TOOLKIT:

Data governance and the G20's 2025 themes: Solidarity, Equality and

September 2025

A contribution to the Task Force on Artificial Intelligence, Data Governance and Innovation for Sustainable Development

This knowledge resource contributes to the G20s debates on data governance issues and highlights convergent understandings between members of relevant engagement groups, Sherpa and Finance tracks.

Knowledge partner: UNESCO

CONTENTS

1. EXECUTIVE SUMMARY 3
2. TOOLKIT AIM 3
3. DATA GOVERNANCE AND KEY G20 CONCERNS 5
4. UNPACKING TERMS AND CONCEPTS 11
5. DATA GOVERNANCE AS A CENTRAL PILLAR OF AI AND DIGITAL TRANSFORMATION 16
6. DATA GOVERNANCE: THE GENERAL AND THE PARTICULAR 21
7. WHY: ALIGNMENT WITH A VISION AND PURPOSE OF DATA GOVERNANCE 22
8. WHAT GOVERNANCE COVERS ACROSS THE DATA LIFECYCLE 27
9. HOW: KEY STEPS IN DOING DATA GOVERNANCE 32
10. WHO: PEOPLE AND ROLES 36
11. WHERE AND WHEN: ROADMAP TO FACILITATE IMPROVED DATA GOVERNANCE 41
12. ACRONYMS AND ABBREVIATIONS 44
13. APPENDIX A: BACKGROUND 45
14. APPENDIX B: DATA GOVERNANCE AND DIGITAL PUBLIC INFRASTRUCTURE 46

1. EXECUTIVE SUMMARY

This toolkit is a resource for G20 and other countries to govern data in the context of artificial intelligence technologies (AI) as applied to the core focus areas of the 2025 G20 presidency. It complements and builds on the global data governance toolkit produced by the Broadband Commission for Sustainable Development Working Group on Data Governance, as well as insights from a data governance dialogue held at the second meeting of the G20's Task Force on AI, Data Governance and Innovation for sustainable Development (AITF). It further incorporates insights from the 16 responses by participants in the 2025 G20 to a survey of their data governance practices.[1] Countries are at different stages of developing data governance frameworks. While some are consolidating fragmented laws and policies into a single national framework, others are strengthening sectoral measures and interoperability. In several cases, governments are preparing comprehensive data governance policies that integrate privacy, interoperability, transparency, and innovation into a single framework. This demonstrates how multiple instruments—such as access to information, personal data protection, digital government laws, and data-sharing decrees—can evolve into a more coordinated and balanced approach.

UNESCO as knowledge partner to the South African G20 Presidency produced this customized toolkit as part of its wider work on data governance.

2. TOOLKIT AIM

This resource offers tools for data governance in selected domains of direct relevance to the G20's concerns, and how these intersect with AI in these areas. The primary envisaged users of these tools are policymakers and civil servants in G20 countries and beyond. The tools will also be of value to data practitioners from the private sector and civil society interested in how data governance impacts key areas of social and economic life in regard to the development and application of AI technologies. More information on the background of the Toolkit and its scoping is provided in Appendix A.

A targeted survey of G20 stakeholders, completed in July 2025, underscores a diverse spectrum of national awareness and readiness in data governance. Of 16 respondents, over half reported medium awareness and sensitivity among civil service leadership (nine out of sixteen selected "medium"), four identified high or very high sensitivity, and two assessed their leadership as having low awareness. This points to measurable progress in embedding data governance

[1] Responses came from African Union, Argentina, Brazil (two responses), European Union, Germany, Italy, Korea, Mexico, Netherlands, Nigeria (two responses), Norway, Russia, Saudi Arabia, South Africa, Spain.

issues at the leadership level but also highlights remaining gaps and the need for furrher capacity building and cross-government alignment.

Key governance challenges prioritized by respondents were:

1. **Dealing with privacy abuses and Cybersecurity** (scored as a top 3 challenge by nearly all respondents, with multiple marking it as "1").
2. **Unlocking data for public value** and **Cross-border data flows**, both consistently ranked as being of high national relevance.
3. **Storage constraints, and data science skills** were also noted as pressing, though relatively fewer countries scored these as their principal concern.

On **institutional arrangements**, most participating countries report a distributed approach to governance; responsibilities span ministries and agencies dedicated to privacy, competition, data markets, and consumer rights. Several respondents cited ongoing efforts to consolidate frameworks and better coordinate across traditional policy silos.

Regarding **AI and new challenges**, while the survey shows national strategies often include an AI component and recognize the opportunities of advanced analytics, most responses focus on upgrading data quality, interoperability, and capacity-building, rather than highlighting explicit governance structures for agentic AI.

Many countries employ hybrid approaches to institutional design, combining distributed instruments with central coordination. Independent data protection authorities play a key role in safeguarding privacy and providing oversight on emerging areas such as AI. Central committees, interoperability programmes, and open data policies complement this work by promoting integration and accountability. Civil service awareness is often rated as moderate, reflecting ongoing capacity-building efforts.

Good practices identified include:

- Cross-agency digital infrastructure
- National data and cloud policies,
- Standardized data-sharing guidelines and privacy authorities
- Interoperability platforms that connect federal systems, reducing administrative burdens and improving efficiency.
- National data catalogues centralizing datasets for transparency, reuse, and accountability.
- Portability frameworks in the financial sector, fostering innovation and consumer choice.

However, respondents also flagged major fragmentation, sectoral silos and a lack of clear, accountability mechanisms for inter-agency / cross-sector challenges, such as by AI agents.

In response, the toolkit gives attention to:

- Navigating cross-border data and complex jurisdictional landscapes,
- Advancing secure and ethical AI (including with a view toward developments like Agentic AI),
- Fostering whole-of-government coordination and establishing clearer lines of accountability.

As shown in the survey, G20 countries are at varying stages in recognizing and addressing advanced data governance challenges, with particular gaps in coordination and capacity to address new forms of autonomous AI. The toolkit's recommendations are shaped to reflect these realities—focusing on practical frameworks, interoperability, ethical oversight, and targeted capacity building—so all G20 members can better navigate today's evolving data governance environment.

3. DATA GOVERNANCE AND KEY G20 CONCERNS

This toolkit begins with the recognition that data governance issues, both cross-cutting and sector-specific, are relevant to multiple G20 working and engagement groups. Data governance is a key enabler across priorities such as the digital economy, inequalities, climate action, disaster response, trade and finance. Appendix B explores data governance in the context of Digital Public Infrastructure (DPI), highlighting the vital role of the transversal G20 Task Force on Artificial Intelligence, Data Governance and Innovation for Sustainable Development (AITF).

In line with the priorities of the 2025 South African G20 presidency, the table below maps selected agenda topics to artificial intelligence and data governance, with further details provided later in this document.

G20 Topic	**Data challenge**	**AI aspect**	**Data governance dimension**
Promoting Solidarity, Equality, and Sustainability;	Assessing gaps and fitness for purpose in data sets	Risk of limited or biased data that lead to AI	**Collection & Access:** Foster the collection of inclusive and representative data that reflects the realities of diverse communities,

addressing Inequalities	relevant to these objectives. Shortfalls in quality data to map and address these issues, Transforming data into effective policies and actions that contribute to sustainable development and inclusion.	reinforcing inequalities. Over-concentration of AI data in a few geographies, and overemphasis on the private competitive aspects in the data lifecycle at the expense of collective collaboration and public benefit.	languages, and knowledge systems. Ensure that data sourcing strategies address gaps and biases, supporting global inclusivity in AI development. **Use & Processing:** Mandate bias audits and regular representation checks in AI training data and algorithms. Promote practices that incorporate the principle of equal opportunity throughout all stages of the data lifecycle, mitigating risks of discrimination and exclusion in AI outcomes. **Policy & Oversight:** Require inclusive stakeholder engagement—nationally and internationally—when developing data governance policies. Align frameworks with human rights principles by strengthening laws and enforcement mechanisms on data privacy, access, and protection.
Equitable, inclusive, and sustainable artificial intelligence (AI).	Under-capitalised entities unable to compete with transnational corporations. Inability to leverage alternative and more inclusive data sets.	Value of stimulating competition in AI markets through Digital Public Infrastructure (DPI) and targeted support measures. Using AI tools to assess industry concentration and barriers to entry	**Collection & Access:** Promote fair and open access to high-quality data, ensuring that opportunities for data availability extend to a wider range of actors, including those traditionally underrepresented. Establish mechanisms for appropriate data compensation where relevant. **Use & Processing:** Implement stringent safeguards for large and influential data holders and processors to prevent data-related

	Lack of data on environmental impact.		injustices, such as unfair market dominance or misuse of data resources. **Retention & Impact:** Incorporate assessment and reporting of environmental impacts throughout the data lifecycle, from collection to storage and disposal, ensuring sustainable and responsible data governance practices.
Digital Public Infrastructure and transformation	Harmonising standards for data interoperability and public access	Creating common data pools, and tiered access regimes to other data holdings, for purposes of training AI applications	**Standardization/Interoperability:** A dopt common data standards, tiered access regimes (how). **Access:** Maintain transparent and equitable rules for data contributors/users (who/what).
Connectivity for inclusive digital development	Digital divides both exclude large swathes of society from using digital technologies and services, and render these advances less data-rich and representative for those who are connected.	AI technologies are limited by the digital divide, and are unavailable to sectors of society that have the most to benefit. AI can map divides and impact of mitigation measures.	**Collection/Monitoring:** Regularly track and report on digital/devices divides (how/why). **Access:** Address participation/inclusion gaps in national data strategies (who/what).
Digital innovation	Lack of data and data	Sensitising these actors that their	**Access/Sharing:** Facilitate responsible unlock of large

ecosystems: unleashing the potential of MSMEs	science capacity amongst smaller and new economic actors	unique competitive edge may be less in access to AI applications, but rather the limited availability of quality data sets to which these can be applied.	public/private sector datasets for MSME use (how/what). **Consent/Compensation:** Ensure IP rights and transparency for data providers/creators (who/how).
Disaster Risk Reduction & Climate Resilience	Access to and use of data to identify vulnerabilities, to engage early warnings, and to help target responses and evaluate impact	Opportunity of AI aiding data analytics. Risk of automation errors.	**Coordination/Sharing:** Enable rapid data sharing among key stakeholders (who/how). **Quality Control:** Standardize mechanisms for data accuracy, update, and authorized use (how/what).
Debt Sustainability	Obstacles to access and use data to assess debt sustainability	AI to help with predictive analytics. AI spending based on debt may fuel an investment bubble.	**Collection/Update:** Ensure timely, comprehensive debt data collection (who/what). **Access/Use:** Develop predictive tools with transparent methodologies, based on updated data (how).
Climate Action	Hurdles in aggregating and harmonising swathes of data on carbon and methane	AI to retro-engineer data interoperability.	**Collection/Harmonization:** Promote adoption of shared standards for climate/emissions data (how). **Accountability:** Require environmental cost reporting by data processors/centres (why/who)

	emissions, captures, offsets, etc.		
Agriculture and food security	Unrealised opportunities in harnessing data-for sustainable agriculture	Constraints on design and deployment in AI innovation and in technology transfer.	**Interoperability/Sharing:** Advance standards for agri-data platforms (what/how). **Rights:** Protect farmer privacy and support data-sharing protocols that respect local interests (who/why).
Addressing polycrises (Climate, Energy, Food, Debt)	Obstacles in integrating data governance across sectors/issues. Lack of platform transparency as a fetter to evidence-based action around information integrity.	AI could help to bridge different data sets, and analyse interdependencies in ways that promote data-based transparency and insight	**Cross-sector** **Interoperability:** Establish multi-agency, multi-sector frameworks for secure, consistent data sharing including data exchanges (how/who). **Transparency:** Encourage open, accountable models for public-facing analysis (why).
Harnessing critical minerals for inclusive growth and sustainable development	Insufficient data on mineral reserves, extraction processes, and supply chain transparency can lead to exploitation,	AI can support supply-chain analysis, and aggregation of geological, aerial, technical, economic and trade data.	**Collection/Transparency:** Mandate data collection and open access for geological, economic and environmental datasets (who/what). **Monitoring:** Set reporting rules for social/environmental impacts (how/why).

	environmental damage, and unequal distribution of benefits.		
Heritage restitution and culture	Data sovereignty Respect for cultural diversity in data policies. Intellectual property issues.	The quality of AI development and deployment depends on diverse, quality and sustainably data sets.	**Rights/Stewardship:** Embed cultural rights and data sovereignty in national/international data frameworks (who/what/why). **IP/Honouring Origin:** Ensure data policies respect indigenous and local ownership (how/who).

Many G20 countries already have overarching, cross-cutting and/or topic-specific frameworks relevant to the data issues cited in the table above. Many also have relevant approaches to issues such as Open Data, data protection, digital ID and payment systems, e-health, etc. Accordingly, this toolkit recognises there is not a tabula rasa within the G20. Rather, it offers a supplementary perspective to what exists.

It therefore presents a sample of checklists and templates that can help G20 state actors, and others, to take stock of their existing data governance activities from a holistic perspective, while also recognising specific and distinctive issues at the domain level, including the related intersections with AI. The toolkit recognises the fallacy of "one size fits all", and also the risks of centralising control around a lever that is as powerful as data. The assumption is that there exist high-level and overall frameworks for data governance in G20 countries, and while these may need to be revised and updated, specific tools can assist in localised assessment of gaps, frictions and obstacles, and also propose ways to overcome these. The intention, therefore, is to support G20 members in making positive change and improvements especially at national levels. In this way, the resources in this toolkit are designed to integrate with both existing and new elements within individual countries' data governance practices.

G20 participants to the July 2025 online survey commented on challenges to their country's work on data governance. They listed cases of for emergency responses, climate monitoring, public debt transparency, targeting social assistance policies, financial inclusion, and critical

infrastructure. One country stated, however: "There is no overarching, integrated data governance strategy explicitly linking these domains [sector specific data governance] to national priorities in sustainability or social equity. As a result, the potential of data to inform and coordinate cross-sector action remains underdeveloped. Most efforts remain siloed, though a few initiatives signal emerging integration".

The African Union stated that its data governance efforts are directly aligned with the G20's 2025 focus on inequality, disaster resilience, climate action, debt sustainability, and critical minerals. "Improved data systems support climate resilience by strengthening early warning systems and environmental monitoring. In the critical minerals sector, data governance ensures transparency, traceability, and fair benefit-sharing. Similarly, in tackling inequality and debt, better data enables targeted social protection and public finance management." The African Union added that it was committed to working with G20 partners to embed equity and resilience into global digital and development frameworks.

One country respondent said: "[T]here is a complex impact on overall issues rather than just specific areas. The development of data governance is expected to be significant in the economic and industrial sectors where data analysis and utilization are most active."

4. UNPACKING TERMS AND CONCEPTS

About Data: For this toolkit, datafication can be taken to mean the transformation of digitalised signals into a raw material for further processing. Not all digitization processes result in "data", and nor should they. Governance can authorize, promote – and interdict – areas of datafication. Once created, data becomes part of a lifecycle and governance issues continue to apply throughout. (The concept of a data lifecycle is elaborated in section 8 below).

About Governance: Adapting the perspective of the World Summit on the Information Society, this encompasses the development and application of shared principles, norms, rules and decision-making procedures. The core sense is to draw attention to the spectrum of control factors - within which rules (such as laws and regulations) exist within a wider frame of variables that direct and shape the phenomenon being governed (data in this case).

Applied to data governance, the Broadband Commission Data Governance Toolkit: Navigating Data in the Digital Age defines data governance as follows: "The processes, people, policies, practices and technology that seek to govern the data lifecycle toward meeting the purpose of increasing trust, value and equity, while minimizing risk and harm in alignment with a set of core principles". (italics added). Drilling down further, data governance can be understood as being about the control (or lack thereof) of how data is generated, managed, used and re-used.

A comprehensive approach to data governance will therefore be founded on an overarching vision for data within society, which in turn can inform more specific data governance that highlights common elements in common, but which also accommodate appropriate implementation across different domains and institutions. Of general relevance is that different governance considerations may apply to the different categories of data as outlined in the box below:

- **Personally identifiable information (PII)** covers data such as names, addresses and identity numbers, as well as facial, iris, fingerprint and gait records, and implicates the right to personal privacy.
- **Non-personal data** refers to aggregated or anonymized datasets, or those resulting from environmental and other sensors.
- **Unstructured data** is that which lacks a predefined format or organizational framework, requires specialized tools and strategies to unlock its potential. It can be processed to can constitute either PII and non-personal data, or a mix of both.

Data governance faces the threat of unstructured and non-personal data being engineered to reveal personal data, since in practice the distinctions between these categories is not always clearcut.

Synthetic data refers to artificially generated information produced through computational methods. While it is designed to replicate the statistical patterns and relationships found in real-world datasets, it excludes any direct identifiers or personal information from individuals. The value of synthetic data lies in its ability to enable privacy-preserving analysis and model training using data that maintains the utility of the original. However, because synthetic data is fundamentally detached from its source, there are inherent challenges regarding fidelity and authenticity. These concerns become more pronounced if new synthetic datasets are generated from previous synthetic data, as each subsequent iteration increases the risk of drifting further from real-world conditions, potentially undermining data validity and reliability.

While deeply intertwined and often co-dependent, data governance and AI governance are distinct. Data governance can be seen as a broad foundational layer, that extends beyond AI governance, while AI governance can be visualised as covering a vertical set of considerations where data is one, critical, layer alongside others such as compute, algorithms and capacity.

AI governance for its part can be understood as:

- The framework of policies, processes, people, roles and tools that include but which also go beyond the data itself to encompass the character, behaviour and impact of the AI models, applications and systems·

- With regard to data within the AI lifecycle of development, deployment, monitoring and retirement, the governance of AI may address issues such as system fairness, explainability, accountability, risk, performance and ethics. Beyond the data component (to which AI systems also contribute new data), there are issues of governing algorithms, compute technology and environmental impact.

This characterisation shows that AI governance cannot exist effectively without robust data governance. If the underlying data is biased, incomplete, of poor quality or improperly secured, any AI system built on it will inherit and likely amplify those flaws. Thus, a public or private institution for example cannot just procure an AI system without also giving attention to how that institution governs data in terms of quality assurance, interoperability, privacy protection, etc.

In brief, good data governance provides the essential data foundation which AI governance needs to ensure responsible and effective AI systems. This is an underlying truth that applies transversally, with considerations as per the domain under consideration (for example, crisis response, MSME development, etc.)

In a nutshell: data governance versus digital transformation

Digital transformation is a broad process that involves the adoption of digital technologies—including data systems and AI—across organizations and societies. While data governance is concerned with the policies, standards, and practices that ensure data is collected, managed, and used responsibly, digital transformation goes further. It encompasses not only the effective use of data but also addresses the impact of digital technologies within broader economic, political, social, and environmental contexts.

Digital transformation requires organizations to navigate new opportunities and risks, carefully weighing investment costs against potential benefits at institutional, enterprise, and societal levels. Achieving successful digital transformation also depends on continuously evaluating whether existing data governance frameworks remain robust and adaptable in the face of rapid technological and contextual change.

In sum, data governance is a foundational component of digital transformation, but the latter extends to cultural, organizational, and systemic changes that shape how digital tools—including data—are integrated into everyday practice.

A number of G20 countries are engaged in new initiatives in data governance to address emerging challenges or opportunities from AI.

Participants in the G20 online survey in July 2025 cited examples of responses to data governance in the light of AI, grouped below into four main areas:

1. National Strategies and Frameworks

- New national AI strategies increasingly emphasize secure, interoperable data sharing and ethical AI use across sectors (e.g., healthcare, justice, transportation), often with explicit requirements for consent, data quality, privacy, and technical oversight.
- Ongoing reviews and simplification of existing data protection frameworks, such as the EU's GDPR, to address emerging challenges presented by AI systems.
- Development of sectoral guides (e.g., Generative AI Guide for Civil Servants) and national systems for ethical and technical standards in AI deployment, covering data quality, developer responsibility, and safety.
- Adoption and/or planning of new policies on data and cloud, including frameworks for open data and secure cloud infrastructures.

2. Legislation, Regulation, and Oversight

- Introduction of laws mandating privacy-by-design principles and Data Protection Impact Assessments specifically for AI systems.
- Regulatory responses to specific AI applications, such as oversight of generative AI tools (e.g., ChatGPT) over data and transparency concerns.
- Legislative proposals covering a wide range of AI-related topics, including governance mechanisms, education, intellectual property rights, and data handling.
- Active discussions about regulating deepfakes, mandatory AI model testing, and labeling of AI-generated outputs.

3. Institutional Capacity Building

- Establishment of dedicated agencies (e.g., national AI agencies) tasked with AI governance and coordination.
- Capacity-building is supported by national training programmes, with many countries developing dedicated courses in data science, governance, and analytics. National AI strategies are also closely linked to data governance, embedding ethical, inclusive, and sustainable practices across sectors.

4. Awareness, Education, and Ethical AI

- Launch of national initiatives and campaigns to increase awareness around the responsible and ethical adoption of AI, with a focus on aligning AI deployment with human and fundamental rights.
- Encouragement of multi-stakeholder consultations to shape AI and data policies, involving experts, industry, and civil society.

  Recognition of the value—and challenges—of leveraging unstructured data for improved AI model development and innovation.

These developments illustrate the breadth of new efforts G20 countries are taking to update and reinforce data governance frameworks in response to rapid AI advancements, focusing on trust, transparency, legal clarity, capacity building, and the promotion of fundamental rights.

The AU responded that it had launched several recent initiatives to integrate new data governance mechanisms that are tailored to AI's challenges:

1. A continental AI roadmap and White Paper (April–June 2024)—aligning AI ethics, data governance, workforce development, and digital infrastructure
2. In July 2024, the Continental AI Strategy formalised a multi-tiered governance approach emphasising data protection, cross-border data sharing, and ethical frameworks for AI systems
3. The AUDA-NEPAD "Shaping Africa's AI Future" summit (Aug 2024), establishing unified policy pathways and launching a Digital Readiness Index
4. A phased AI-in-healthcare regulatory framework (late 2024)—supporting ethical deployment of AI in diagnostics and telemedicine
5. In May 2025, a High-Level Policy Dialogue on AI reaffirmed commitment to data sovereignty, inclusion, and investment, urging Member States to adopt AI regulations aligned with continental priorities.
6. Through a May 2025 UNESCO–AU stakeholder consultation, the AU contributed African perspectives to the Broadband Commission's Data Governance toolkit, emphasising AI ethics, trust and practical implementation.

5. DATA GOVERNANCE AS A CENTRAL PILLAR OF AI AND DIGITAL TRANSFORMATION

Data governance is a field that intersects with, but is also fundamental to, the governance of digital transformation and especially AI governance. It encompasses technical, policy, human rights, ethical, regulatory, and institutional arrangements that shape the data lifecycle from creation, collection, storage, ownership, use, protection, access, sharing and deletion. With clear purpose and effective execution, data governance is essential for governance which assures that AI systems serve public interest and which balance innovation with guardrails. It is not a one-way relationship, however: AI itself has roles to play in the shaping of data governance. For more insight on the macro picture, readers seeking further information and links to a suite of tools in different domains of data governance are recommended to consult the Broadband Commission Data Governance Toolkit: Navigating Data in the Digital Age. What follows below provides context for this specific toolkit.

Data governance covers a number of inter-related fields such as:

- Cross border and jurisdictional issues
- Unlocking data for public value
- Privacy abuses
- Cyber security
- Data storage and constraints
- Stakeholder awareness and data science capacity
- Intellectual Property issues
- Data ethics

These fields have major bearing on the use of AI for the purposes of the G20. It is readily apparent that Issues of data misuse or exclusion can result in a loss of public trust in AI systems. Likewise, legitimate privacy and intellectual property concerns about data extraction, commercialization and use can hinder potentially positive uses of data and emerging technologies for the wider public benefit. As stated in the AITF issue note "Making data available for AI", "Data's true potential derives from its good public characteristics. It is a non-rivalrous and potentially non-excludable (through open data requirements) resource that can support multiple uses by different users simultaneously without being depleted. As such [it] is a critical input downstream in the wider economy but also upstream in the production of advanced data-driven technologies such as various Artificial Intelligence systems."

The AITF issue note continues: "An effective valuation framework for AI should account for how data contributes to model functionality, risk exposure, and downstream impacts. Integrating such valuation into governance mechanisms can enhance accountability, support proportional regulation, and ensure that the value generated by data is recognised and distributed fairly across

the AI value chain." It calls for a public policy, planning and regulatory perspective that balances commercial valuation of data perspectives with valuation of data in public resources allocation and potential to realise public value creation.

In this frame, it is important to keep in mind the dynamic changes in the environment of data governance. Amongst developments current in 2025 are:

- Massive increases in synthetic media and synthetic data (especially from Generative AI systems) and questions about how these impact upon data quality and the value of scaling data for AI systems
- Changes in AI to include "AI agents" which raise novel questions for data governance
- Intensified debates and legal cases over ownership, copyright, compensation and provenance
- New challenges to data security and personal privacy, including from AI technologies
- Heightened attention to data sovereignty and free flow with trust
- Changes in data portability and interoperability (the extent to which unstructured data becoming more processable with the use of generative AI).

G20 participants reported a wide range of active measures to advance data quality, interoperability, and portability:

Enhancing Citizens' Access, Control, and Data Portability

- Models such as personal data vaults are being explored to give citizens more control over their personal data.
- Data Subject Access Requests empower individuals to rectify, update, and port their personal information.
- Electronic identity systems and Public Digital Identity Cards enable secure, interoperable digital services across sectors.
- Enforcement of legal frameworks, such as the EU's GDPR Article 20, guarantees the right to data portability.

Building Interoperable Digital Ecosystems

- National Data Platforms and Government Service Buses facilitate secure, real-time data exchange between public agencies, using standardized formats and access protocols.
- National Interoperability Frameworks, aligned data standards, and guidance for API development support seamless integration, reuse, and sharing of data across digital infrastructures.

- Proactive metadata management—including the implementation of National Metadata Profiles—is improving consistency and discoverability of datasets.
- Initiatives such as Transparent Information Management and Exchange modernize how the state shares information both internally and with citizens and businesses.

Supporting Holistic Data Governance and Open Data

- National Data Catalogues and Public Data Portals centralize and describe public datasets, making them more accessible and usable for government, business, civil society, academia, and citizens.
- Participation in the Open Data Maturity Assessment advances standards for interoperability, portability, and data quality.
- Federal and regional open data repositories, or "data lakes," structure data using common exchange standards (enabling portability and robust analytics).

Guaranteeing Quality and Professionalism

- Technical rules and guidelines in public administration are in place to ensure data accuracy, completeness, and timeliness.
- Certification schemes, such as those under the Framework Act on the Promotion and Use of the Data Industry, verify and uphold data quality.
- The National Institute of Statistics and Geography exemplifies rigorous quality assurance with standardized data collection, processing, dissemination, and regular audits—ensuring impartiality and reliability.

Sector-Specific Initiatives

- In healthcare, a National Health Data Space connects regional health systems and improves clinical data access and research through standardized, ethical data sharing.
- Open Finance initiatives standardize and secure the sharing of financial data between authorized entities, based on customer consent.
- A Centre of Excellence for Data Sharing & Cloud supports interoperability within sectoral data initiatives and federated cloud systems.
- Geographic data accuracy and interoperability are enhanced through dedicated APIs for geographic normalization.

Promoting Innovation and Efficient Data Use

- National Data Banks, including data lakes and marketplaces, enable structured, standardized access to a broad array of datasets via APIs, fostering integration and reuse.
- Digital identity projects, such as the ID4D Project, harmonize and validate citizen data from multiple sectors in line with international standards.

Together, these efforts reflect robust progress across the G20 toward high-quality, interoperable, and portable data ecosystems, with a focus on citizen empowerment, responsible innovation, and effective governance.

An example of the imperative to keep data governance abreast of such changes is in the realm of laws and institutions that govern the right to freedom of expression and access to information and the right to privacy. Many jurisdictions operate silos in these realms, which has the effect of pitting privacy against access to information and granular data, rather than considering the ways in which these can be balanced. A number of countries also remain within a paradigm of confining access to information (and data) regimes to apply only to the public sector's holdings. This is notwithstanding the existence of compelled disclosures such as company filing registries. The Aarhus Convention provides for public interest access to specific private sector data affecting the environment. The Escazú regional agreement sets out access to information about environmental matters in Latin America and the Caribbean also applies to private organizations that receive public funds or benefits (directly or indirectly) or that perform public functions and services. Data governance interests can unlock data sharing by both public sector and private actors, with due safeguards for the various rights involved. Here is a tool to assess implicated legislation:

10 steps to ensure that access to information and data protection laws are fit-for-purpose in enabling rights-protecting data sharing.

1.	Identify stakeholders in the data privacy, security and access spaces and consult them on the purpose of governance regimes in regard to contemporary objectives of fostering data access, sharing and innovation while respecting personal privacy and data security imperatives.
2.	Assess the actual impact of existing legal regimes, regulatory architecture and licensing options on these objectives.
3.	Identify current gaps such as legal and financial frameworks which could, if introduced, mandate, incentivise or improve data access, data trusts and data co-operatives

4.	Assess challenges in data markets, in terms of competitiveness, individual consent, safety standards, and in terms of data security, storage and transmission.
5.	Implement low-hanging fruit such as data portable formats and Open Data in the public service while remaining compliant with existing laws, and develop flexible licensing regimes.
6.	In the light of above, review if existing laws and regulations are fit for purpose.
7.	Call for and assess public submissions about possible amendments and about levels of readiness and needs for capacity on the data "demand" side.
8.	Embark on appropriate reform of legal and institutional provisions with tailored mechanisms for transparency, oversight and redress.
9.	Operate systematic impact assessments of data governance arrangements, including especially as relevant to algorithmic and AI processes.
10.	Develop/revise standards for harmonised data anonymisation and pseudonymisation.

The example above is domain specific, but in general there is a need to ensure that data governance is regularly reviewed in order to keep abreast of rapid change. The tool below unpacks what's needed – and is particularly relevant to domain specific governance in G20 topics of interest:

1.	Practice foresight, scenario planning, and risk-opportunity assessments
2.	Continuously monitor and audit compliance with governance regimes and assess reasons for shortfalls
3.	Adopt a change management footing and engage stakeholders regularly
4.	Monitor and revise fitness-for-purpose of existing roles and responsibilities in the data governance architecture
5.	Ensure systems that can afford agile updates to data governance

These five steps entail examining how data governance and AI governance impact upon each other, and they also give structure to how AI tools may assist in these different operations.

6. DATA GOVERNANCE: THE GENERAL AND THE PARTICULAR

General (transversal) data governance issues

The AITF dialogue in April 2025 proposed that governance ensure transparent and accountable data management throughout the entire AI value chain, from data collection and annotation to deployment and monitoring. Within this perspective, there are several common and interdependent challenges that apply to data governance regardless of the domain or issue. Here is a checklist of the objects of data governance, and related actions, which cut across sectoral and topical differences:

1.	Data quality and integrity: Ensuring data is accurate, representative, complete, consistent and reliable. Governance needs to facilitate that there is multi-cultural and multilingual data. [2]	☐
2.	Ethical data use: Addressing biases, potential for discrimination, and societal impacts of data processing and algorithmic decision-making	☐
3.	Data access and sharing: Breaking down data silos, facilitating necessary and secure data exchange and overcoming obstacles to sharing.	☐
4.	Data security and privacy: Protecting sensitive information from unauthorized access, breaches and misuse, while complying with data protection laws.	☐
5.	Data standards and interoperability: Addressing the lack of common formats, definitions and technical compatibility between different systems and organizations.	☐
6.	Data ownership and accountability: Clarifying roles and responsibilities for data management, stewardship and governance.	☐
7.	Regulatory complexity and compliance: Navigating a growing landscape of data-related laws and ensuring compliance as well as making these fit-for-purpose and aligned with international human rights law	☐
8.	Data Literacy and capacity: Raising awareness and promoting standards and programmes to deal with Insufficient skills among users and decision-makers to understand, manage and leverage data effectively.	☐

[2] The issue note for the AITF discussion on dialogue reflects proposals for "mandatory data representation audits for public AI systems, progressive data taxation frameworks that require multinational technology corporations to contribute to national AI development funds where they extract data value, and the creation of regional data commons that prioritise local language datasets and culturally relevant training materials".

9. Digital divide and infrastructure gaps: Tackling uneven access to technology, connectivity and digital tools, that limit data participation and benefit.	☐
10. Data lifecycle management: Effecting society-wide and consistent policies for data creation, storage, use, retention, and secure disposal.	☐

Specific data governance Issues per societal domain

While the generic issues are embedded across data governance in general, their particular challenges and implications, vary across different domains. Within a comprehensive data governance framework, there are specific "affordances" or positive capabilities within each domain. These represent the beneficial outcomes and opportunities that can come from tailored and applied data governance. The possibilities will need to comply with the broader societal data governance framework by inherently respecting general principles such as privacy, security, fairness, transparency and accountability, but will also interpret and prioritise these according to the domain-specific character.

Moving from the question of what data governance covers, the next sections address further questions and how these relate to G20 concerns.

7. WHY: ALIGNMENT WITH A VISION AND PURPOSE OF DATA GOVERNANCE

Data governance is not just about using data for efficiency in isolation of human rights and sustainable development. That depends on the extent to which data governance relies explicitly on a high-level vision that reflects a society's objectives and offers an overall framework on how its data assets can contribute to national goals.

Such a framework should:

- Be comprehensive, integrated, and regularly updated
- Offer guidance at both generic and specific levels
- Inform the design of policies, strategies, and rules
- Shape institutional implementation and coordination mechanisms

The 16 responses to the July 2025 G20 survey on data governance revealed that most participants have implemented multi-layered governance structures and legal measures. However, few countries have adopted a truly overarching policy for data governance. According

to one respondent, this lack of an integrated framework has complicated alignment, harmonization, and enforcement across sectors.

Several participants acknowledged additional challenges. One noted that the existence of multiple, fragmented policy documents not only creates complexity but also requires constant revision and updating. Another respondent, while describing their national approach as comprehensive, observed that the various instruments in place do not amount to a unified strategy or framework that addresses the complete data lifecycle, quality management, ethics, innovation, and cross-sectoral data sharing.

As a result, the primary focus in many countries remains on personal data protection and government transparency, rather than holistic data governance.

Finally, a further respondent reported that, in their context, data governance is generally addressed only as part of broader considerations related to AI, rather than as a distinct policy area.

This picture informs the need to continuously strive for a comprehensive, integrated and updated data governance framework that gives guidance at the generic as well as specific levels. Such a framework should be based on a wider vision, and serve to inform more specific policy, strategy and rules, as well as institutional implementation and co-ordination. Among a range of G20 objectives, the following considerations can factor into a purposive and stable vision for a governance framework. As elaborated in the Broadband Commission Toolkit:

Strategic Objectives for a Vision-Aligned Governance Framework

- Achieving the Sustainable Development Goals (SDGs)
- Promoting Open Data and Digital Public Infrastructure (DPI)
  – Enhancing transparency, accountability, and citizen engagement
- Enabling Data Free Flow with Trust (DFFT)
  – Facilitating cross-border data sharing while preserving privacy and sovereignty
- Protecting Vulnerable Groups and Sensitive Data
- Mobilizing Data for Crisis Response
  – Improving preparedness and responsiveness to emergencies
- Harnessing Artificial Intelligence Responsibly
  – Ensuring fairness, accountability, and transparency in AI systems

The rationale for a vision to inform a data governance framework is also highlighted by consideration of key risks and challenges in digital transformation such as:

- Overdependence on a limited number of service providers for data storage and processing
- Resilience challenges, including power outages and natural disasters
- Technological obsolescence of systems and standards
- Cybersecurity threats and digital vulnerabilities

Against this background, producing or revising a vision for a data governance framework will likely set out the purposes of such governance such as fostering innovation, ensuring privacy and enhancing trust in digital transformation. Not all societal visions, nor related data governance frameworks, will include the challenge of inequalities, but this is something put forward as relevant by the 2025 G20.[3]

> Responses from G20 participants to the July 2025 online survey about data governance included the observation: "Data access and interoperability remain challenges, especially for local decision-makers." The point was added: "health data registries could theoretically inform equity-focused policies, but integration with broader social data is still minimal." It was proposed that "governance advances should be envisaged through strategic integration: linking environmental, economic and social data; improving data accessibility across agencies and regions".

Data governance in regard to equality relates to both the exclusions in data, and the purposes to which data is put. It intersects with Digital Public Infrastructure which is essential for distributing digital opportunities fairly and widely. It aligns with issues of gender equality, and cultural diversity in digital transformation.

Digital public infrastructure is increasingly recognized as an enabler of effective data governance, contributing to:

- Trust: legal frameworks for data protection and oversight authorities.
- Interoperability: shared digital platforms serving as backbones for integrated services.
- Inclusion: social registries ensuring marginalized groups are incorporated into policy design and benefits

[3] In relation to inequality, the 2025 G20 introduced in the DEWG Issue note the perspective that: "Data justice requires redressing not only the possible harms that can occur from the use of data-driven technologies deployed in developing DPIs but also to address the possible uneven distribution of opportunities that can arise in deployment of DPI solutions, such as procurement, R&D, and research."

- Innovation: open data initiatives and portability frameworks stimulating entrepreneurship and new services.

To assess if a framework does justice to this issue in a granular way, the following tool shows how objectives apply to this focus and enable related policies and practices such that line ministries shape purposive data governance across each other.

Checklist: data governance for addressing inequalities

Data governance focus area	**Does the governance framework contain action points such as:**
1. Promote inclusive data collection and representation	Mandate the collection of disaggregated data (e.g., by gender, ethnicity, income, geographic location) to reveal existing inequalities and identify underserved populations. Call for guidelines and incentives for data sourcing that ensures representation of marginalized groups and prevents data gaps that perpetuate invisibility.
2. Mandate algorithmic bias audits to assess possible roles of underlying data:	Develop and enforce frameworks for algorithmic fairness, ensuring that data used for AI development is representative and that models do not inadvertently perpetuate or amplify existing inequalities.
3. Foster equitable data access and benefit sharing	Establish data trusts or co-operative models that empower communities, especially in developing regions, to collectively own, manage, and benefit from their data. Create policies that ensure fair value exchange for data shared from individuals or communities, particularly where data is used to generate significant economic value.
4. Invest in digital infrastructure	Prioritize investment in accessible digital infrastructure and connectivity in underserved domestic regions and developing countries to bridge the digital divide and to enable data participation and improve access to social services and economic opportunities.
5. Harmonize cross-border data flow regulations with equity focus	Promote action for international data governance agreements that balance legitimate cross-border data flows with strong data protection and sovereignty principles, preventing data exploitation. Develop mechanisms to ensure that data transfer agreements explicitly address equitable benefit-sharing and do not disadvantage

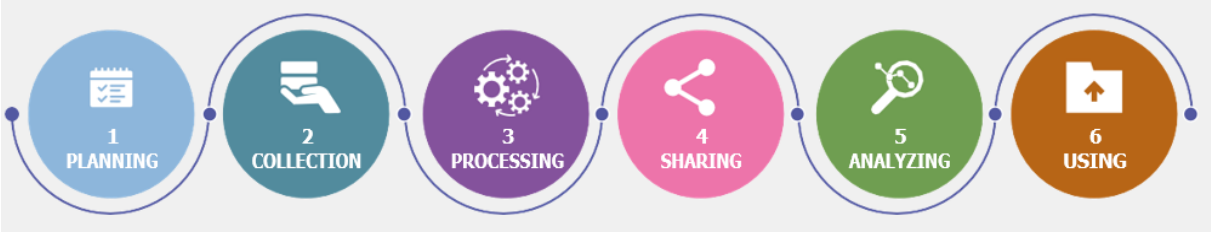
	countries that may have less developed data infrastructure and/or governance frameworks.
6. Establish ethical guidelines for data use in development	Mandate national and international ethical guidelines for the use of data in development initiatives, ensuring that data collection and analysis actively contribute to poverty reduction and social equity. Use data insights to inform equitable resource allocation for social welfare programs, ensuring that funds are directed to areas and populations with the greatest need Implement oversight mechanisms to ensure that data-driven development programmes respect universal human rights and avoid unintended negative consequences for vulnerable populations.
7. Support data sovereignty and local data governance models	Recognize and respect the rights of indigenous communities and marginalized groups to govern their own data according to their cultural values and self-determination principles. Encourage and support the development of localized data governance frameworks that reflect the specific socio-economic contexts and needs of diverse communities.
8. Enhance data literacy and critical data skills for all	Provide for public awareness campaigns to educate citizens, especially those in vulnerable groups, about their data rights, potential risks, and opportunities in the data economy. Ensure data governance strategies account for varying levels of digital literacy and access among beneficiaries of social welfare programmes, advocating for diverse data collection and service delivery channels (e.g., physical access points, community workers). Integrate critical data thinking and digital citizenship into educational curricula to empower individuals to navigate the data-driven world more effectively. Implement comprehensive data literacy programs and technical training for individuals and institutions in areas with limited data capacity, fostering local data expertise.
9. Implement transparent and accountable data practice	Ensure transparency in data governance processes, including clear policies on data collection, use and sharing by both public and private entities. Establish accessible grievance and redress mechanisms for individuals and communities who believe their data rights have been violated or who have experienced harm due to data-driven systems.

10. Prioritize funding for data-driven solutions to inequality	Direct public and private investment towards research and development of data solutions to address structural inequalities (e.g., precision public health, inclusive financial services). Create incentives for organizations to use data responsibly and innovate in ways that reduce, rather than perpetuate, social and economic disparities. Establish clear data metrics and reporting frameworks to measure the social impact and effectiveness of interventions, promoting evidence-based policy-making.

8. WHAT GOVERNANCE COVERS ACROSS THE DATA LIFECYCLE

There are different conceptions of this "lifecycle". For this G20 Toolkit, the "cycle" is not to be understood sequentially, but rather as constituting a set of complimentary lenses to be applied, when appropriate, to the dynamics of data in the digital economy.

Figure 1: Data Lifecycle



Source: Authors.

Data governance decisions are critical at every aspect of the data lifecycle to ensure that data is treated effectively, ethically and in compliance with societal frameworks. The data lifecycle generally encompasses aspects such as creation/collection, storage, processing/analysis, use/sharing, archiving, and destruction. Drawing in part from the Broadband Commission's toolkit, the following points give insights on this issue in general, preparing the stage for assessing the issue- and domain- specific applications later in this document.

- Producing plans is an ongoing consideration in the data lifecycle: This involves setting clear objectives, establishing partnerships, and designing a strategic roadmap for establishing and evaluating data governance at general and specific levels.

- Collections and retention of data: The parameters of data generation and acquisition are relevant to the sourcing phase, and these call for governance of datafication that ensures transparent, legal and ethical methods of both collection and storage, as well as relevant

provisions for archiving and destruction. It is of particular relevance to have in place comprehensive records management policies, in alignment with the UNESCO's 2015 Recommendation concerning the preservation of, and access to, documentary heritage including in digital form.

- Ownership: In this dimension, the intellectual property of the data sets needs prior governance to clarify legal entitlements to use and re-use such data, as well as synthetic data arising from such use.

- Data transfers, data markets and data sharing: Governance here sets boundaries and incentives for cross-organizational collaboration, trust-building and data exchange.

- Processing: This covers how data is organized and classified, and subjected to a range of manipulations that afford its transformation into new combinations of data, or into ensuring training advances in AI algorithms that are used in further data processing. These processing actions may result in synthetic data that is fed into further rounds of processing.

- Analysis: Going a step further than processing, this entails actions to generate meaningful insights or outcomes that inform new algorithms or to plan technology responses (as in agentic AI). Analysis may also sometimes include the combination with other data sets and the generation of further sets to achieve insights (eg. new documents are inserted into the system).

- Application/ use: Data-driven insights are translated into effective decision-making (which may be automated and feed specifically into agentic AI) and other actionable outcomes by actors or machines. Governance of the sharing of data is an issue here.

- Transparency: Certain phases and results in the data lifecycle may be technically opaque or actively hidden. Governance will determine what should be transparent and disclosed by default, as well as the conditions for public information requests concerning data across the lifecycle

Data governance is implicated in both the opportunities and risks of Generative AI:

- As AI systems become more capable of creating their own data, making predictions, and generating new insights from vast datasets, data governance frameworks need to evolve. Generative AI also introduces new complexities related to data sourcing, model transparency, ethical use and regulatory compliance that traditional data governance models may not fully accommodate.

- For example, data provenance and quality are emerging as critical issues to avoid perpetuating biases or producing harmful outputs. In addition, data governance must seek to address issues such as copyright, and liability and accountability for AI-generated content (e.g., misinformation or wrong advice).
- The degradation of synthetic data, based upon earlier synthetic data, over time creates a risk that data governance needs to tackle.

Technical and Human Capacity: A Core Pillar in Data Governance

G20 survey responses underscore that strengthening technical and human capacity is critical to effective data governance throughout the data lifecycle. Capacity-building efforts, as reported by member countries and partners, span several interconnected domains:

1. Strengthening National Research and Innovation Ecosystems

- Countries are investing strategically in national research systems, including improving conditions for international researchers and funding high-performance computing infrastructure.
- Many participants support research, development, and innovation in digital technologies and services central to the data economy.

2. Workforce Upskilling and Education

- Several countries report dedicated resources to upskill public sector employees, incorporating data science and AI training into workforce development and higher education curricula.
- Initiatives led by national statistics institutes, often in collaboration with universities, provide hands-on capacity-building and explore innovative methodologies (e.g., through Big Data Labs).
- Sub-regional ecosystems are fostering expertise in high-performance computing, digital health, and climate data.
- Universities are offering specialized degree programs in data science and AI and are adapting curricula to meet public sector needs, sometimes with a strong interdisciplinary focus linking data, policy, economics, and the social sciences.

3. National Strategies and Targeted Programs

- National Digital Transformation Strategies often direct investments and policies toward expanding data science capacity, funding R&D, supporting AI-focused programs, and creating innovation networks.
- Shared research infrastructure platforms are being launched to lower the barriers for data science collaboration and access.
- Several countries support scholarships and interdisciplinary laboratories as part of a comprehensive national AI plan.

4. Large-Scale Digital Skills and Talent Programs

- Ambitious initiatives such as the 3 million Technical Talent (3MTT) Program aim to train millions of citizens in digital skills, including data science, AI, and software development, over a short period.

5. Institutional and Governmental Capacity Building

- Open Data Directorates and similar agencies are leading data science capacity development within public administrations—providing targeted training in technical and methodological topics such as open data management and API utilization.
- Federal government data laboratories are cited as exemplary in advancing technical capacity in government.

6. Regional and Global Initiatives

- The European Union prioritizes digital skills development through its European Data Strategy, the Digital Decade Policy Programme, and the Digital Europe Programme, all of which fund education and foster "test-before-invest" digital innovation hubs that train and advise organizations on data science and AI adoption.
- The African Union, working through AUDA-NEPAD and its Human Capital and Artificial Intelligence Initiative, invests in data science capacity-building and policy training, as seen in initiatives like the AI Policy and Regulatory Training for African Policymakers in Abuja, 2025.

These collective efforts reflect a growing understanding within the G20 and partner regions: that robust and sustained investment in human and technical capacity is essential for effective, inclusive, and future-ready data governance. As highlighted in recent G20 policy briefings, such capacity-building is foundational both for leveraging data as an engine of innovation and growth and for ensuring that data governance frameworks are coordinated, ethical, and aligned with broader social and economic goals.

9. HOW: KEY STEPS IN DOING DATA GOVERNANCE

Mechanisms are essential preconditions for an effective data governance framework. Adapting from the Broadband Commission Toolkit, the following have relevance:

To effectively implement data governance principles and decisions throughout the data lifecycle, a robust framework should deploy a range of complementary mechanisms. Key elements include:

1. **Policies and Guidelines:** Foundation-setting principles that govern data collection, use, sharing, and disposal.
2. **Technology and Governance by Design:** Embedding governance requirements directly into IT systems and digital architectures.
3. **Standards and Common Vocabulary:** Facilitate interoperability and consistent understanding across sectors and borders.
4. **Codes of Conduct:** Voluntary or enforced behavioral norms for data handling by individuals and organizations.
5. **Licensing Arrangements:** Define permissions and restrictions for data reuse, redistribution, and derivative works.
6. **Data Stewardship and Institutional Arrangements:** Assign responsibility for data management, ensuring alignment with governance objectives.
7. **Audit and Compliance Mechanisms:** Monitor adherence to policies, with periodic reviews and accountability measures.
8. **Training and Cultural Change Initiatives:** Build capacity and foster a data-aware organizational culture.
9. **Contractual Mechanisms:** Embed clear responsibilities for access, sharing, use, and handling of data—including API terms and third-party access—in contracts and agreements.

Special Focus: Data Governance Integration in Procurement

For G20 countries, integrating data governance requirements directly into procurement processes is an increasingly critical mechanism. This includes requiring vendors to disclose:

- The provenance and quality of data used to train systems.
- Risk assessments, including evidence of stress testing and red-teaming conducted on systems.
- Intellectual property rights of any new datasets or outputs generated by the system.
- Intended use and management of metadata and derived data arising from processing activities.
- Provisions for public sector data-sharing, ensuring that arrangements for access and re-use are addressed in contract terms.

This approach not only ensures alignment with national and international governance objectives, but also supports greater transparency, accountability, and public value when engaging with technology providers.

Data Governance implementation:

Data governance requires sustained investment and assessment of compliance costs and the returns on spending. Financing mechanisms can include public budgets, grants, debt, partnerships and voluntary participation. Some expenses can be shared, such as in areas of capacity building and cyber security. There is a need to anticipate and amend resourcing in the face of new developments around the data lifecycle. Because data governance is not a once off activity, criteria will be needed to integrate financing into wider budgetary processes and covering both generic and domain-specific outlays. The checklist below may be of value in assessing data governance costs, beyond software and technology, as linked to a number of the G20 priorities in 2025:

G20 2025 Priority	**Amongst the data governance costs to be considered**
Inclusive economic growth, industrialisation, employment, and reducing inequality	1. Create data tools for MSMEs to access markets and finance 2. Roll out data literacy programs for youth and informal workers 3. Build open data platforms for marginalized communities 4. Develop ethical AI guidelines for employment and labour rights 5. Monitor progress via disaggregated data dashboards
Agriculture and food security	1. Support interoperable agricultural data platforms for precision farming 2. Establish data-sharing protocols protecting farmers' rights 3. Develop climate-resilient data tools for smallholders 4. Implement blockchain for supply chain transparency 5. Promote open data standards for crop yields and pricing 6. Train farmers in data collection/analysis 7. Create data rights frameworks for agricultural communities 8. Build early warning systems using satellite/soil data 9. Audit food system data biases annually
Artificial Intelligence and digital technologies for	1. Host annual Data Governance Dialogues 2. Develop open-source AI tools for public interest applications 3. Support global AI talent exchange 4. Fund R&D into resource-efficient AI systems 5. Create sandbox environments for ethical AI testing

sustainable development	
Climate action and just energy transition	1. mplement standardized environmental data reporting 2. Build open platforms for climate finance tracking 3. Develop AI tools for carbon and methane emissions monitoring 4. Mandate corporate climate data disclosures and monitor enforcement 5. Negotiate global data-sharing agreements for emissions 6. Integrate indigenous ecological knowledge into datasets 7. Audit energy companies' environmental data practices
Heritage restitution and culture	1. Enact cultural data sovereignty policies 2. Develop frameworks for cultural data rights management 3. Build open databases for heritage restitution claims 4. Establish indigenous data governance protocols 5. Conduct cultural diversity audits in AI training data 6. Digitize heritage artifacts with community consent 7. Conduct annual audits of cultural data biases
Debt sustainability and global financial architecture reform	1. Develop AI tools for debt sustainability analysis 2. Build interoperable debt data platforms 3. Train officials in financial data governance
Addressing polycrises (climate, energy, food, debt)	1. Develop interoperability frameworks for crisis data 2. Create multi-agency data-sharing platforms 3. Standardize crisis impact metrics 4. Build real-time polycrisis dashboards 5. Negotiate cross-border data-sharing agreements 6. Train officials in crisis data management 7. Monitor misinformation risks in crisis datasets 8. Develop predictive analytics for crisis prevention 9. Conduct annual polycrisis simulation exercises 10. Establish post-crisis data review processes
Promoting solidarity, equality and sustainability	1. Develop inclusive DPI implementation roadmaps with attention to data governance issues 2. Establish community data trusts for resource distribution 3. Align all data governance with SDG tracking 4. Enforce corporate data transparency registers 5. Monitor equity impacts through disaggregated metrics

The HOW questions require that data governance framework anticipate the use of tools applicable to different aspects in the life cycle. The Broadband Commission Toolkit sets out links to valuable resources for governing data collection, storage, protection, access management, evaluation of assets and processing. Looking ahead, the toolkit also flags future issues:

Emerging Developments

- **Decentralized Storage Networks:** (Blockchain-based) systems decentralize data storage, to enhance resilience and security.
- **Data Mesh:** Decentralized data architecture to promote team autonomy and scalability.
- **Edge Computing:** Enables local data processing, reducing latency and ensuring real-time decision-making.
- **Data Products:** Pre-prepared, reusable, and modular datasets designed for specific use cases to streamline analysis and decision-making.
- **PETs in Processing:** Technologies like federated learning and secure multi-party computation to ensure secure collaborative processing.

Governance here can also draw upon UNESCO's resource on Open Data and AI. Data sharing arrangements need careful attention to ensure trust, human rights-alignment, and secure collaboration between state agencies, and arrangements within external partners. Where data markets are involved, with specialist data broker companies aggregating sets from different sources, triangulation needs governance if there is to be respect for privacy rights. Gatekeeping markets (such as digital advertising exchanges) may need governance interventions as well as regulation of micro-targeting through exploiting mixes of acquired (and live) data points. Guidelines on access to data, produced by Research ICT Africa and CETIC.br for the G20 Digital Economy Working Group, constitute a useful tool for this aspect of data governance.

10. WHO: PEOPLE AND ROLES

Ensuring Decision Provenance in Data Governance

For governments developing, implementing, or refining a data governance framework, a critical requirement is establishing **decision provenance, t**he ability to trace how data-related decisions are made, by whom, under what authority, and through which processes. This is essential for fostering accountability, trust, and coherence across institutions and society.

To ensure decision provenance, a data governance framework should include:

- **Defined Institutional Roles and Mandates**
  Clear articulation of which entities—governmental, corporate, or non-profit—are empowered to make decisions, enforce policies, and oversee implementation. For instance:
  - A Data Protection Commission should have a clearly defined legal mandate and responsibilities, along with the authority of Data Protection Officers.
  - A Right to Information regulator must have a well-scoped remit, and where functions are converged (e.g. privacy and access), the framework should clarify how overlapping authorities are resolved.
  - The role of data stewards must be positioned within the broader governance ecosystem, ensuring alignment with institutional responsibilities.

- **Mechanisms for Coordination and Oversight**
  Provenance depends on transparent, traceable decision-making across the entire system. This requires:
  - Inter-ministerial task teams and formal mechanisms for cross-agency coordination
  - Ongoing stakeholder engagement that documents inputs from civil society, the private sector, and the public
  - Procedural clarity around how decisions are proposed, reviewed, adopted, and enforced

- **Mapping the Data Governance Landscape**
  Decision provenance also involves identifying where decisions are made throughout the data lifecycle. A robust framework should highlight:
  - The role of private actors, including technology providers and platforms
  - Governance of cybersecurity, including mandates during crises or cyber incidents

- The responsibilities of electoral bodies, especially where data influences democratic processes
- Points of convergence, conflict, or ambiguity across institutional mandates

Ultimately, decision provenance provides the backbone of a joined-up governance architecture. It allows societies to not only understand who governs data and how, but also to ensure that decisions are made transparently, responsibly, and in line with national values and development priorities. A key consideration is whether formal mechanisms exist to coordinate data governance across ministries, regulators, and non-state actors, ensuring traceability and accountability throughout.

Institutional Arrangements and Coordination for Data Governance

Survey responses from G20 participants reveal a diverse landscape of regulatory authorities, sector-specific regulations (such as those governing health data), and a broad range of soft law and voluntary practices. Many countries operate within complex institutional environments, where legal mandates and authority over data governance are distributed across multiple bodies.

A recurring theme is the **need for greater integration and coherence**. As one respondent noted, "While efforts are ongoing to improve coordination, there is still a need for a more integrated framework that aligns these diverse policies under a unified data governance strategy." This challenge is particularly notable in regional blocs such as the European Union and the African Union, where relationships between supranational and national institutions must be carefully managed. Varying degrees of autonomy are granted to regulatory authorities, and coordination processes often differ substantially among and within countries.

Current coordination mechanisms include:

- Platforms enabling secure and standardized data exchange among public administrations.
- Requirements for federal agencies to submit Open Data plans.
- Centralized coordination by ministries at the national level, often complemented by decentralized structures at the regional or local level, such as dedicated data offices responsible for implementing local strategies and fostering inter-institutional collaboration.
- In some settings, ministerial departments actively coordinate with industry, academia, and local governments to address major issues, ensure security, and manage data systematically across its lifecycle.

- Open Data Directorates in some countries facilitate engagement with civil society, organizing discussions with open data experts and promoting state–citizen collaboration.
- Institutional architectures exist for Chief Information/Data Officers and Directors-General to connect ministries, agencies, and local governments, while in other cases, deputy ministerial or agency heads provide overall guidance and facilitate coordination on data management among various participants.

However, gaps remain. In several cases, there are **no formal mechanisms for coordination** between state and non-state agencies, although Ad hoc or consultative bodies may exist. Respondents frequently cited **ongoing efforts to streamline roles, reduce overlaps, and enhance synergy** among ministries, regulators, and non-state actors as part of a broader push toward a more integrated data governance ecosystem. In some countries, coordination mechanisms are currently being redefined and structurally adapted in response to evolving needs and circumstances.

In summary, while models of institutional coordination vary, G20 countries widely recognize both the challenges and the necessity of improving coordination to ensure consistency, accountability, and effectiveness in data governance.

Summary of Characteristics of G20 Data Governance Architectures:

- **Hybrid Models:** Most G20 countries do not follow a purely centralized or decentralized model but a hybrid (federated) approach where a central authority sets broad rules, and sector-specific bodies or agencies implement them.
- **Varying Degrees of Centralization:** Some countries exhibit higher degrees of harmonization and centralization in data protection, while others have a more fragmented, sectoral, or state-level approach.
- **Focus on Trust and Innovation:** G20 discussions often emphasize balancing data protection and trust with enabling innovation and data flows for economic growth. This tension often shapes the roles and mandates of different institutions.

In essence, the institutional architecture for data governance across G20 countries is a dynamic network of interconnected, yet often independently operating, entities that collectively aim to manage data as a strategic asset while safeguarding fundamental rights. A general overview of the institutional architecture relevant to data governance in G20 countries shows the following kinds of institutional mechanisms:

ore Regulatory and Policy-Setting Bodies	
1. Parliamentary committees; Ministries of Digital/ Communications/Innovation	Formulating national digital strategies, policies, and legislation related to data, digital economy, cybersecurity and AI.
2. Data Protection Authorities (DPAs)	Primary enforcers of data privacy and protection laws
3. Information Regulators	Primary adjudicators of data inventories and access
4. National statistical commissions	Sets standards for official statistics, collects and disseminates public data, ensuring quality and confidentiality.
5. Government Chief Data Officers (CDOs) / Cross-governmental IT agency managing data governance	Promotes government-wide data quality, interoperability and strategic data use. May negotiate cross-border data transfer accords in conjunction with foreign ministries / departments. Oversight of state technology systems, including procurement, and ensuring support for data security, pooling and API access.
6. National Human Rights Institutes	Advise, monitor and educate on the rights dimensions.

Sector/issue specific Regulators	
1. Financial Regulators	Data management, security and reporting to ensure financial stability, combat money laundering and protect consumers.
2. Health regulators/ministries	Sets standards for health data privacy, interoperability of electronic health.
3. Competition authorities	Addresses data monopolies and anti-competitive practices related to data access, and ensures fair competition in data-driven markets.
4. Copyright authority	Oversees intellectual property rights related to data.
5. Consumer rights authority/ ethics office/ ombudsman	Competence to settle data disputes unresolved by other institutions.

Cross-Cutting and Supporting Institutions	
1. Cybersecurity agencies	Protects critical data infrastructure, responding to cyber threats, and develops cybersecurity standards that underpin data security.
2. National research entities and academia	Includes focus on open data and data science development. Contributes to theoretical and practical understanding of data governance, develops ethical frameworks, conducts independent analysis, and builds capacity.
3. Government training college / institution dedicated to advancing data literacy within government and the wider society	Implements initiatives to raise awareness and capacities around data rights, risk and opportunities Liaises with education and training providers to integrate data literacy into their offerings.
4. Crisis response units	Sources and uses data in emergency contexts
5. Multi-stakeholder forums and industry associations	Facilitates dialogue between government, private sector and civil society.

The Broadband Commission Data Governance Toolkit: Navigating Data in the Digital Age provides detail on civil service roles that implement data governance. In summary, the functions for officials in data governance include:

1. Development and coordination of governance standards and practices

2. Ensuring compliance and adjudication of disputes

3. Facilitation and management of data governance (such as nominated data stewards to be responsible for managing quality, security, access and use)

4. Review, evaluation, training and guidance

11. WHERE AND WHEN: ROADMAP TO FACILITATE IMPROVED DATA GOVERNANCE

Data governance is not a one-time effort. it requires continuous review and adaptation. As technologies evolve rapidly, from the growing influence of agentic AI to the transformative potential of quantum computing, governance frameworks must remain agile and responsive. This section of the Toolkit introduces a self-assessment tool designed to help governments and institutions evaluate and update their data governance practices in light of emerging challenges and opportunities.

AI agents prompt a new look at data governance

Agentic AI amplifies the need to move beyond static data governance. This is because this use of AI entails decentralization and autonomous automation that ingests, infers and links large volumes and varieties of data. Agentic systems can independently access and process data from a range of often siloed sources. Traditional governance issues such as requiring data minimization and purpose specification are put into question. This makes it challenging to track exactly what data is being used, for what purpose, meaning that the risk of unintended data exposure or misuse increases significantly. This also complicates the issue of alignment with ethics and privacy regulations.

The agent-active technology operates proactively and adapts in real-time to reach goals which are more primary than any built-in compliance with data governance regimes. To the extent that AI agents are designed to respond to individuals' agency, they will operate on the basis of mammoth troves of data about each individual and also draw on extensive computer memory to provide tailored services.

For these reasons, data governance becomes even more complex than ever, and entails dealing with dynamic real-time changes in the data lifecycle. It calls for continuous monitoring, real-time risk assessment, and dynamic policy enforcement.

At the same time, agentic AI could be deployed to assist in addressing some of the challenges. This could be in at-scale automation of metadata tagging data quality checks and detecting anomalies, the tracking of provenance and the monitoring of rule compliance. Technically, it can work across hybrid data sets and cloud storage environments. Certain Agentic AI tools could highlight data bias where this affects algorithmic outcomes. Efficiencies and innovations could be enhanced through agentic use of AI systems. However, if humans are not involved in

oversight, review and appeal, there is a danger of data decisions being taken within a "black box" beyond any effective governance.

A selection of G20 experiences:

Assessing the Effectiveness of Data Governance: G20 Approaches and Metrics

In response to the question, "Does your country utilize specific metrics or processes to assess the effectiveness of its data governance across various sectors?" G20 participants described a diverse array of methodologies and indicators, which can be grouped as follows:

1. Comprehensive Frameworks and Models

- Many countries use structured methodologies, such as Data Maturity Models and indices, to evaluate the benefits, quality, and effectiveness of data governance initiatives across the public and private sectors.
- For example, a National Data Index may assess entities on data quality, compliance, sharing, and privacy, while the European Interoperability Framework enables countries to track progress in digital governance, AI adoption, and data trust.

2. Performance Metrics and Key Indicators

- Digital agencies often apply key performance indicators (KPIs), such as platform usage, data integration levels, and the impact on public services.
- Sector-specific assessments include metrics on electronic health record integration in healthcare, the effectiveness and fairness of AI in justice, and annual ratings of digital public service provision by region.
- The business sector is implementing digital maturity assessments and compliance metrics for data protection and cybersecurity.

3. Data Quality, Openness, and Transparency

- Tools like the Open Data Barometer, Open Government Partnership metrics, and Open Government Index are used to measure the availability, accessibility, and usability of government data, often with multi-dimensional and quantifiable indicators.
- Ongoing efforts include the development of indices to evaluate the presence and quality of open data in public administration.

4. Oversight, Compliance, and Accountability

- Data Protection Agencies rely on quantitative and qualitative indicators—such as internal dashboards for user engagement and system performance—and facilitate transparency through public reporting.
- The institutionalization of Data Protection Officers (DPOs) in government and private organizations fosters accountability and improved data handling.
- Metrics include the number and value of fines or sanctions imposed for privacy and data protection violations, and auditing systems for data breaches or regulatory compliance.

5. Specialized Auditing and Sectoral Assessment

- Research centers benchmark algorithmic transparency, bias detection, and fairness, especially for AI systems.
- Some agencies conduct risk assessments, impact assessments for high-risk AI, and dedicated audits for compliance and ethics.
- Feedback from civil society and NGOs—often monitoring and raising issues publicly—serves as an informal yet vital form of assessment.

6. Regional and International Peer Review

- The African Union's AUDA-NEPAD promotes metrics and peer review processes such as the Sectoral Digital Readiness Index, which measures progress in data infrastructure, regulatory frameworks, digital skills, and ethical AI deployment across Member States.
- AUDA-NEPAD also conducts AI regulatory readiness assessments, particularly in healthcare, to evaluate legal, institutional, and data protection capacities for digital health.

7. Emerging and Evolving Metrics

- Some countries are developing new metrics to assess the impact of emerging legislation (e.g., the EU Data Governance Act and Data Act) and to measure national data sharing and trading activity.

Across the G20, the evaluation of data governance effectiveness is increasingly systematic, multi-dimensional, and responsive to sectoral challenges and innovation. Continuous improvement is underpinned by a combination of quantitative indicators, qualitative assessments, peer review processes, and engagement with the wider public.

12. ACRONYMS AND ABBREVIATIONS

AI	Artificial Intelligence
AITF	G20 Task Force on Artificial Intelligence, Data Governance and Innovation for Sustainable Development
AU	African Union
AUDA-NEPAD	African Union Development Agency – New Partnership for Africa's Development
API	Application Programming Interface
CDO	Chief Data Officer
DPI	Digital Public Infrastructure
DPA	Data Protection Authority
DPO	Data Protection Officer
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation (EU)
ID4D	Identification for Development
IP	Intellectual Property
KPI	Key Performance Indicator
MSME	Micro, Small, and Medium Enterprise
NGO	Non-Governmental Organization
OECD	Organisation for Economic Co-operation and Development
PETs	Privacy Enhancing Technologies
PII	Personally Identifiable Information

R&D	Research and Development
SDGs	Sustainable Development Goals
UNESCO	United Nations Educational, Scientific and Cultural Organization

13. APPENDIX A: BACKGROUND

In 2025, under South Africa's presidency, the G20 agreed on a dedicated task force that links the issues, as evident in the title "Artificial Intelligence, Data Governance, and Innovation for Sustainable Development". At the second meeting of the task force in April 2025, a data governance dialogue examined how best to manage and govern data, with a focus on data quality, privacy, security and its ethical use. Also discussed there was the need to align principles, standards and practices for data governance, in order to fully unlock the benefits of data sharing and cross-sectoral interoperability.

G20 discussions in 2025 that further implicate data governance include the topics of "Data Free Flow with Trust" and enhancing data access and sharing in the Digital Economy Working Group. All this is in a context where data is becoming a major differentiating factor for underpinning competitiveness and as a factor for increasing opportunities and imperatives for co-operation and collaboration. This is in regard to building foundational AI models, but also – when application tools are increasingly available to all – in deploying these instruments for particular purposes.

The value of data governance for inclusion and equality was underlined in the AITF's dialogue on data governance at its second meeting held on April 10-11 2025. This toolkit operates on the observation at the dialogue (and reported in the issue note "Making data available for AI") that AI systems require well-governed data to function ethically and responsibly, thereby avoiding the perpetuation of biases, amplification of discrimination. The dialogue noted that the misuse of data that can breach people's rights. The toolkit provides ways to take further the dialogue's call for support to data commons initiatives and inclusive data infrastructure, embracing also the countries in the Global South and across Africa, in order to reduce data asymmetries and foster innovation on fair and equal terms.

Aligning with the AITF dialogue, this toolkit echoes the affirmation that the objective of data governance should be to align principles, standards and practices, and optimise them for human rights and sustainable development. The resource is intended to advance the dialogue's stress on strengthening international cooperation in data governance, enhancing the harmonisation of standards, setting up mechanisms to promote and facilitate trustworthy cross-border data flows to support AI innovation globally, and addressing digital and data divides that limit the potential for AI systems to benefit humanity. It further gives practical suggestions to the dialogue's

discussion on managing and governing data, including aspects such as data quality, privacy, security and its ethical use.

Finally, this kit also rests upon data governance perspectives within wider international developments such as the Global Digital Compact, and the Governing AI for Humanity Report released by the UN Tech Envoy High-Level Advisory Board on Artificial Intelligence, and is cognisant of the UN's Commission on Science and Technology for Development working group on data governance. The tools are further constructed in cognisance of regional initiatives such as the EU's General Data Protection Directive and Data Act, the OECD Guidelines on the Economic Regulation, and the African Union Data Policy Framework – all of which implicate significant norms for the purposes and principles of data governance.

This current toolkit thus complements this broader work on data governance, as well as generic data governance tools such as those elaborated in the Broadband Commission Toolkit developed by UNESCO, the ITU, UNDP and the AU. This specific kit is not intended to duplicate these, nor to be exhaustive. Instead, it focuses upon a small number of selected issues relevant to the G20.

The strategic focus of this toolkit responds to the findings of an online survey of G20 participants completed in July 2025. Of 16 responses, 12 signalled privacy abuse and cybersecurity as among their top data-governance issues. Nine respondents alerted that unlocking data was in their top list.

Almost half of them registered cross-border data transfers in the top tier. Six placed the issue of capacity and skill within their priority issues, while nine others ranked this issue within the category of second level concern. Ten scored intellectual property issues as being of medium level concern. Only three respondents flagged storage and processing among their top challenges.

These findings inform the toolkit's content and emphasis.

14. APPENDIX B: DATA GOVERNANCE AND DIGITAL PUBLIC INFRASTRUCTURE

The toolkit takes close cognisance of G20 work on data governance to date particularly in regard to the promotion of Digital Public Infrastructure and related data issues. The 2024 Ministerial Declaration of the Digital Economy Working Group links data governance to the challenge of governance frameworks for AI. The Leaders Declaration observed that "To ensure safe, secure, and trustworthy AI development, deployment and use, the protection of human rights, transparency and explainability, fairness, accountability, regulation, safety, appropriate human oversight, ethics, biases, privacy, data protection and data governance must be addressed". One

example of the intersection of this Toolkit is with the G20's prior work is on Digital Public Infrastructure (DPI).

The G20 New Delhi Leaders' Declaration defines DPI as a set of shared digital systems that must be secure, interoperable, built on open standards, and promote access to both public and private services for everyone. Examples of DPI include digital IDs and other digital registries, as well as electronic signatures, and public key Infrastructure.

Such DPI is reliant on core databases as authoritative sources of data that is fundamental for administrative processes and services, for example covering people, company record offices, licenses, buildings, locations, roads and vehicles. These infrastructures are often managed by different public sector organisations such as tax authorities, company offices, land registers, statistical agencies and environmental agencies, sometimes in fragmented fashion.

In this context, effective data governance constitutes the foundational structure for enabling interoperability and reusability across these offices, their systems and use-cases. Data governance here can facilitate that data about ID and other registries be reused in sectors like health, education, financial inclusion, land records and many others. In this respect, DPI and data entails challenges and opportunities in:

- Data privacy, security and consent
- Data quality and integrity and use
- Interoperability and portability

Views on data governance as relevant to advancing Digital Public Infrastructure

Responses to an online survey of G20 participants reveal these observations about their countries:

- Development of national digital infrastructure, including a number of common solutions that are being used for data sharing.
- Demonstration of how robust privacy laws can coexist with open, trust-based data ecosystems—vital for secure and accountable DPI.
- National initiatives covering digital ID, digital payments and an app, and these showcase practical implementations of DPI components.
- Promotion of open APIs, interoperable systems, and strong public-private collaboration—critical for scalable DPI solutions.
- A human-centric approach to AI and data use aligns with the G20 goals of ensuring DPI benefits all segments of society.

- An initiative for Transparent Data Management and Exchange supports the creation of high-value public services, with a focus on user-centricity and interoperability. A centralized corporate data warehouse will aggregate both internal administrative data and potentially relevant external datasets, helping to eliminate data silos across ministries and enabling more cohesive, efficient, and informed public sector decision-making.
- Promotion of literacy about data governance with guides and courses in the school of government.
- An autonomous structure and legal mandate for the data authority illustrates how such independence can build public trust and ensure continuity across administrations.
- The creation of a National Alliance for Artificial Intelligence, which resonates with G20 emphasis on inclusive governance models and participatory approaches to DPI development.
- Data management is linked to advancing digital public infrastructure through the creation of conditions for international cooperation in the field of artificial intelligence, digital commerce and data protection.
- The creation of a unified public services platform demonstrates a successful model of providing public services in electronic form based on centralized data management and protection.
- An integrated ecosystem of DPI that combines technological strengthening with open standards and technical capacity development shows digital infrastructure as a multidimensional public good. This ecosystem includes platforms that generate public and private value from a State-Citizen and State-State perspective.