ISSUE BRIEF:

Guidelines for access to data for researchers, promoting data sharing with and by the public and private sectors

August 2025

Digital Economy Working Group (DEWG)

Priority 4: Equitable, Inclusive and Just AI

Deliverable 2.8.3 Guidelines for access to data for researchers, promoting data sharing with and by the public and private sectors

Knowledge resource for the G20 Digital Economy Working Group[1]

Knowledge Partners: Research ICT Africa[2] and the University of Pretoria's Data Science for Social Impact Research Group and Data Science Law Lab.

[1] This document responds to feedback on an earlier version produced for the second DEWG meeting in April 2025, and it updates a revised one that was circulated in May.

[2] RIA acknowledges support from UNESCO for the work in developing these Guidelines. The ideas and opinions expressed here are those of the authors; they are not necessarily those of UNESCO and do not commit the Organization. The paper also benefited from ongoing engagements with African Access to Data Alliance and the African Union Rapporteur on Freedom of Expression and Access to Information.

Executive Summary

Access to high-quality, diverse, and relevant data that is privacy-preserving and secure is crucial for the ethical and effective training of Artificial Intelligence (AI) models as well as for the use of AI applications. Much has been written about transborder data transfers, open data initiatives, data commons, and secure data-sharing platforms, but less attention has been given to governing access to public-interest data in the first place. This document provides focused guidelines which G20 policymakers may wish to consider in order to advance data sharing in both public- and private sector holdings with the aim of unlocking public value within the digital economy. This gives attention to the data value cycle that underpins AI, as well as to other matters such as transparency and accountability.

The guidelines are contextualised within broader frameworks, experiences and co-conditionalities. They offer possible actions relevant to the levels of policy, institutions, and specificities of the public and private sectors, as well as to licensing, technical matters, and data literacy. This knowledge resource contributes to redressing the highly uneven distribution of opportunities associated with advanced data driven technologies. Appendices provide information on foundational issues; background on normative and legal frameworks; and a companion pilot project which seeks to apply the guidelines to access local language data sets for Large Language Modelling in South Africa.

I. Objectives

The purpose of these non-binding guidelines is to assist policy-makers in unlocking public- and private sector data as a public resource, particularly for researchers, start-ups and micro-, small- and medium- enterprises (MSMEs). Within a holistic concept of human rights-based data governance, the guidelines facilitate:

- greater data openness by default;
- elaborating conditions for compulsory disclosures; and
- encouragement of voluntary agreements through equitable licensing.

These three mechanisms can facilitate increased data flow within and between public, private and civil society sectors for public benefit. The optional actions are informed by the need to build trust as a key component to encourage data access and various safeguards are suggested in this respect. The intersection of the access dimension with data flow dimensions is also covered. The points further give special attention to encouraging data access arrangements for key stakeholder groups to engage in both public and private value creation (and combinations thereof). In this vein, the guidelines propose using tools such as Creative Commons and other open and public licensing provisions to facilitate trust and confidence among data holders when sharing data, allowing for a range of possible permissions, including those for subsequent commercial purposes. This vision envisages tiered and custom provisions to unlock data for researchers, data scientists, start-ups, entrepreneurs and MSMEs. Also in the guidelines are pointers for increasing incentives and data literacy in order to give practical impetus to both the supply- and demand-sides of data access.

To be noted is an illustrative case study being developed as a companion initiative to these guidelines. This pilot initiative in South Africa aims to bring together relevant stakeholders to contribute data sets in African languages, enabling the development of an AI Large Language Model (LLM). This outcome, in turn, will enable actors to innovate applications and services on the back of core digital public infrastructure. Details of the pilot are given in Appendix 3.

Data sharing is not an end in itself, but needs to be assessed against the costs and benefits for individuals, organisations and communities[3], for which these guidelines use the frameworks of public value and public interest. The foundational issues surrounding data access are discussed in Appendix 1.

It is against this backdrop that these principle-based possible actions, drawing upon international experiences and normative standards (see Appendix 2), offer implementable steps for policymakers and other stakeholders to fast-track changes in data governance, thereby effecting greater data access as a precondition for data use in value creation.

II. Working definitions[4]

- Data designates signals and records in structured or unstructured formats, including text, images, sound and video. AI model parameters, weights and algorithms may all be considered as data. Much "raw" data can be processed to produce meaningful results, including becoming an information resource. Information itself may be treated as data for further knowledge conversion operations.
- Data access refers to the principle and practical availability of data for actors to retrieve and/or process it, usually subject to various conditionalities. It does not necessarily imply the transmission of data sets, since options include processing data on site with only the outcome finding then being downloaded.
- Data sharing means the provision of data access by a data subject or a data holder to a data processor (the data user) for the purpose of the joint or individual use of such data, based on voluntary agreements or by law. Sharing may be done directly or through an intermediary and may take place under diverse licence conditions.
- Data holders are entities or individuals who, according to applicable laws or regulations, have the authority to allow others access to this data, with accountability for data processing operations then normally passing to the recipients of such access.
- Data brokers are entities that function as commercial third parties transacting between data holders and data processors. (Data intermediaries are normally non-profit third parties).

[3] For community benefit, see RIA's research on African data trusts

[4] These draw from, inter alia, definitions in the OECD Compendium compiled for the 2024 G20, and those from the European Data Governance Act

- Data literacy is the ability of stakeholders to recognise and act on the opportunities and risks at stake in data sharing, based on their knowledge and skills as well as on their understanding of applicable legal, ethical and institutional parameters.
- Public interest is a criterion that designates shared benefits to society as a whole (for example, public services and infrastructure) rather than advancing only individual, group or private interests. The concept implies that such benefits should be promoted and protected by all stakeholders, and especially by public authorities. Deciding what amounts to public interest entails weighing up competing assessments of potential impact and considering trade-offs over time. The power to do such adjudication is often contested, meaning that legitimate systems are needed for decisions on data access.
- Public value refers to the achievement of public interest purposes, such as sustainability or inclusivity, and can emerge out of alignments between public and private sectors and civil society. The concept is broader than, but not necessarily incompatible with, private value creation which is about the growth or profits made by businesses.[5] Private value creation may support public value where it contributes to public interest purposes.

III. Principles for the guidelines

- International human rights standards form the basis of value for all data governance.
- Transparency and accountability are required from duty-bearers to rights-holders.
- Fairness, the right to equality and non-discrimination are applicable to whom gets data access, and to the data to which access is given.
- Individuals' rights such as to justifications for limitations of any of their rights, for explainability of automated decisions, and for redress apply to data access.
- Environmental rights also implicate the technology used in data access and processing.
- The rights of communities, particularly indigenous groups, to control their own data, ensure that data governance respects cultural heritage, traditional knowledge, and collective ownership.
- Since all stakeholders, including marginalised communities, should have a voice in shaping data governance policies, participatory decision-making in policy and practice around data access, sharing and processing is to be valued.
- Safeguarding of individuals and communities from potential harms associated with data breaches, surveillance and misuse, emphasising cybersecurity, data minimisation, and ethical AI practices.

IV. The guidelines

The 30 points below are presented in the context of other outputs of the G20 Digital Economy Working Group which highlight essential co-conditions such as: digital infrastructure, including public infrastructure, access to compute, and data science

[5] See Mariana Mazzucato. The Value of Everything: making and taking in the global economy

expertise. Effective policy provisions that advance these co-conditions will optimise the points below in regard to transparency that entails meaningful access to data[6]. The possibilities for action that follow are categorised in terms of how they relate to policy, institutions, the public and private sectors, data quality and data literacy.

A. <u>Producing pro-access policy</u>

1. Optimise legal, policy and regulatory systems (including self- and co-regulatory systems) that govern data issues to ensure a balance whereby human rights-based access to and sharing of data are facilitated.
    - Ensure legal compliance and respect for privacy, cybersecurity, intellectual property as well as other data protections, and make provision for public-interest overrides to be considered through independent process.
    - Require that these considerations apply to data holdings in both public and private sectors.
    - Elaborate provisions for access by academia, civil society, start-ups, and MSMEs, with both voluntary and required data access arrangements being differentially accommodated under governance and licensing frameworks.
2. Create a clear and predictable legal regime for data access, including a range of licensing frameworks that cover purpose and use specifications, incorporating both limits and allowances for any further sharing[7], and which also include provisions for legal disclosure of data partnerships, along with penalties in the event of data breaches or purpose violations. Consider further:
    - Giving civil law effect to Creative Commons licenses and other public open licences. .[8] These extend from licences for the full public domain; and combinations of attribution required; sharealike continuity; no derivative work; non-commercial; and commercial. The licences can apply to individual data items as well as entire databases.
3. Provide tiered access frameworks in terms of degrees of data openness, as applicable to different actors, and ensure that these contribute to equitable opportunities for marginalised communities and MSMEs.
4. Integrate environmental and social sustainability provisions into strategies and programmes for data access, including energy-efficient data infrastructures, thereby aligning with the Sustainable Development Goals.
5. Commission and support research[9] into the barriers to availing and accessing data, and into mitigations thereof, and
    - Require impact assessments for major data-sharing initiatives with clear metrics and accountable reporting mechanisms.

---

[6] A repository of meaningful transparency initiatives relevant to this is available here.

[7] This is elaborated in this resource.

[8] The OECD also points to Microsoft's three data sharing agreement templates; the Linux Foundation's Community Data Licence Agreements, and the Open Knowledge Foundations. Other cases entail equitable open licences such as the NOODL License and third-party legal arrangements for stewardship models including options such as Data Trust Agreements and Data co-operative accords.

[9] An example is the report by RIA regarding Africa, which assessed challenges of company policies, legalities, technical hurdles, commercial concerns, and how data-holding entities calculated harms and mitigations

- Create multi-stakeholder advisory councils to support the assessment of data access governance, including issues such as policy impact, compliance, enforcement, and revision.

6. Adopt a critical and judicious approach through recognising that data sets are not neutral but generated for particular purposes, which may not be suitable for different purposes, and recognise that:
    - Intensified data use and data access are not a panacea[10] for all challenges,
    - There are substantive debates about data colonialism[11] and data justice[12].
    - There are a range of meanings of "open" in relation to "open data", and avoid assuming that different actors mean the same thing by the term.
    - Without policy interventions, the promotion of data commons initiatives will be of greatest benefit[13] to those already in a position to use such assets[14], reducing the wider significance and utility of such access, and that this calls out for specific mitigations.
    - The jurisdictional[15] authority and compliance issues raised by trans-border dimensions[16] in data sharing, while also promoting inter-operational data access approaches across G20 members and more broadly.

B. Innovating institutions to be fit for purpose

7. Promote institutionalised mapping of data holdings and data set inventories through encouraging or requiring the publishing of dataset descriptions and codebooks in both public and private sectors. Further:
    - Commission identification of data silos, with a view to identifying categories and cases where further public value can be unlocked.
    - Convene policy discussions to prioritise high-impact domains (for example, giving special attention to local language broadcasting, including by public broadcasters; commuting with real-time transport data; and environmental monitoring) as well as thresholds which do not impose inappropriate burdens on MSMEs.
8. Assess markets that commercially trade in data, and assess how these might be governed in the wider public interest of expanded data access.
9. Ensure relevant regulators enjoy constitutional independence from possible governmental and commercial interference. In addition, ensure:

[10] As argued here.
[11] As presented here.
[12] The Data Justice Lab unpacks this focus.
[13] Insight can be found here.
[14] See blog post by Jeni Tennison: Individual, collective and community interests in data
[15] An example here is a case before the South African Information Regulator concerning jurisdiction over international social media platforms.
[16] Such as evident in the EU's single-market approach and the aspiration in the African Union's Data Policy Framework.

- Independent adjudication mechanisms and procedures to validate data requests based on public interest criteria and for cases of compelled disclosure of defined data sets.
- For contested decisions on data access, there are independent mechanisms for appeal and dispute resolution[17].

C. Driving public sector data sharing

10. Foster transparent uptake of data pooling within the public sector for both internal operations. Consider actions to:
    - Specify data sharing provisions in regard to relevant procurement contracts with the private sector,
    - Facilitate partnerships with actors outside of government.
11. Foster data pools with non-state partners that are topic-specific and integrated, such for the purposes of transit planning, by assembling data from telecoms, banks and fuel providers.
12. Ensure open data by default in public entities and for restricted data provide systems to consider requests for access and proposals for data partnerships.
    - As appropriate, structure these systems via tiered authorisation (e.g., open to all public; restricted to regulators - such as for automated audit purposes[18]; or law enforcement with due oversight; and, data sets permitted for vetted researchers.
13. Engage different levels of government, such as municipalities, and establish linked-up public data repositories and trusts that can aggregate datasets from multiple agencies such as in health, education, transportation, and other public services[19].

D. Guidance for governing access to private sector data

14. Assess the imperative of expanding the requirements of private sector companies to proactively disclose core data where appropriate (eg. as specified in financial and trading regulations, and in health and environmental regulation).
15. Require that private companies operate free data-sharing agreements with public, private and civil society actors, in regard to data that can provide insights for immediate crises such as pandemics or environmental disasters, law enforcement and prevention of incitement to violence and monitoring systemic risks to rights[20] as related to the business operations of the data holder. In parallel, encourage voluntary efforts for data sharing that contribute to knowledge building, public health, children's rights, safety issues and environmental monitoring.

[17] This point is highlighted by the International Panel on the Information Environment.
[18] This point is highlighted in this reference
[19] Contextual analysis is important in considering the viability of data trusts, as shown in this RIA report.
[20] For example, the Digital Services Act repository of data for monitoring systemic risks to rights by Very Large Online Platforms

16. Create frameworks for compensated access for non-emergency public interest uses. Develop systems for tiered compensation of data holders (for example, distinguishing between non-profit research and commercial use of public-private datasets).
17. Advance codes of conduct where companies agree to allow research access via independent third-party intermediaries and also commit to not take adversarial action against public-interest good faith researchers.[21]
18. Implement frameworks that promote the use of privacy-enhancing technologies (PETs), such as differential privacy, federated learning, and secure multi-party computation, to enable responsible data sharing while minimising risks of re-identification and misuse of sensitive information.

E. Addressing technical issues

19. To facilitate seamless data access and integration, inter alia for the purposes of public interest research, crisis response, and regulatory oversight, consider steps to:
    - Promote activities dealing with the development of standards and protocols that facilitate data portability across various data systems.
    - Ensure that state data holdings are optimised for interoperable formats and standardised APIs.
    - Establish regulations that require private sector data to be shared in standardised, machine-readable formats with clear metadata to ensure interoperability across sectors.
20. For effective cybersecurity frameworks[22]
    - Ensure measures to protect data integrity, as well as to prevent unauthorised access, and build public trust in digital systems.
    - Join international mechanisms for threat detection, response and mitigation of potential risks.
21. Promote technical options for data sharing, such as "Trusted Research Environments", and popularise different technical architectures such as well-regulated and managed sandboxes and experimental data repositories.

F. Promoting data quality

22. Encourage assessment of data in terms of potential bias and exclusions, which may affect the utility of access and sharing.
23. Ensure specific disclosure rules for industry sectors and related regulators that specify when or what data should be excluded from the data sets they keep and share.
24. Ensure there is appropriate liability for problems when such arise out of data that does not meet quality criteria of completeness, consistency and reliability, and require public datasets to adhere to FAIR principles (Findable, Accessible, Interoperable, Reusable).

[21] As per the EU's 2022 revised Code of Practice on disinformation.

[22] This is stressed by, amongst others, the European Banking Federation.

25. Incentivise activities related to dataset cleaning, pre-training and standardisation, in order to address undetected errors and exclusions[23].

G. Enhancing incentives for data access

26. Consider tax breaks or procurement advantages for companies contributing to public-interest data pools. Consider also cross-subsidisation such as revenue from high-value dataset licensing (e.g., geospatial data) to fund municipal data trusts. Further, encourage pilot partnerships between willing actors such as in data access for specific AI applications of public interest.
27. Establish rules for data-co-operatives, and consider “Data Altruism" or related certification for companies sharing data for public use.
28. Provide civil servants with awards for excellence in data access issues.

H. Building data literacy

29. Through a range of interventions, promote awareness and skills around data literacy amongst MSMEs, including digital start-ups, and amongst both researchers and data scientists, as well as regulators. These measures include financial incentives and training support, as well as reduced-cost access to data sets, computing power and technical expertise.
30. Invest in education and training programs to enhance data literacy among citizens and civil society, so that they have improved agency to access and utilise data effectively.

[23] An example in the case of Facebook is dissected here.

Appendix 1: Foundational issues in access to data

The G20's Digital Economy Working Group (DEWG) affirmed in its 2024 ministerial decision that data is a major input for economic development, which can ensure accountability and transparency. The OECD Compendium, introduced by the G20 Brazilian presidency in 2024, makes the case for data access and data sharing across public institutions and with the private sector for public interest. The attention to data sharing also aligns with the G20's interests in Digital Public Infrastructure as articulated by the DEWG in 2023. It builds on the 2015 G20 agreement on open data principles in the context of combatting corruption.

Benefits of data access:

Extensive research, including cases covered in the OECD Compendium, shows the value of data access as a source of public value-creation. Data sharing has helped ensure improvements in public interest goods such as healthcare services, social protection and education outcomes. There are examples of pooled data assets with tiered access arrangements operating on request-based protocols. One study reveals the potential for informed policy when industries share data with researchers, as well as for fostering an ecosystem of IT developers. The practice of Open Science is another case in point. Additional benefits of data access can be seen in promoting even playing fields, such as in supporting competition law enforcement, as well as countering corruption and advancing wider transparency (eg. data showing beneficial ownership, data on public tender awards).

Key users of access to data include researchers, regulators and MSMEs. Policy attention is needed to strengthen these users who operate on the "demand side" of the data access equation, optimising benefits that have significance for the wider public interest. However, these particular Guidelines focus on opening up the data "supply" side as an essential component within wider data governance frameworks. Data sharing, whether voluntary or mandatory, requires strong governance frameworks and clear accountability, given the multifaceted mixture of risks and opportunities entailed. Enhancing the availability of data can often require reforms in policy, strategies and practice across both public and private sectors, including in licensing frameworks for data sharing arrangements. To this end, these Guidelines offer practical steps for policymakers in G20 and other countries which recognise the value of heightened access to data.

Inhibitors for data sharing:

A prelude to advancing access is identifying and mitigating the barriers for a range of data holders to open up, and likewise in enhancing appropriate incentives for them to do so . This applies to barriers such as the following:

- Access to data is sometimes seen as incompatible with the protection of personal privacy, but where such tensions exist, balances can be struck that cater for both interests.
- Intellectual property issues around data openness have intensified over the past year, with social media companies, publishers and academics, amongst others,

experiencing AI companies' using their data assets without consent. Resolutions are needed to move forward with clarity.

- An ongoing obstacle for data access in many states is their lack of capacity in terms of policy, infrastructure and practices for collecting, using and sharing data within their internal operations, and additionally in terms of availing data access to external actors. This limits the potential for such states to leverage data for public value creation.
- Cultural mindsets that favour secrecy, and hence exclusive possession and control of data, also exist across state organs and state-owned enterprises. In the private sector, many business models are intensely data-centric but operate on similar assumptions that favour opacity and limit the identification of common interests in greater openness.
- Some observers perceive a "data winter" emerging, at a time when even the holders of massive data sets are recognising that synthetic data is not a substitute for relationships to unlock and use high-quality data generated by other actors. Meanwhile, an analysis of a tracker shows that progress in the nexus between platform transparency, data access and research could do with renewed attention.

Language and data access:

Data access is deeply implicated in Artificial Intelligence (AI) foundation models and applications. Key constraints on the local development of various forms of AI have been identified as the availability of data, computing and advanced data skills. Among these, the primary problem for many data scientists, especially on the African continent, is the lack of access to the local language data necessary to build and deploy small or large LLMs in indigenous languages. In a context where data is already lacking by virtue of many forms of Indigenous knowledge not being extensively written, recorded or digitised, the impact is severe for existing inequalities in AI-powered services in smaller languages. Conversely, access to that digital-ready linguistic data which does exist, could potentially boost local economies and empower African researchers, tech start-ups and developers.

To date, much licensing of data use has been a matter of private ordering whereby researchers, individuals and data holders (whether private or state-owned enterprises) take steps in their private capacities to deny or allow access. However, particularly for much of the data held by public and quasi-public agencies (such as state-owned enterprises), parameters can be put in place to promote access through effective licensing provisions where legitimate interests prevent default disclosure by these bodies.

Appendix 2: Norms and legal standards

Relevant to these Guidelines are many normative encouragements for data sharing, such as from the UN Secretary-General and the Global Digital Compact. The Governing AI for Humanity Report released by the UN Tech Envoy High-Level Advisory Board on Artificial Intelligence entails further normative push in this direction. UNESCO's 2002 guidelines on the right to information in national frameworks cover public sector information and data access, as does the organisation's 2023 publication of Data sharing to foster information as a public good. The Open Government Partnership brings together 75 countries with interests in open systems of governance.

The US-EU Trade and Technology Council in 2023 adopted shared principles on access to data from online platforms for researchers. The OECD in 2021 agreed on a Recommendation on Enhancing Access to and Sharing of Data, and in 2025, it published Enhancing Access to and Sharing of Data in the Age of Artificial Intelligence. Under the auspices of 55 countries, the Observatory on Information and Democracy seeks to advance access to data from digital intermediaries. In their Policy Guidelines on Data Justice, the Global Partnership on Artificial Intelligence highlights the importance of equitable access to data to redress the uneven distribution of opportunities associated with data-driven technologies. Another normative document is a draft code of conduct on researcher access to data, which includes a model data-sharing agreement that would ensure privacy compliance and the need for an independent mediatory body to adjudicate data access, as authorised by the EU's General Data Protection Regulation (GDPR). That code was adopted by platforms after extensive negotiations with the European Digital Media Observatory. Also in the ecosystem is the Digital Public Goods Alliance unites national and international organisations in facilitating the discovery and deployment of open-source technologies.

These thrusts, to greater or lesser extents, locate data governance within a human rights framework with particular relevance to the right of access to information (and by extension, to data), as well as the right to the protection of personal privacy. The rights to intellectual property, to culture and language, and to non-discrimination are also implicated, as are consumer rights. The norms around data sharing can be further framed through second-generation rights, which encompass environmental and social rights, as well as through the Sustainable Development Goals.

A number of legal elaborations exist about access to data held by both the public and the private sector. These include: the EU's Data Governance Act and Data Strategy, as well as developments under the GDPR, the Artificial Intelligence Act and the Data Act. Instruments that recognise that private entities have access obligations under certain circumstances include: the Aarhus Convention; the 2022 Organisation for Economic Co-operation and Development Declaration on Government Access to Personal Data Held by Private Sector Entities; the Escazú agreement covering Latin America and the Caribbean; and the Model Law on Access to Information for Africa.

Also highly pertinent to these Guidelines are African instruments in the form of the African Union Data Policy Framework and the African Union Continental AI Strategy. Additionally

significant is the recent resolution on the importance of data access by the African Commission on Human Peoples' Rights, developed with the support of the African Alliance for Access to Data, which includes stakeholders such as the African Open Science Platform and the African Network of Information Commissioners. Also relevant is UNESCO's Accra statement on the topic, and the Broadband Commission's Working Group on Data Governance.

Pertinent to data access are legal regimes in numerous countries about the right to information, many of which align to Article 19 of the International Covenant on Civil and Political Rights with emphasis on the word "receive" in the right of freedom of expression: "this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers…". Freedom of information laws may already provide for a right to access (raw) data, while others may invite revision and elaboration to explicitly encompass data as well as rights vis-a-vis private sector data in public interest.

Appendix 3: Pilot on access to data in South Africa

This pilot will function as part of the evolution and practical application of the guidelines. It aims to enhance access to datasets in African languages that can enable the development of a local AI LLM. In turn, this outcome will enable actors to innovate applications and services on the back of a digital public infrastructure. Currently, the legal, regulatory and contractual frameworks (copyright, data protection, contract, technological protection measures, etc.) that govern data in South Africa are complex, creating bottlenecks and uncertainty that inhibit data access. For instance, there is significant uncertainty in the copyright limitations and exceptions such as fair dealing (or fair use, should the Copyright Amendment Bill in South Africa be passed) and the extent to which it enables (or will enable) access to data for AI development purposes. Regulation can affect, enable or restrict data access for research and even commercial purposes.

The pilot will assess the legal, regulatory and governance enablers and constraints in the South African context concerning access to linguistic data sources for AI development in the public interest. It will address siloed conceptions of public value creation.

The project will engage with issues such as broader commitments to data commons and data lakes in the Global Digital Compact, the Digital Public Goods Alliance (UN SG Office), the WSIS+20 review, AUDPF, and South Africa's national Digital Public Infrastructure (DPI) framework. Finally, the pilot will incorporate rights and data governance compliance with the SA Information Regulator and propose any governance amendments to enable data access in ways that ensure that it can occur with anonymised, rights-preserving and public interest purposes.

The work will focus on assessing the structure that governs the empirical reality of negotiating access and use of public data in South Africa, and then recommending data models that can be used to optimise public benefit as well as data literacy and awareness initiatives that can increase data access and expand the appreciation of the value of public data and trust between citizens and public institutions.

To conduct this assessment and make recommendations as to data models that governments can use to optimise public benefit and increase data access for data-driven value creation, the research will be founded on the right of access to information granted under the Constitution of South Africa – section 32. This human right equips data practitioners with a legal basis to request and receive access to and use data held by the State and other persons. Beyond the legal language in the Constitution, there is the institutional structure that governs the empirical reality of negotiating data access.

It is through, first an understanding, and then an assessment of this context, that it becomes possible to indicate to what extent (if at all) public value (most significant benefit to the public) is optimised; what data sharing models are in place; what capacity exists to efficiently and securely manage data; and how such capacities are being deployed. Based on the result of this assessment and the comparative analysis of the possibilities in the jurisdiction under focus, it is then possible to recommend appropriate data models; measures to strengthen capacity in public institutions to efficiently and securely manage data and; data literacy and

awareness initiatives that can increase data access and expand appreciation of the value of public data.

In the case of South Africa, the modalities for enjoying the constitutional right of access to information are codified in its Promotion of Access to Information Act. In the area of copyright law in these countries, exceptions are placed on otherwise copyright-protectable materials in a bid to promote the right of access to information. Viewed from the perspective of data governance in the technology era, this right of access to information offers a human rights solution to a policy problem: how to open and share data resources between governments and the private sector while avoiding the misuse of data. The right of access to information has been an important lever to allow data practitioners to seek data access towards data-driven value creation and the overall emergence of a vibrant data economy.

Working with and/or focusing on selected public bodies in the custody of large amounts of data, the pilot will undertake a data access request simulation exercise in which the implementers will contact and request access, pursuant to the relevant access to information statutes. The data access request will be based on a number of scenarios created in consultation with data practitioners and MSMEs involved in various projects that require access to and use of public data.

The implementers will analyse these frameworks and the results of the simulation exercise examining features such as (i) who can request access and from who; (ii) what rights and powers have they been given; (iii) mechanisms for requesting access including whether they are manual, automated, static/fixed or otherwise; (iv) the existence of a clear, practicable procedure for requesting access; (vi) mechanisms for challenging full and/or partial access refusals; (vii) remedies for access refusals; (viii) what data models are represented by or resulting from existing mechanisms for access requests and receipts; etc. The intention is to draw lessons and make recommendations on appropriate data models, propose initiatives to strengthen the existing capacity to efficiently and securely manage data and initiatives for data literacy and awareness.

This output linked to the DEWG will provide experience for G20 member states who wish to align with a key contemporary opportunity in the language space, and it can thereby contribute to wider worldwide momentum to further unlock data as a public resource available to all.