

DATA AS A VALUABLE ASSET: POWER, MARKET STRUCTURE AND SOVEREIGNTY

Sarah Nicole, Sciences Po

Stephanie Diepeveen, ODI

Global, King's College London,
UK

Sumedha Deshmukh,

University of Cambridge, Ada
Lovelace Institute

WORKSTREAM

2026

T20
U.S.

An informal engagement
of global think tanks

Data As A Valuable Asset: Power, Market Structure and Sovereignty

Sarah Nicole (Sciences Po), Stephanie Diepeveen (ODI Global, King's College London),
Sumedha Deshmukh (University of Cambridge, Ada Lovelace Institute)

Abstract

This policy brief argues that while compute captures much current policy attention, data is AI's most critical input and a critical area where the G20's parallel tracks are well positioned to progress effective global digital governance. Without data, large language models would not exist. How data is collected, used, owned, and governed is fundamentally a question of power and sovereignty. Today's leading AI companies have relied on mass data extraction to train their systems. This approach favours incumbent hyperscalers and those first to market, concentrating innovation among a handful of tech giants controlling vast data repositories and the infrastructure to process them. Further, having exhausted available web data, leading AI companies now increasingly turn to synthetic data, a closed loop that risks degrading model quality over time and further entrenches incumbents who control the infrastructure to generate and process it.

Data underpins some of the fundamental political and economic challenges facing G20 members – and non-G20 members – within an increasingly digital and AI-focused economy. Despite being critical to the digital economy, the stakes that surround data access, capture, and use have not been reflected in global governance efforts. While recent T20s, notably Brazil and South Africa, have advanced data justice discussions, there remains a tendency to frame data in financial terms (advertising, logistics, or model training), with public discourse narrowing to personal data, copyright, and privacy. This framing obscures deeper questions about value.

Data is not a neutral or uniform commodity. It is produced, shaped, and interpreted through specific technical, social, and institutional processes which inform decisions, structure power, and reflect embedded values. Its value is not intrinsic but contextual and relational: the same dataset can support multiple, even conflicting, uses depending on who accesses it and for what purpose. Governance models must resist reducing data to a single economic logic and honour the diversity of meanings embedded in its production.

DISCLAIMER: This Policy Note, authored independently and subjected to peer review, presents the views and opinions of its writers. These do not necessarily reflect the official policy or position of the T20 U.S. Summit Secretariat.

Key recommendations for the G20 (to be elaborated alongside the full brief) include: (1) **Advance definitional clarity on data** to design more comprehensive policies aligned with the realities of the AI economy. (2) **Develop market-shaping policies** by addressing data as a mechanism of market capture, considering a range of potential business models, and countering structural advantages held by incumbent players.

Challenge & Analysis

1. Data: more than a commodity

[Data is often framed in economic terms](#)¹ – from advertisement business models to financial transactions – and public discourse tends to narrow its framing further by focusing on [personal data, copyright, and privacy](#).² This individual and economic framing, while important, obscures deeper questions about how data, and the knowledge from processing and aggregating it, holds value.

Data is not a neutral or uniform commodity. It is produced, shaped, and interpreted through specific technical, social, and institutional processes. It is active: it informs decisions, structures power, and reflects the values embedded in its creation and use. Recognising the contextual and relational nature of data reveals forms of value that extend well beyond extraction or optimisation.

Crucially, [data's value is not intrinsic – it depends on how the data is contextualised, interpreted, and applied](#).³ A single dataset can support multiple, even conflicting, uses depending on who accesses it and for what purpose. Health data, for instance, might help a hospital network refine diagnostic algorithms, allow an insurer to price risk, or feed a government's public health surveillance systems. Recognising this multiplicity is key to building governance models that honour the diversity of relationships and meanings embedded in its production and use.

2. Data at the center of the digital economy, and as a site of market concentration

[Data is both an essential productive resource and an increasingly scarce strategic asset](#).⁴ Unlike traditional inputs such as capital or labour, data is largely non-rival: it can be used repeatedly, by multiple actors, without being depleted. This property has made it [foundational to business models](#)⁵ built on network effects, where the accumulation of data by early movers improves their products, attracts more users, generates still more data, and further entrenches their advantage.

¹ Qin et al., “Commercial Value of Data Resource”.

² G’sell, “An Overview of the European Union Framework”.

³ Coyle et al., “The Value of Data”.

⁴ Oderinwale and Kazlauskas, “The Economics of AI Training Data”.

⁵ Rikap, “Varieties of corporate innovation systems”.

The result is a self-reinforcing cycle of market concentration – and potential anti-competitive practices – that predates AI but that AI has made significantly more acute, since Large Language Models (LLMs) require data at a scale only a small number of firms and states can realistically assemble and process.

This dynamic positions data not simply as a commercial input but as a [strategic precondition for participation in the digital economy](#),⁶ comparable to energy or telecommunications networks in earlier industrial eras. Governments, companies, and individuals who lack access to – or control over – large quantities of data and especially high-quality data are correspondingly constrained in their capacity to develop competitive digital industries, shape the AI systems that increasingly mediate economic and civic life, and capture the value that their own data generates.

How is data defining the future of AI development?

Data scarcity derives from the limited availability of new, diverse, and high-quality information capable of improving frontier AI models. [Most of the web's publicly accessible data has already been absorbed into existing models](#),⁷ so each additional round of scraping adds less than the last. Progress nonetheless depends on exactly this kind of data.

While much of the [policy attention – and discussion of sovereignty – is on compute scarcity](#),⁸ rather than data, the advancement of recent Chinese open-weight models has demonstrated that less compute-intensive training runs can still produce high-benchmark frontier models. Yet, [no comparable breakthrough has reduced the need for high-quality data](#).⁹ In this context, AI labs are turning instead to [proprietary](#)¹⁰ and [domain-specific](#)¹¹ datasets that sit beyond the reach of the wider market. One partial substitute is [synthetic data](#)¹² – artificially generated data that mimics the statistical properties and patterns of real-world data – which can help offset the shortage of human-generated and environmental sources. Leaning on it too heavily, however, carries risk. Errors already present in a model can be amplified rather than corrected, diversity can narrow, and ["model collapse"](#)¹³ can set in – a process in which each new generation of models trains increasingly on the output of prior models rather than on original human data, so that the range of information they capture shrinks over successive generations and their reliability across varied contexts degrades.

Embodied and robotics data offers a longer-term alternative, but such data cannot simply be scraped: it has to be generated through physical interaction and sits largely inside proprietary

⁶ Belli et al., “Data sovereignty and data transfers”.

⁷ The Economist, “AI firms will soon exhaust”.

⁸ The Economist, “Sovereign AI”.

⁹ Muenninghoff et al., “Scaling Data-Constrained Language Models”.

¹⁰ Open AI, “OpenAI Data Partnerships”.

¹¹ Google DeepMind and EMBL-EBI, “AlphaFold”.

¹² Schaffelder and Gatt, “Synthetic eggs”.

¹³ Shumailov et al., “AI models collapse”.

industrial and commercial systems, [making it far slower and costlier to accumulate at scale](#).¹⁴ Data also must be made ‘legible’ to machine learning systems through processes of structuring and labelling data that have relied on [outsourced, hidden and lower-income human labour, often in poorer countries](#).¹⁵ Meanwhile, as generative AI spreads, [more of the web becomes AI-generated](#),¹⁶ feeding a recursive loop in which future models increasingly train on past models’ outputs. Filtering and attribution may help, but the trend holds: trustworthy, human-generated, real-world data keeps growing scarcer.

3. The governance gap

Governance efforts remain focused on the cycles of technology rather than on the underlying inputs that drive those cycles. China’s newly launched World Data Organization (WDO) is a notable exception, explicitly targeting data as an input and unlocking data’s value for the digital economy. Unlike time-bound mechanisms such as the UN CSTD Working Group on Data Governance, the WDO’s standing institutional form signals a [bid for long-term influence over data governance](#) rather than a temporary mandate.¹⁷

Still, governments frequently do not invest in data as a public good, or lack the institutional means to do so in ways that advance public value. Initiatives such as the UK’s National Data Library point to what a more deliberate public investment model could look like, but such examples remain the exception rather than the norm. Such initiatives often lack the political support and funding to be sustained over the long term. In addition, there has been a growing push to invest in public data that contributes to public and private uses. By comparison, there has been little consideration of how private data might be used to contribute to public aims. Some frameworks like the EU Data Act and Indian Non-Personal Data Framework have taken initial steps to open pathways for public interest use of private data, but there are more opportunities to consider how this might be done at scale, and in an enforceable manner.

Second, institutional fragmentation compounds the problem. Governments from Brazil, Indonesia, Germany, Japan, France, South Africa, and others [have confirmed that addressing such fragmentation](#),¹⁸ in order to build more integrated, open, and collaborative approaches to digital infrastructure development, has emerged as a consistent priority. Silos do not only constrain civil servants – they fragment the life cycle of data itself. When data cannot flow across administrations, its value is diminished: services go undelivered, inefficiencies compound, and citizens are often required to repeatedly provide information that governments already hold

¹⁴ Davison, “The data gap”.

¹⁵ Anwar and Graham, “Digital labour at economic margins”.

¹⁶ Cloudflare, “Traffic - Bot vs. Human”.

¹⁷ Nicole, “The World Data Organization”.

¹⁸ Nicole et al., “Final Report – Building Digital Infrastructure”.

but cannot share. At the same time, this data can be misused – so global governance bodies must consider how to balance effective service delivery with effective purpose limitation.

Third, where governments do attempt to value data for policy and regulatory purposes, they often [lack robust methodologies for doing so](#) ¹⁹ – a gap some are now attempting to close by using large language models themselves to estimate data's value, an approach that raises its own questions about circularity and reliability given the models' dependence on the very data being assessed. Some have proposed frameworks based on the amount of data, potential uses of the data and/or the openness of the data – but these have yet to be fully developed and adopted.

Fourth, and taken together, these gaps mean that the institutions best placed to govern data – nationally and internationally – are structurally unequipped to do so at the pace or scale the AI economy now demands. At the international level, there is a clear absence of permanent international institutions dedicated to data governance, combined with a broader tendency for these debates to run in silos, particularly with AI governance and data governance discussions continuing to run on separate tracks despite their deep interdependence. This is precisely the gap the G20's parallel tracks are positioned to help close.

Recommendations

The gaps identified above – definitional ambiguity, unchecked market concentration, misdirected policy attention, and a governance architecture too fragmented to keep pace with AI development – are not separate problems requiring separate fixes. They are symptoms of the same underlying shortfall: global governance has not yet treated data as the strategic asset it has become. The following recommendations respond directly to this shortfall.

1. **Advance definitional clarity on data** to design more comprehensive policies aligned with the realities of the AI economy.

Without a shared vocabulary, policymakers risk producing policies that are technically precise but practically incoherent, applying privacy-style protections to industrial data, or treating public-interest datasets as if they were purely commercial assets. The G20 is well placed to lead this work precisely because it already runs parallel political and finance tracks such as the Digital Economy Working Group within the Sherpa Track as well as the Finance Track. These tracks need to align technical, economic, and political perspectives on the question, in ways that support more aligned and comprehensive definitions of data in other international and national fora.

This could include agreement on a definition of data as an economic good, as well as principles for value creation and capture. This could form the basis of potential re-distributive mechanisms

¹⁹ Coyle and Manley, “What is the value of data?”.

for data-related revenue. Definitional clarity could also advance tracking data value creation and capture for AI, which would [materialize past G20 ambitions to make data governance the foundation for equitable AI](#).²⁰

Such thinking may necessitate a global body or board that is responsible for the global governance of data and its value. The G20 and T20 could provide concrete research on the options for this type of mechanism, and the G20 could be the ideal forum for advocating for – or elevating – this body.

This definitional clarity will also allow the G20 to play a leadership role in uniting data governance and AI governance rather than leaving them siloed.

2. **Develop market-shaping policies** by addressing data as a mechanism of market capture, considering a range of potential business models, and countering structural advantages held by incumbent players.

Data's non-rival nature does not make its accumulation harmless. As the analysis above sets out, the same network effects that make data valuable also make it a powerful mechanism of market capture: firms that accumulate data early gain [compounding advantages](#) that smaller competitors and new entrants cannot easily overcome.²¹ Further, it is possible to [limit or exclude](#) some from access to data, for example, if it is produced within proprietary systems.²²

Addressing this requires the G20 to treat data access and control as a live competition-policy question, rather than only a privacy one – as seen with the EU's Digital Markets Act (DMA), which complements its existing GDPR. The G20 (with the assistance of the T20) could map market failures across the data and AI value chain. This would be an important first step in considering the wide range of business and governance models that might answer current and future market dynamics.

For instance, where a few firms hold a dataset that others cannot copy, there could be mechanisms for licensing or access regimes to prevent exclusive control. Where users and competitors face high switching costs, governments or international bodies could mandate interoperability and data portability to lower the cost of moving between services. Where small actors lack bargaining power, there could be support for data trusts and cooperatives that pool power and let smaller players scale. Where the market under-provides public-interest data, there might be investments in public data infrastructure - as seen with the UK's National Data Library.

²⁰ G20 Task Force on Artificial Intelligence, “Chairs statement”.

²¹ Mulligan and Godsiff, “The Value Dynamics of Data Assets”.

²² Coyle et al., “The value of data”.

Coordination among G20 members is itself a market-shaping tool - particularly in terms of increasing collective bargaining power. Acting alone, a single country has weak leverage over the terms on which its data is used and valued. Acting together, G20 members can set those terms with a greater likelihood of success, given the size of the combined market. As discussed, much private data value is built on public foundations — public research, public infrastructure, and data that citizens generate and, in effect, pay for. Yet the value tends to flow one way, captured by private firms with no duty to contribute back. Coordinated G20 action could set the principle that public inputs earn a public return. Fiscal policy is an important vector to consider, as tax revenues from data-related businesses are unevenly distributed globally – even though much of the data value chain sits outside of the countries with the largest AI and data monetary gains.

It also requires attention to market structure. Anti-monopoly and competition policy are increasingly important mechanisms for checking corporate power that has long-term economic and political effects – and the G20 and T20 might consider ways in which the data-based business model challenges conventional thinking around antitrust and market assumptions. G20 working groups concerned with trade and with the digital economy can take on a leading role in collating and rethinking market assumptions in the context of data-driven economic activity.

Taken together, these two recommendations and the implementation roadmap in the annex offer the G20 a path forward: define data clearly enough to govern it well and correct the market dynamics that concentration has already produced. Yet, it is important to note that without a standing international institution – which [the World Data Organization seems to be trying to fill](#)²³ – data definitional work and market-shaping policy alike will remain ad hoc, dependent on whichever forum happens to take up the issue in a given year, and vulnerable to being abandoned once political attention moves elsewhere.

²³ Nicole, “The World Data Organization”.

Bibliography

Anwar, Mohammad Amir, and Mark Graham. "Digital Labour at Economic Margins: African Workers and the Global Information Economy." *Review of African Political Economy* (2020): 47 (163): 95–105. doi:10.1080/03056244.2020.1728243.

<https://www.tandfonline.com/doi/full/10.1080/03056244.2020.1728243>

Belli, Luca Britto Gaspar, et al. "Data Sovereignty and Data Transfers as Fundamental Elements of Digital Transformation: Lessons from the BRICS Countries". *Computer Law & Security Review*, volume 54, 2024[10.1016/j.clsr.2024.106017], Available at SSRN:

<https://ssrn.com/abstract=4903196> or <http://dx.doi.org/10.2139/ssrn.4903196>

Cloudflare. "Bot vs. Human Worldwide." Cloudflare Radar. Accessed September 3, 2026.

<https://radar.cloudflare.com/traffic#bot-vs-human>.

Coyle, Diane, Stephanie Diepeveen, et al. "The Value of Data: Summary Report." Bennett Institute for Public Policy, Open Data Institute, and Nuffield Foundation, February 2020.

https://www.bennettschool.cam.ac.uk/wp-content/uploads/2020/12/Value_of_data_summary_report_26_Feb.pdf.

Coyle, Diane, and Manley, Annabel. What is the value of data? A review of empirical methods. *Journal of Economic Surveys* (2024): 38, 1317–1337. <https://doi.org/10.1111/joes.12585>

Davison, Antonia. "The Data Gap That's Holding Back Robotics." *IBM Think*, November 27, 2025. <https://www.ibm.com/think/news/the-data-gap-holding-back-robotics>.

G20, Task Force on Artificial Intelligence, Data Governance and Innovation for Sustainable Development. "Chair's Statement." Cape Town, South Africa, September 30, 2025.

<https://g7g20-documents.org/database/document/2025-g20-south-africa-sherpa-track-digital-economy-ministers-ministers-language-chairs-statement-task-force-on-artificial-intelligence-data-governance-and-innovation-for-sustainable-development>.

Google Deepmind and EMBL-EBI. "AlphaFold Protein Structure Database." EMBL-EBI.

Accessed September 3, 2026. <https://alphafold.ebi.ac.uk/>.

G'sell, Florence. "An Overview of the European Union Framework Governing Generative AI Models and Systems." SSRN, May 20, 2024. <https://ssrn.com/abstract=4762804>.

Muennighoff, Niklas, Alexander M. Rush, et al. "Scaling Data-Constrained Language Models." *Journal of Machine Learning Research* 26 (2025): 1–66.

<https://www.jmlr.org/papers/volume26/24-1000/24-1000.pdf>.

Nicole, Sarah. "The World Data Organization: China's Global Bid for AI's Scarcest Resource." Oxford China Policy Lab. Accessed September 3, 2026.

<https://oxfordchinapolicylab.org/research/the-world-data-organization-china-s-global-bid-for-ai-s-scarcest-resource>.

Nicole, Sarah, Vidisha Mishra, et al. "Building Digital Infrastructure Ready for the AI Era: Insights from 13 Middle Power Governments." Global Solutions Initiative, June 1, 2026.

<https://www.global-solutions-initiative.org/publication/building-digital-infrastructure-ready-for-the-ai-era/>.

Oderinwale, Hamidah, and Kazlauskas, Anna. "The Economics of AI Training Data: A Research Agenda." arXiv, October 28, 2025 (revised August 2, 2026). <https://arxiv.org/abs/2510.24990>.

OpenAI. "OpenAI Data Partnerships." *OpenAI*, November 9, 2023.

<https://openai.com/index/data-partnerships/>.

Qin, Haolang, YuLin Liu, and YiTing Hu. "Commercial Value of Data Resource: An Overview of Current State of Research and Future Prospects." *Sage Open*, October 9, 2025.

<https://doi.org/10.1177/21582440251379511>.

Rikap, Cecilia. "Varieties of Corporate Innovation Systems and Their Interplay with Global and National Systems: Amazon, Facebook, Google and Microsoft's Strategies to Produce and Appropriate Artificial Intelligence." *Review of International Political Economy* (2024): 31 (6): 1735–63. doi:10.1080/09692290.2024.2365757.

<https://www.tandfonline.com/doi/full/10.1080/09692290.2024.2365757>

Shumailov, Ilia, Zakhar Shumaylov, et al. "AI Models Collapse When Trained on Recursively Generated Data." *Nature* 631 (2024): 755–759. <https://doi.org/10.1038/s41586-024-07566-y>.

Schaffelder, Max, and Albert Gatt. "Synthetic Eggs in Many Baskets: The Impact of Synthetic Data Diversity on LLM Fine-Tuning." arXiv, November 3, 2025.

<https://arxiv.org/abs/2511.01490>.

The Economist. "AI Firms Will Soon Exhaust Most of the Internet's Data." *The Economist*, Schools Brief, July 23, 2024.

<https://www.economist.com/schools-brief/2024/07/23/ai-firms-will-soon-exhaust-most-of-the-internets-data>.

The Economist. "Sovereign AI Independent of America and China Is a Pipe Dream." *The Economist*, July 16, 2026.

<https://www.economist.com/international/2026/07/16/sovereign-ai-independent-of-america-and-china-is-a-pipe-dream>.

Annex

Implementation Roadmap

Diagnosis: Data underpins some of the fundamental political and economic challenges facing G20 members within an increasingly digital and AI-focused economy, but global governance efforts fail to accurately reflect the stakes that surround data access, capture and use. The institutions best placed to govern data are structurally unequipped to do so at the pace or scale the AI economy now demands. The G20 has a critical role to play in shifting global governance models for data to more accurately capture how data is valuable and engage with relevant market-shaping mechanisms.

Stakeholders (Who?):

- Lead: The Leaders Summit must play a leading role in elevating data within G20 discussions, and providing the political impetus for sustained and purposeful research and debate into data as an asset in the G20 working groups and formal engagement groups (e.g. T20, B20).
- Execution: G20 working groups on Trade, Innovation and the Digital Economy to provide sustained attention to defining and governing data, in coordination with official engagement groups to explore different issues and perspectives (e.g. B20, T20).

Action Plan, Implementation Steps and Timeframe (How?):

1. Short term (6 months - 1 year):

- a. Obtain the political will and leadership to centre data in G20 negotiations and activities through the shared recognition of the importance of data and challenges of global data governance at the 2026 Miami Summit
- b. Build on the shared political will to set out clear roles and responsibilities within the UK G20 working groups to debate and present more accurate and complete definitions of data and its value within a digital and AI-driven global economy. And, to report on the outcomes of this to the 2027 G20 Leaders Summit.
- c. Liaise with other global bodies interested in data governance to elevate it as an issue in its own right and agree to a more comprehensive and accurate definition of its value, endorsed and supported by the G20 (e.g. IGF, UNGA, Global Digital Compact)

2. Medium term (2-4 years):

- a. Develop an established programme of work on data governance within the G20 working groups, which reports annually to the G20 Leaders Summit
- b. Continue to work closely with partner forums to agree and utilise a shared - accurate and comprehensive - approach to data and its value in the global economy

- c. Through the G20 working groups, identify and endorse an appropriate body for global data governance, focused on market shaping mechanisms and an equitable and stable global economy

Spaces for Articulation (Where?):

- In addition to the G20 Finance Ministers meeting and Leaders Summit, and Digital Economy Working Groups within the Sherpa Track, relevant forums for action and alignment with a variety of stakeholders include: IGF, WTO, World Bank/IMF Annual Meetings, UNGA, and others.

Monitoring and Accountability: To be reviewed annually as part of G20 Working Group meetings and reported on for the annual G20 Leaders Summit, potentially in collaboration with other global stakeholders.