeup of the working groups brought together those "holding the pen" on drafting policy, legislation and regulation with their colleagues in other jurisdictions, alongside the legislative staff and regulators who would ultimately be enacting and overseeing their work. This gave members the opportunity to work through problems behind closed doors that would have otherwise either created lingering resentments or led to serious log jams. One of the most exciting conversations came about when the regulator and civil service members in the transparency working group nearly talked over one another in sharing their personal experiences and difficulties with drafting and applying new legislation.

Key Lessons Learned

- Building relationships among those accountable for results yields meaningful collaboration that can effect real change.
- Work with the right people, not the right positions.

People Need a Safe Space to Connect

From the start, the GPGN recognized the value in limiting its membership to government (broadly defined) and a small number of outside partners. Whether they were regulators, civil servants or political staff, network members reported feeling that they could speak openly about their challenges and successes without needing either to explain "how government works" or sharing too much with organizations that might use this information for their own ends. Although bringing together a broad array of partners through a whole-of-society approach is essential to addressing the challenges that we face, the GPGN recognized early on that there was nowhere for government representatives to build connections that did not also include either industry representatives or those seeking external funding for their projects.

Network members also reported being heard and not simply being talked at. The group intentionally did not open up space for transatlantic superpowers to push their agendas. Instead, the GPGN tried to apply a global focus, bringing together a wide range of experts, including some from the Global Majority, to ensure that all members understood the benefits of creating tools that had global application. For example, immediately following the January 6 insurrection, American colleagues were able to confide in and learn from their South American and African colleagues who had worked through similar crises in the past. Subsequent work with colleagues in Ukraine and Sub-Saharan Africa validated this working approach by highlighting the frustrations that partners had previously found in having Global North experts parachuting in to offer largely academic perspectives.

Key Lessons Learned

- Truly safe spaces allow people to share their setbacks and concerns instead of just focusing on successes.
- In many cases, Global Majority partners have been facing these issues for longer periods of time and have developed effective, nuanced solutions that can be applied more broadly.

Allowing All Partners to Maintain Their Autonomy

The GPGN aimed to demonstrate the importance of building partnerships among individuals approaching these issues from different perspectives, including from parallel disciplines, regions or political approaches. Members of the network recognized the value in hearing about others' experiences with similar challenges in other regions and from other policy areas (for example, network members working to combat online hate and learning about regulatory approaches taken to address data privacy). Others noted that the network meetings were sometimes the only place where they could speak openly with legislative staff, regulators and civil servants from their own countries. Some were able to use the meetings to build their own internal bridges within their own bureaucracies, having learned about initiatives moving forward in other parts of their own governments through network conversations. The GPGN was asked by members to enable their leaders to meet through international visits and to foster collaboration on new legislative measures.

There is a vast array of actors working in this space. Instead of either duplicating efforts or attempting to crowd out existing players, the network made a concerted effort early on to work with as many partners as possible to build on their successes, learn from their mistakes and amplify their important work. When learning of a new, innovative approach to one of the issues covered by network members, the GPGN team would invite the researchers to present their findings to the network. For example, when launching its working groups, the network recruited leading experts to enable these groups to build on work already under way and contribute to global conversations. Leveraging the Atlantic Council's previous work on transparency, for example, provided regulators who were designing transparency mechanisms in their own countries with a set of questions that they could use to help focus efforts. Similarly, when the performance measurement group was tackling the challenge of identifying "success" in this space, the GPGN worked with the Social Science Research Council to apply their HuMetrics tool, initially designed to identify values-based approaches for universities (Agate et al. 2020), to help members think through the overall goals of their own national efforts.

Key Lessons Learned

- Being able to spot a problem does not mean that you are the right organization to address it.
- Before starting something new, it is always worth determining whether others have already been working to address similar challenges.
- We do not need to conform or entirely align with one another to achieve related goals: recognizing and building from difference is a strength.

Where Do We Go from Here?

The pressure of addressing what feels like relentless, ongoing crises across the information space, coupled with historic, structural and vertical governance and accountability issues, has democratic governments and CSOs largely tackling information threats in a disjointed fashion, often ignoring underlying data governance challenges. Connecting well-intentioned government and CSO networks across geographic, policy and operational silos, as well as creating the space for community building, information sharing and learning, is needed to direct global efforts toward a coherent, system-wide response. We must ensure an open, inclusive democratic future for the information environment, with effective data governance at its centre. To borrow terminology from law enforcement, collaboration, when well-managed, can create a "force multiplier effect." Or, in musical terms, what makes a choir great is not the best voices, but instead the ability of its members to work together to create something that exceeds the sum of their parts.

To that end, we need to double down on efforts to bring together those working across policy, operational and geographic divides, and especially to ensure a central place for data governance in those conversations. Aligning governance efforts focused on all aspects of the information environment is essential to effecting meaningful, measurable change. Based on learning from the GPGN and subsequent efforts to bridge divides, it is CONNIE's intention to help by linking activities and plans, reducing duplication, supporting faltering efforts, and amplifying and celebrating emerging bright spots.

Democratic societies must move past tactical-level decisions aimed at developing individual solutions to global systems challenges. The current approach risks putting the underlying goals of a free, open, transparent and democratic online (and offline) space at risk. As with the wars on drugs and terrorism, the time has come in which democratic societies have to discard their initial approaches and reset for what lies ahead. We need to stop reacting to the past in order to build what we want to achieve together for the future.

Notes

1 The information environment is defined as "the space where human cognition, technology, and content converge" (Wanless and Shapiro 2022, 3).

2 See the European Commission's definition: https://digital-strategy.ec.europa.eu/en/policies/dsa-vlops.

3 As articulated, for example, by Taylor Owen (2018).

4 Information in this section is drawn from interviews conducted with GPGN members between 2022 and 2023 and enhanced by the author's experience leading the taxonomy working group for the 2023 Summit for Democracy's Information Integrity Cohort, and the Carnegie Endowment for International Peace's multi-stakeholder crisis response network supporting the integrity of Ukraine's information environment in 2022–2023.

Works Cited

Agate, Nicky, Rebecca Kennison, Stacy Konkiel, Christopher P. Long, Jason Rhody, Simone Sacchi and Penelope Weber. 2020. "The transformative power of values-enacted scholarship." *Humanities and Social Sciences Communications* 7. https://doi.org/10.1057/s41599-020-00647-z.

Guterres, António. 2024. "Video Message." UN DESA Global Forum on Data Governance and Digital Transformation UNU Artificial Intelligence Conference. April 24. www.un.org/en/desa-en/un-desa-global-forum-data-governance-and-digital-transformation-unu-ai-conf.

MacCarthy, Mark. 2022. *Transparency Recommendations for Regulatory Regimes of Digital Platforms*. Conference Report, GPGN Transparency Working Group. Waterloo, ON: CIGI. www.cigionline.org/publications/transparency-recommendations-for-regulatory-regimes-of-digital-platforms.

Owen, Taylor. 2018. "Ungoverned Space: How Surveillance Capitalism and AI Undermine Democracy." In *Data Governance in the Digital Age*. March 20. Waterloo, ON: CIGI. www.cigionline.org/articles/ungoverned-space.

Wanless, Alicia and Jacob N. Shapiro. 2022. *A CERN Model for Studying the Information Environment*. November. Washington, DC: Carnegie Endowment for International Peace. https://carnegieendowment.org/research/2022/11/a-cern-model-for-studying-the-information-environment.

About the Author

Chris Beall has created and leads Project CONNIE, a new initiative connecting the parallel, often disconnected global networks supporting the integrity of the democratic information environment. CONNIE will provide "intelligent facilitation," linking current and planned activities and capacity-building efforts across policy, geographic and operational silos, reducing duplication, amplifying emerging bright spots and bolstering faltering efforts.

Previously, Chris was a senior fellow at the Carnegie Endowment for International Peace, where he led a multi-stakeholder crisis response network bringing together civil society, government and industry partners working to protect the integrity of Ukraine's information environment. He also led the taxonomy working group for Latvia and Canada's 2023 Summit for Democracy cohort on information integrity, and both created and led the GPGN at CIGI. The GPGN brought together civil servants, regulators and legislative staff from democracies around the world addressing aspects of digital platform governance. Chris has held leadership positions in a range of departments and agencies within the Government of Canada, including as the founding director of the Digital Citizen Initiative at Canadian Heritage. He holds a doctorate from the University of Oxford.

Trade Agreements and Data Governance

Patrick Leblond

Owing to their rapid rise in the last decade or so, cross-border data flows and digital trade are increasingly becoming governed by trade agreements. This is because national regulations restricting the flow of data (personal, business and government) across borders are considered an important impediment to trade (Aaronson 2019; Cory and Dascoli 2021). For instance, Magnus Rentzhog (2015), in a study of Swedish companies from a wide range of sectors, found that moving data across borders easily was crucial for the well-functioning of these firms' global value chains. Restrictions on cross-border data flows are particularly problematic for trade in services (Ferracane and van der Marel 2021).

Trade agreements recognize that policy makers face a tension between, on the one hand, generating the economic benefits associated with unfettered data flows across borders and, on the other hand, providing a trusting environment for individuals, firms and governments to conduct their business. They aim to ensure that national regulations affecting data flows are not disguised protectionist measures that discriminate against foreign providers of digital goods and services in favour of domestic ones.[1] As such, the core principles of national treatment, most favoured nation and transparency apply here as well.

Although trade agreements have the potential to limit the ability of governments to regulate data and the digital economy domestically (Leblond 2021a), they have not prevented national regulation from impeding cross-border data flows between their member states so far. For instance, data-localization measures have been increasing rather than decreasing in the last decade (Cory and Dascoli 2021; Organisation for Economic Co-operation and Development [OECD] 2023). Moreover, they have become more restrictive: "more than two-thirds of identified measures involved a storage requirement with a flow prohibition" (OECD 2023, 18). National data-flow regulations are also becoming "increasingly complex and fragmented" (ibid., 17). Even if they do not prohibit the flow of data across borders, they make it more costly for firms in terms of compliance, thereby hurting international trade (Evenett, Fritz and Giardini 2023).

There are more and more trade agreements with provisions addressing international data flows; however, they have not been effective at fostering data free flow with trust in support of international trade (World Economic Forum 2020). This essay explains why and offers an alternative way forward.

Are Trade Agreements Effective at Governing Cross-Border Data Flows?

This section examines provisions regarding cross-border data flows found in trade agreements. To begin, it is worth mentioning that there has been a moratorium on imposing customs duties on electronic transmissions at the World Trade Organization (WTO) since 1998.[2] All trade agreements with an e-commerce or digital trade chapter also contain an article that prohibits the imposition of customs duties on electronic transmissions.

The Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP)[3] is the first major trade agreement with provisions on cross-border data flows. It prohibits restrictions on cross-border data transfers for business purposes and requirements to localize the storage of data domestically, respectively (articles 14.11 and 14.13).[4] It does not specify what types of data are covered, except to say those that are necessary for business purposes. However, articles 14.11 and 14.13 also allow the parties to impose restrictions on data transfers to achieve a "legitimate public policy objective." Restrictive measures on cross-border data flows cannot, however, be disguised protectionism that favours one or a set of domestic firms at the expense of their foreign competitors. Moreover, any restriction must be commensurate with the objective that it is meant to achieve; it cannot be stronger or more encompassing than what is strictly required to be effective (the necessity test).

The CPTPP also aims to ensure that signatories have laws and regulations that provide a minimum level of personal information protection (article 14.8). However, the provisions are very flexible in terms of accommodating different national approaches. They simply call on the parties "to take into account principles and guidelines of relevant international bodies,"[5] which is a well-established approach in trade agreements, without specifying any particular body.

Thus, there is a fair degree of ambiguity as to the extent to which the CPTPP prevents governments from imposing restrictions on data flows between member states. Ultimately, it is left to the CPTPP's state-to-state dispute settlement mechanism (DSM) to decide. And, given the absence of internationally agreed regulatory standards, the basis for a DSM panel decision is uncertain.

The CPTPP served as the template for the United States-Mexico-Canada Agreement (USMCA).[6] The USMCA adds to the CPTPP in that, with respect to the requirement to have a legal framework to protect personal information, it mentions explicitly the Asia-Pacific Economic Cooperation (APEC) Privacy Framework and the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.[7] Like the CPTPP, however, the USMCA parties are not required to take them into account; they are only encouraged to do so.

In addition, the USMCA provides some limit on the extent to which data protection legislation or regulation can constrain the transfer of data between the member states: restrictions must be "necessary and proportionate to the risks presented."[8] The CPTPP has no such limit. Nevertheless, it still leaves open the question of determining if and to what extent restrictions are "necessary and proportionate." As with the CPTPP, answering such a question is left to a panel under the USMCA's DSM (chapter 31). Hence, the USMCA does not really reduce the uncertainty relating to restrictions on cross-border data flows that could be imposed by national data-protection laws and regulations.

Finally, unlike the CPTPP, the USMCA does not allow the parties to invoke a "legitimate public policy objective"[9] exception to impose data-localization requirements to firms

from the other two parties to the agreement (except, like with the CPTPP, when a digital good or service is provided to a government). Here, the USMCA is clearer in supporting cross-border data flows than the CPTPP. However, as always, the effectiveness depends on the parties' willingness to pursue a dispute under chapter 31.

If the USMCA has some stricter provisions than the CPTPP regarding data flows, the more recent Regional Comprehensive Economic Partnership (RCEP) has looser ones.[10] The RCEP's language is such that it allows member states to impose whatever national regulatory restrictions on the flow of data across borders. With respect to provisions in favour of allowing the cross-border transfer of information by electronic means, the RCEP retains the CPTPP's "legitimate public policy" exception, including for data-localization measures, but makes it self-judging. This means that any measure restricting the cross-border flow of data is legitimate if a party to the RCEP says so. And the other RCEP members cannot dispute it: first, because it is explicitly stated that they cannot; and second, because the RCEP's state-to-state DSM does not apply to chapter 12 on digital trade (unlike the CPTPP and USMCA).

If the RCEP makes toothless the provisions prohibiting restrictions on cross-border data transfers for business purposes as well as requirements to localize the storage of data domestically, the recently released "stabilized" text of the WTO's plurilateral agreement on electronic commerce simply ignores them.[11] Compared to the RCEP's fudged approach, it has the benefit of being honest about the fact that the more than 80 parties to the negotiations cannot agree on how to address cross-border transfers.

Until recently, the European Union did not include provisions on cross-border data in its trade agreements, other than the prohibition of customs duties on electronic transmissions. This is because international flows of personal data are covered by the General Data Protection Regulation (GDPR).[12] Under the latter, the European Union allows for personal data to flow freely with countries whose personal data (or privacy) protection regimes are deemed "adequate" by the European Commission.[13] In cases where the regime in another country outside the European Union is not considered adequate, organizations can still flow personal data out of the European Union if appropriate safeguards vis-à-vis the organization receiving the personal data are in place. These safeguards can be provided to the European Union through approved standard contractual clauses, binding corporate rules, codes of conduct and certification mechanisms.

This is why both the Canada-EU Comprehensive Economic and Trade Agreement[14] and the EU-Japan Economic Partnership Agreement (EPA)[15] did not include provisions on the free flow of data. In the latter case, article 8.81 originally stipulated that the two parties would reassess the situation within three years after it entered into force. On October 28, 2023, the European Union and Japan announced that they had concluded a deal on cross-border data flows to amend the EPA (European Commission 2023). Under this amendment, the European Union and Japan "shall not adopt or maintain measures which prohibit or restrict the cross-border transfer of information."[16] It also provides an exhaustive list of such measures (for example, no data localization, no requirement to use computing facilities in one of the parties' territory, no need for authorization to transfer data to the other party's territory). Like other trade agreements, the amendment contains an exception clause to pursue a "legitimate public policy objective" if it is non-discriminatory and satisfies a necessity test. It also states that personal data-transfer instruments, such as those found in the GDPR for safeguarding, are allowed so long as they are generally applied.

At the end of November 2023, the European Union announced that it had agreed to a digital partnership with Canada. In this case, the text is notable for the fact that it does not say anything about the free flow of data other than "both sides intend to

exchange information on their respective data governance frameworks and discuss the interoperability of Canadian and EU Data Spaces."[17] The difference in the European Union's approach to Canada and Japan with respect to cross-border data flows remains a mystery at the time of writing, especially since the EU-New Zealand trade agreement, which came into force on May 1, 2024, has provisions on cross-border data flows that are in line with the EU-Japan EPA amendment (and all three countries' data protection and privacy protection regimes are deemed adequate by the European Commission).

On April 21, 2022, the Global Cross-Border Privacy Rules (CBPR) Forum was established by Canada, Japan, the Republic of Korea, the Philippines, Singapore, Taiwan and the United States, although the United States is the lead on this initiative.[18] The CBPR Forum is meant to promote "trusted global data flows" by building on the APEC CBPR System, from which firms can voluntarily obtain data-privacy certifications that demonstrate their compliance with the privacy rules. With this certification, firms can transfer personal data freely between the forum's members. The forum is also expected to pursue interoperability with other data protection and privacy regimes such as the GDPR.

Another recent significant trade agreement governing cross-border data flows is the Digital Economy Partnership Agreement (DEPA) set up by Chile, New Zealand and Singapore in June 2020 (it entered into force in January 2021).[19] It has been dubbed the "world's first digital-only trade agreement" (Taheri, Adams and Stern 2021). It consists of 16 modules to facilitate digital trade and cooperation on digital issues and technologies. For the most part, these modules "adopt or refine existing measures addressing digital trade facilitation," especially from the CPTPP, to which all three countries also belong (Ciuriak and Fay 2022, 3). Module 4, which deals with data issues, is a good example of a module that adopts existing CPTPP commitments (on cross-border transfer of information by electronic means and on location of computing facilities without going further); however, it does not go beyond those previous commitments (ibid., 4).

In sum, trade agreements governing cross-border data flows are a growing "digital noodle bowl" that makes it increasingly difficult for firms, especially small and medium-sized ones, to keep up with (Honey 2021). These agreements' ability to ensure the free flow of data across borders is also in question because of their ambiguous language, which ultimately leaves it to a few people on dispute-settlement panels (when those are made available by an agreement) to decide what are "legitimate public policy objectives" and "necessary and appropriate" measures. Perhaps not surprisingly, no dispute on digital trade matters has yet been initiated under any agreement. This is likely due to the parties preferring to leave things as they are: better the devil you know (with the national regulatory space that goes with it) than the one you do not know (i.e., regulations and standards decided by non-democratically elected panel arbitrators that significantly restrict national regulatory space). Finally, for the most part, trade agreements governing cross-border data flows do not have provisions for the parties to develop common standards and regulations, only referring to the latter or encouraging the parties to have them in place nationally.

Is There a Better Way to Govern Cross-Border Data Flows?

The growing digital noodle bowl leads to two unsatisfactory scenarios. In the first scenario, member states are allowed to adopt whatever regulations they deem necessary to protect individuals, consumers, businesses and governments at the national level, but at the expense of cross-border data flows. In the second scenario, data is freer to flow across borders (as a result of trade agreements limiting national data regulations' scope of applicability), but at the expense of trust in data-driven markets.

These two scenarios are derived from Patrick Leblond and Susan Ariel Aaronson's (2019) data trilemma, which states that the following three elements cannot hold simultaneously: free flow of data across borders; national data protection laws and regulations that are distinct from those of other countries; and a high level of trust in the data environments among individuals, consumers, businesses and governments. Only two of the three elements can occur at the same time. Strong national data protection laws and regulations should lead to high trust levels but, to do so, they risk imposing restrictions on cross-border data flows. Alternatively, if policy makers want to ensure the free flow of data across borders while maintaining national data policies, then they may have to accept weaker data protection measures, which could negatively affect trust. Finally, if policy makers want data to flow freely across borders while ensuring a high degree of trust surrounding the collection and use of data, then they either adopt another jurisdiction's regulatory standards (in order for data to flow freely with this jurisdiction and others with the same recognized standards), or they cooperate with governments in other countries to develop and enforce common, high-quality protection standards and regulations for personal as well as non-personal data.

Leblond and Aaronson (ibid.) argue that the best approach to obtaining freely flowing data across borders and high trust levels among consumers, businesses and governments in data-driven markets is to create a single data area with its own standard-setting and monitoring body, which could be called the "International Data Standards Board" (IDSB). Such an IDSB would be responsible for setting standards that regulate the creation, processing, use, distribution and transfer of data, both personal and non-personal, within the single data area. It would also be responsible for monitoring that the states that are members of the single data area apply and enforce the common standards adequately. IDSB members would allow data to flow freely between them as they would apply the same standards as well as cooperate closely in terms of not only developing standards but also sharing information and enforcing compliance. This single data area would welcome and support (financially and technically) any country willing to adopt and enforce the common regulatory standards. The IDSB's frequent assessments would determine if a member state is able to continue taking full part in the single data area or not. In case of inadequate application or enforcement, the other members of the single data area would be allowed to restrict data flows to the member state that is not in good standing until proper actions have been taken to remedy the situation.[20] By developing common standards that are effectively enforced, the member states of a single data area managed by an IDSB would overcome the ineffectiveness of existing trade agreements to flow data (personal and business) freely across borders with trust, and thus allow them to derive the economic benefits associated with the data flows.

Such an international body should draw inspiration from international financial standard-setting bodies such as the Basel Committee on Banking Supervision, the International Organization of Securities Commissions and the International Accounting Standards Board (Leblond 2021b). The main challenge is how to get an IDSB and its attendant single data area off the ground.

One could envisage a Bretton Woods-like international conference to set up an IDSB. After all, the Bretton Woods conference led to the creation of three international bodies: the International Monetary Fund, the World Bank and the International Trade Organization (whose charter was never ratified by the Americans). Admittedly, today's cross-border data context does not compare to the world economic situation that prevailed at the end of the Second World War. This is why the probability of an international conference à la Bretton Woods to create an IDSB is low for the foreseeable future.

A more promising avenue to get things off the ground would be for the Financial Stability Board (FSB), which is responsible for coordinating international financial standards across Group of Twenty countries and works in tandem with the above-mentioned organizations, to set up an international body responsible for making national data regulatory regimes interoperable so that data can cross borders securely and with minimal friction for cheaper and faster payments globally. Successful international cooperation in financial services, notably payments, could then provide the basis for a broader application to other industrial sectors since they face similar data-transfer issues. The FSB is already looking into "developing recommendations [...] for promoting alignment and interoperability across data frameworks applicable to cross-border payments, including data privacy, operational resilience, AML/CFT [anti-money laundering/counterterrorism financing] compliance and regulatory and supervisory access requirements" (FSB 2023, 2). Perhaps the creation of an IDSB by the FSB and other international financial regulatory bodies could be one of the recommendations coming out of this mapping exercise.

A third way to go could be for the Global Privacy Assembly, which is the forum for more than 130 data protection and privacy authorities from around the world, to push its members to create an IDSB with standards-setting and enforcement powers over its members (for details, see Leblond 2021b). The advantage of such an approach is that the resulting single data area would immediately apply to a broad array of sectors, not just financial services.

In conclusion, trade and digital agreements have not helped facilitate international data flows. The main reason is that when these agreements contain provisions covering data, the language leaves too much room for discretion and interpretation, which allows governments to impose regulatory restrictions on cross-border data flows with impunity. The creation of an IDSB to manage an international single data area could remedy the current situation if it could issue clear, detailed standards and had the means to enforce them effectively by excluding members who do not abide by the rules from participating in the single data area. The challenge the world's policy makers face is how to set up such an international body. This essay has offered some approaches that could be feasibly pursued in the foreseeable future.

Notes

1 For an excellent discussion of digital protectionism, see Aaronson (2019).

2 The moratorium has been renewed every two years during the WTO's ministerial conference (MC). According to the WTO's Director-General, Ngozi Okonjo-Iweala, it is unlikely that it will be renewed at the next MC in 2026 (Bounds 2024).

3 *Comprehensive and Progressive Agreement for Trans-Pacific Partnership*, 8 March 2018 (entered into force 30 December 2018) [*CPTPP*], online: <www.international.gc.ca/trade-commerce/trade-agreements-accords-commerciaux/agr-acc/cptpp-ptpgp/index.aspx?>.

4 Except for data held by government entities, which is excluded from chapter 14's application.

5 *CPTPP*, *supra* note 3, art 14.8 (Personal Information Protection).

6 *United States-Mexico-Canada Agreement*, 29 January 2020 (entered into force 1 July 2020) [*USMCA*].

7 The APEC Framework is modelled on the OECD Guidelines.

8 *USMCA*, *supra* note 6, c 19, art 19.8(3).

9 *Ibid*, c 19, art 19.11 2(a).

10 The CPTPP also served as a template for the RCEP, which entered into force on January 1, 2022.

11 *Regional Comprehensive Economic Partnership*, 1 January 2022 (entered into force 15 November 2020). The stabilized text can be found here: https://docs.wto.org/dol2fe/Pages/SS/directdoc.aspx?filename=q:/INF/ECOM/87.pdf&Open=True.

A previously leaked text from December 2020 (www.bilaterals.org/IMG/pdf/wto_plurilateral_ecommerce_draft_consolidated_text.pdf) included provisions prohibiting restrictions on cross-border data transfers for business purposes and requirements to localize the storage of data domestically, using language very similar to the CPTPP's.

12 The European Union has a regulation in place that ensures the free flow of non-personal data within the European Union; however, it does not apply to non-personal data flowing outside the European Union.

13 See https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en.

14 Entry into force on September 21, 2017.

15 Entry into force on February 1, 2019.

16 EC, *Annex to the Proposal for a Council Decision on the conclusion, on behalf of the European Union, of the Protocol amending the Agreement between the European Union and Japan for an Economic Partnership regarding free flow of data*, COM(2023) 774 final, art 8.81(2), online: <https://data.consilium.europa.eu/doc/document/ST-16002-2023-ADD-1/en/pdf>.

17 See https://ised-isde.canada.ca/site/ised/en/canada-european-union-digital-partnership.

18 Since then, Australia and Mexico have become members (see www.globalcbpr.org/about/membership/).

19 On May 3, 2024, the Republic of Korea became DEPA's fourth member. Canada and China have applied to become members.

20 Restrictions on participating in the single data area could be limited to the type of data where standards are not being applied or enforced properly.

Works Cited

Aaronson, Susan Ariel. 2019. "What Are We Talking about When We Talk about Digital Protectionism?" *World Trade Review* 18 (4): 541–77. https://doi.org/10.1017/S1474745618000198.

Bounds, Andy. 2024. "Ecommerce tariffs will kick in from 2026, says WTO chief." *Financial Times*, March 27. www.ft.com/content/aea64aa4-fde2-46f3-9376-c56b8e94263b.

Ciuriak, Dan and Robert Fay. 2022. *The Digital Economy Partnership Agreement: Should Canada Join?* CIGI Policy Brief No. 171. Waterloo, ON: CIGI. www.cigionline.org/publications/digital-economy-partnership-should-canada-join/.

Cory, Nigel and Luke Dascoli. 2021. "How Barriers to Cross-Border Data Flows Are Spreading Globally, What They Cost, and How to Address Them." July. Washington, DC: Information Technology & Innovation Foundation. www2.itif.org/2021-data-localization.pdf.

European Commission. 2023. "EU and Japan conclude landmark deal on cross-border data flows at High-Level Economic Dialogue." Press release, October 27. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_5378.

Evenett, Simon J., Johannes Fritz and Tommaso Giardini. 2023. "Deterring Digital Trade Without Discrimination." *American Journal of International Law Unbound* 117: 104–9. https://doi.org/10.1017/aju.2023.15.

Ferracane, Martina Francesca and Erik van der Marel. 2021. "Do data policy restrictions inhibit trade in services?" *Review of World Economics* 157: 727–76. https://link.springer.com/article/10.1007/s10290-021-00417-2.

FSB. 2023. *Stocktake of International Data Standards Relevant to Cross-Border Payments*. September 25. www.fsb.org/wp-content/uploads/P250923.pdf.

Honey, Stephanie. 2021. "Untangling the Digital Noodle Bowl: The Case for DEPA." *TradeExperettes*, July 7. www.tradeexperettes.org/blog/articles/untangling-the-digital-noodle-bowl-the-case-for-depa.

Leblond, Patrick. 2021a. "Uploading CPTPP and USMCA Provisions to the WTO's Digital Trade Negotiations Poses Challenges for National Data Regulation: Example from Canada." In *Big Data and Global Trade Law*, edited by Mira Burri. Cambridge, UK: Cambridge University Press, 301–15.

———. 2021b. "Governing cross-border data flows beyond trade agreements to support digital trade: Inspiration from international financial standards-setting bodies." In *Addressing Impediments to Digital Trade*, edited by Ingo Borchert and L. Alan Winters, 169–94. London, UK: CEPR Press. https://cepr.org/system/files/publication-files/60026-addressing_impediments_to_digital_trade.pdf.

Leblond, Patrick and Susan Ariel Aaronson. 2019. *A Plurilateral 'Single Data Area' Is the Solution to Canada's Data Trilemma*. CIGI Paper No. 226. Waterloo, ON: CIGI. www.cigionline.org/publications/plurilateral-single-data-area-solution-canadas-data-trilemma/.

OECD. 2023. *Key Issues in Digital Trade Review: OECD Global Forum on Trade 2023 "Making Digital Trade Work for All."* October. Paris, France: OECD. www.oecd.org/en/publications/key-issues-in-digital-trade-review_b2a9c4b1-en.html.

Rentzhog, Magnus. 2015. *No Transfer, No Production – a Report on Cross-border Data Transfers, Global Value Chains, and the Production of Goods*. Kommerskollegium 2015:4. Stockholm, Sweden: National Board of Trade Sweden. www.kommerskollegium.se/globalassets/publikationer/rapporter/2016-och-aldre/no-transfer-no-production-a-report-on-crossborder-data-2015.pdf.

Taheri, Rachelle, Olivia Adams and Pauline Stern. 2021. "DEPA: The World's First Digital-Only Trade Agreement." Asia Pacific Foundation of Canada, October 7. www.asiapacific.ca/publication/depa-worlds-first-digital-only-trade-agreement.

World Economic Forum. 2020. "Data Free Flow with Trust (DFFT): Paths towards Free and Trusted Data Flows." White Paper. May. www3.weforum.org/docs/WEF_Paths_Towards_Free_and_Trusted_Data%20_Flows_2020.pdf.

About the Author

Patrick Leblond is a CIGI senior fellow. He is an expert on economic governance and policy, with a focus on North America, Europe and, increasingly, China. He has published extensively on financial and monetary integration; banking regulation; international trade, including digital trade; and business-government relations.

Prior to his current position of associate professor and holder of the CN-Paul M. Tellier Chair on Business and Public Policy at the University of Ottawa, Patrick was an assistant professor of international business at HEC Montréal and the director of the Réseau économie internationale at the Centre d'études et de recherches internationales de l'Université de Montréal. Patrick also holds the designation of chartered professional accountant and previously worked as a senior accountant and auditor at Ernst & Young in Montreal. He went on to work as a senior consultant, first in economic and financial consulting with Arthur Andersen & Co., and then later in business strategy consulting with SECOR Consulting.

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

About the Series

In 2018, the essay series Data Governance in the Digital Age anticipated some of the data governance issues that have emerged, such as surveillance capitalism and the economics of data, but did not cover data valuation in depth. Data is increasingly central to economic activity and how we make sense of the world, but it is still not valued in either national or corporate balance sheets. There is no accepted methodology to measure data's value – value that depends on its usefulness in a particular context, which is framed by individual or societal perspectives, governance rules and regulations, and input from different stakeholders. Four themes are explored in this essay series: the current state of global data governance; different perspectives on notions of value; governance frameworks to unleash the value of data; and mechanisms for governance cooperation.

Centre for International
Governance Innovation

cigionline.org