Data Governance Working Group Report

December 2023 - New Delhi Summit

GPAI / THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

This report was developed by the Global Partnership on Artificial Intelligence's Working Group on Data Governance. Its contents reflect the opinions of the GPAI Experts involved and do not necessarily represent the official views of GPAI Members. GPAI is a separate entity from the OECD and accordingly, the opinions expressed and arguments employed therein do not reflect the views of the OECD or its Members.

The GPAI Working Group Data Governance agreed to declassify this report and make it publicly available.

Co-Chair's Welcome

Jeni Tennison
Executive Director
Connected by Data

Maja Bogataj Jančič
Founder and Head of the
Intellectual Property Institute

One of the key difficulties in the development of artificial intelligence is ensuring that data is collected, used, and shared in a way that is both responsible and trustworthy. Our daily actions are directly impacted by the increasing availability, collection and use of data. Good data governance has been demonstrated by our Working Group as one of the key aspects to establishing AI systems that are both inclusive and innovative, whilst bringing economic growth, and societal benefits for the whole of society.

Our Expert Working Group mandate aims to "**collate evidence, shape research, undertake applied AI projects and provide expertise on data governance, to promote data for AI being collected, used, shared, archived and deleted in ways that are consistent with human rights, inclusion, diversity, innovation, economic growth, and societal benefit, while seeking to address the UN Sustainable Development Goals."**

Since we started our mandate three years ago we have achieved considerable advancement in the field of data governance. We are deeply proud of the progress we have made with our Working Group – people who are as passionate about the responsible use and governance of data as we are!

This year we launched two new projects, and completed a technical demonstration:

- ➔ Our first new project on 'From co-generated data to generative AI - implementing new models of sharing and governance in digital ecosystems' (referred to as the CO-GEN Project) examines how existing concepts of data sharing in co-generation scenarios have been developed so far, and how these initiatives meet the needs of an AI-specific context.
- ➔ Our second new project is about the role of the Government as a provider of data for AI which seeks to support how governments make decisions on whether and how they share the data they steward with AI developers.
- ➔ Our continuing project on Privacy-Enhancing and Adjacent Technologies completed a technical demonstration on the better health use case with Singapore. Building on the findings made this year, our next immediate step is to deploy an awareness campaign to disseminate the results from the demonstration project.

Furthermore, building on the findings of last year's report on Data Trusts in Climate, the Working Group supported a field study in the Lake Chad Basin in Cameroon to scan the local data ecosystem about climate-induced migration. On the basis of the combined set of work on Data Trusts by the Working Group, we have identified Data Institutions as one of the ways to help minimise the challenges of good data governance. From there, the Working Group supported the development of a practical tool - Trustworthy Data Institutional Framework - to develop safe, fair, and equitable sharing of data while empowering individuals and communities to assert their data rights.

Finally, building on our previous work in the field of Data Justice, we developed an awareness campaign to draw attention to fairness in the way that groups are made visible,

represented, and empowered as beneficiaries in the collection and use of data for the development of AI systems.

Our ambition as a Working Group does not stop here. Whilst we are excited to pursue our work around these challenges in 2024, our Working Group identified new initiatives emerging from the GPAI Innovation Workshop held in Montreal last September where Experts from our group co-designed with GPAI State Members new projects that are presented for future adoption in our next year's Work Plan.

In closing, we would like to thank all the Working Group for their dedication, commitment, creativity and hard work. It was a real privilege Co-Chairing this group of brilliant minds who have been a joy to collaborate with and we're eager to see the great work that the Working Group will deliver in 2024.

Introducing the Data Governance Working Group

The Data Governance Working Group (DG for short) brings together 28 experts, including two observers, from 22 countries with experience in technical, legal and institutional aspects of data governance. All our experts have shown tremendous collaboration, creativity, and commitment and as we advanced our work over the past year.

The DG Working Group's overarching objective is that data used for AI-based systems is collected and managed responsibly, in particular considering the situation of parties that are in some way or another associated with the origin and context of the data or that may otherwise be affected by use of the data such as data subjects and communities which data is collected.

It is worth noting that the DG Working Group applies a horizontal lens to its work and projects. This reflects the foundational nature of data governance and suits the group's expertise, alongside maintaining the flexibility and broader use of its work. Our Working Group collaborates with all other Working Groups through the Multistakeholder Expert Group - which we familiarly call the MEG - to advise on the data governance aspects of GPAI's wider project portfolio. Our Working Group experts have volunteered to provide expertise and guidance on all 2023 MEG activities such as the Townhall Meeting, the Experts Working Group convenings and the first Innovation Workshop held in Montreal last September. The Working Group is also eager to pursue its collaboration with the MEG through next year's transversal project on *Safety and Assurance of Generative AI* (SAFE). DG is looking forward to pursuing its work with the Responsible AI Working Group by providing expertise on data governance through the joint collaboration for next year's new project *Repositories of Public Algorithms*.

Approximately 40% of DG Experts are women, a number which we'll work to increase in the future. Most members (61%) come from the science sector, 28% are from the civil society and 11% are from the industry. A better balance should be achieved in the coming months and years as we believe that the collaboration of *all* stakeholders will be necessary to ensure AI governance.

These Experts have either been designated by the members of GPAI or through GPAI's self-nomination process. It's worth mentioning that each Expert acts with full independence inside the Working Group.

Finally, two additional observers take part in DG activities. One of them is a representative of the OECD, and the other one is a representative of UNESCO.

The next page presents the Working Group's Experts and Observers.

DG's Members

Experts of GPAI's Data Governance Working Group

Jeni Tennison (Co-Chair) – Connected by Data (United Kingdom)
Maja Bogataj Jančič (Co-Chair) – Intellectual Property Institute (Slovenia)

Aleksandra Przegalińska – Kozminski University (Poland)
Alžběta Krausová – Institute for State and Law (Czech Republic)
Andrea A. Jacobs – Code Caribbean (Antigua and Barbuda)
Asunción Gómez – Technical University of Madrid (Spain)
Bertrand Monthubert – Ekitia (France)
Ching-Yi Liu – National Taïwan University (Taïwan)
Christiane Wendehorst – European Law Institute / University of Vienna (Austria / EU)
Dani Chorin – Israeli Government (Israel)
Emmanuel Vincent - INRIA (France)
Jae Moon - Yonsei University (Korea)
Jhalak Mrignayani Kakkar – Centre for Communication Governance (India)
Josef Drexl – Max Planck Institute (Germany)
Kim McGrail – University of British Columbia (Canada)
Marc Rotenberg – Centre for AI and Digital Policy (United States)
Massamba Badiane - Ministry of Digital Economy and Telecommunications (Senegal)
Mikael Jensen – D-Seal (Denmark)
Paul Dalby – Australian Institute of Machine Learning (Australia)
Radim Polčák – Masaryk University (Czech Republic)
Ricardo Baeza-Yates – Universitat Pompeu Fabra & Northeastern University (Spain)
Robert Kroplewski – Minister for Digitalisation of the Information Society (Poland)
Sarah Shoker – OpenAI (United States)
Shameek Kundu – TruEra (Singapore)
Teki Akuetteh Falconer – Africa Digital Rights Hub (Ghana)
Ulises Cortés – Barcelona Supercomputing Center/ Universitat Politècnica de Catalunya (Spain)
Yeong Zee Kin – Infocomm Media Development Authority (Singapore)
Zümrüt Müftüoğlu – Yildiz Technical University (Turkey)

Observers

Christian Reimsbach-Kounatze – OECD
Jaco Du Toit – UNESCO

Working Group Timeline

JANUARY

First Working Group meeting (11th) – kick-off meeting to welcome new self-nominated experts and member-nominated experts. This meeting also presented the objectives for the three approved projects including a first brainstorming session to start off the two new projects coming ahead for the year.

FEBRUARY

Second meeting of the Working Group (8th) – project updates followed by two Experts presentations on their work to share expertise and knowledge on responsible AI use, development and governance.

APRIL

Third meeting of the Working Group (19th) – project updates followed by a plenary discussion to identify the potential risks and benefits of Generative AI including defining the ideal approach for the governance of these systems.

MAY

Fourth meeting of the Working Group (10th) – plenary discussion to identify and define what key messages the DG Working Group will bring to the attention of GPAI State Members to responsibly assess the development of Generative AI. This meeting allowed the Experts of the group to address the questions raised by the State Members in preparation for the Townhall meeting held the following week May 15th.

➔ GPAI Townhall meeting on Generative AI held live May 15th 2023

JUNE

Fifth and sixth meeting of the Working Group (1st & 28 th) – project updates and presentation of the proposals to be included in the work plan 2024.

AUGUST

Seventh meeting of the Working Group (30th) – project updates and presentation to prepare the Innovation Workshop in-person in Montreal end of September including planning on the next steps coming for the Summit.

OCTOBER

Eighth meeting of the Working Group (12th) – two workshops, one about the technical demonstration for the Privacy-Enhancing and Adjacent Technologies project followed by a second workshop on the project on Government as a provider of Data for AI.

NOVEMBER

Ninth meeting of the Working Group prior to the Summit – question and answer session in preparation for the GPAI Summit 2023.

DECEMBER

Presentation of finalized outputs including projects included as part of the Work Plan 2024 at New Delhi Summit (12th - 14th)

Progress Report 2023

Overview of the Data Governance Expert Working Group Projects

From co-generated data to Generative AI –
New Rights and Governance Models in Digital Ecosystems (CO-GEN)

This new project examines how existing concepts of co-generation have been developed so far and how these initiatives address an AI-specific context. For this year, the project draws a first analysis of the world of co-generated data, content, and AI systems by identifying whether the current variety of existing rights and legal protections take into account the specific requirements of AI as a technology.

The role of Government as a Provider of Data for AI

This new project proposes to support governments to make decisions about whether and how to share data they steward with AI developers. The intended impact is to increase the availability of publicly held data for AI grounded in the principles of human rights, inclusion, diversity, innovation and economic growth by helping governments to prioritize their efforts and to reduce their concerns about the risks of sharing public data for AI by providing clear guidance, use cases and examples that demonstrate how it can be done safely and responsibly.

Enabling data sharing for social benefit through data institutions

This project was established to support the creation of real-world data institutions that enable data sharing for social benefit. It supports a set of recommendations and tools that empower individuals and communities to enact their data rights, ensuring that data-sharing activities reflect the diverse interests of all in society. The end goal is to help GPAI realise the potential of data trusts as a tool to promote the safe, fair, legal and equitable sharing of data, in service of the UN Sustainable Development Goals.

Privacy-enhancing and adjacent technologies
This project aims at demonstrating the viability of AI systems in helping achieve the UN Sustainable Development Goals such as global health, climate action, and the future of work in harmony and with flourishing of human dignity, by providing a means to safely develop, use and share data while preserving privacy, sovereignty, personal integrity, IP rights, and security. The project also seeks to overcome challenges to data usability commonly faced when working with PETs by publishing practical guidance and lessons learnt from the recent demonstration system. This can support innovation by helping smaller organisations or corporations to compete more effectively with large (and sometimes monopolistic) data-rich organisations that have access to massive datasets within their organisational boundaries.

Advancing research and practice on data justice –
Awareness campaign 2023
This project was established to fill a gap in Data Justice research and practice that provides a frame to help policymakers, practitioners and users to move beyond understanding data governance narrowly as a compliance matter of individualised privacy or ethical design, to include considerations of equity and justice specifically as it relates to redressing the uneven distribution of opportunities, and harms, associated with AI and ML currently. The objective is to make significant progress in getting more equitable access to, greater visibility and fairer representation of those individuals and communities marginalised from data used in the development of AI/ML systems, through the adoption of more just principles into AI policy and practice. The Working Group developed a set of recommendations at the end of 2022 and building on these this year an awareness campaign was released to sensitize the world in the way that groups are made visible, represented, and empowered as beneficiaries in the collection and use of data for the development of AI systems.

From Co-generated Data to Generative AI – New Rights and Governance Models in Digital Ecosystems

Project Co-Leads

Christiane Wendehorst
Professor of Law,
University of Vienna
Scientific Director of the
European Law Institute

Kyoko Yoshinaga
Associate Professor
Graduate School of
Media & Governance,
Keio University

Context

Data created by multiple individuals or organizations as opposed to only the data holder is referred to as co-generated data. Co-generation happens throughout the data ecosystem, including when we use conversational AI technologies, interact with voice assistants, and browse social media. There may be several data co-generators in each of these scenarios, some more or less involved in the co-generation process or conscious of its possible financial, social, legal, or ethical ramifications. With the increasing accessibility of AI systems using Generative AI, we argue that we should reconsider the current legal frameworks to better address the specificity of co-generated data.

Activities of 2023

For the first phase conducted this year, we partnered with the Open Data Institute and the Aapti Institute to examine the variety of rights and legal protections available to those who have co-generated data content. Who has rights in data, on what basis, and what these rights might mean are the fundamental questions we have addressed as part of this year's research.

In 2023 we conducted a literature review that produced a legal framework for analysis, examined six co-generational scenarios using the framework, and conducted 13 interviews with international experts that offered further in-depth information. **The findings suggest that the current legal frameworks do not adequately address the unbalanced power dynamics between AI technology companies and co-generators, which can impact their ability to act on these rights.**

Key findings

- ➔ The current landscape is dominated by big tech companies that wield huge amounts of power over other co-generators;
- ➔ A harm-led approach to understand the impacts is critical for a robust analysis;
- ➔ Prioritising a multistakeholder approach in developing rights frameworks is critical;
- ➔ There is no one-size-fits-all solution, but a diverse set of potential solutions, and a need to have contextual nuance.

Next steps

In March 2024, we aim to publish our first report examining the current legal landscape of co-generated data, presenting principles on co-generation (including with regard to collective rights and potential rights in technology such as generative AI) and reporting on our key findings. From there, building upon the outcomes of the foundational work completed this year, the project will seek to explore two specific approaches for implementing the developed co-generation guiding principles:

1) Assessing what appropriate licensing schemes could look like when applied to a co-generated context, and
2) Develop an understanding of existing technical solutions which may help enable putting these principles into practice.

The Role of Government as a Provider of Data for AI

Project Co-Leads

Ching-Yi Liu
Professor of Law at National Taïwan University

Jhalak Kakkar
Executive Director at the Centre for Communication Governance at National Law University Delhi

Context

Governments have a crucial role in data production, collation, and collection. Governments hold a variety of data made available through the provision of government services such as civil registration (issuing IDs, birth and death certificates, etc.), healthcare, education, business registration, policing, government research, and national statistics, amongst many others. These data contain information that provides insights into societal concerns, including the operation of healthcare systems, providing social protection, school attendance, etc. These can serve as a crucial basis for creating AI solutions that tackle social challenges, such as gaps in government provision systems, including accessibility issues.

The provision of government data to AI-based systems developers must be undertaken responsibly, considering various foundational principles such as respect for human rights, privacy, consent, inclusivity and ethical use. This requires a number of measures that may be adapted in relation to different country contexts and needs. Such measures include mechanisms to ensure that sensitive data, including personal data, is legally and safely shared; adherence to data standards that support interoperability, ensuring data shared is in a format that is structured, discoverable, reusable and machine-readable; public engagement, participation or awareness programmes to ensure public buy-in for the provision of government data to AI developers; transparency mechanisms to enable accountability and build public trust; and undertaking an impact assessment or similar risk-mitigation measures to prevent against the risk of harm, particularly for underserved or vulnerable communities.

Activities of 2023

For the first phase conducted this year, we partnered with Research ICT Africa to conduct this research. As part of the exploratory phase, a survey of data sharing by governments for the development of AI around the world to generate examples of existing approaches, including relevant legal controls. Four detailed illustrative case studies of current practice were identified from these existing examples. From there, a close analysis of these four case studies was conducted, and the findings suggested that a further analysis of the concept of government as a provider of Data for AI needed to be strengthened to capture the actual phenomenon[1].

Key findings

1) **Developing Narratives**: The concept of government as a provider of data for AI research and development is yet to be readily understood. We need to strengthen the conceptual framing of this phenomenon and the narrative around why this is important

[1] See the 2023 report here.

for advancing the public interest in AI-based solutions that support the realisation of developmental priorities.

2) **Transparency**: There is a paucity of publicly available and accessible details and information on existing/potential illustrative case studies related to governments sharing data with AI developers. It may be because of limited public trust in governments sharing data with AI developers, so where this occurs outside of formal public procurement processes for AI developers, there is limited publicly available information. This was the case in all regions we were examining. However, this was a particular concern in Low- or Middle-Income Countries (LMICs), except for a limited number of countries that are leading through regulatory sandbox initiatives. This is due to the objectives of these sandbox initiatives in developing best practices and implementation guidelines in pilot programs that can be replicated by other countries, which facilitates information sharing.

3) **Capacity building**: Governments currently have limited capacity to adequately assess, approve, share and review data that they hold with AI developers, including institutional oversight capacity. This is particularly so in LMIC contexts.

4) **Data Sharing models**: Government Open Data portals remain a critical venue through which to fairly and transparently share government-held data for AI development, but more needs to be done to ensure these platforms are populated, updated, used and accessible. In addition, new models for data sharing to balance the power between data controllers/processors and data subjects, including data commons, are emerging around the world. Further research is needed to better understand their risks and opportunities, and ensure they are responsibly regulated.

5) **Data Subject rights**: The empowerment of data subjects in new arrangements for data sharing between governments and AI developers stands out as an exemplar. However, they are not without costs and fundamental requirements, such as awareness and understanding about data rights from citizens.

6) **Public Trust**: Ensuring public trust in AI-based systems and government adoption of AI is a critical first step in the sustainability of fair partnerships between governments and AI developers.

7) **Ethical regulation and governance**: As the use of AI by governments increases around the world, new opportunities and risks arise. In order to develop sound governance frameworks progressive regulation will be needed to establish strong oversight of privacy and respect for Human Rights.

Next steps

This project aims to build on the outcomes of the work completed in 2023 by implementing and testing its recommendations, ideally in partnership with a governmental organisation, in the areas of healthcare, energy, and/or benefits data.

Enabling data sharing for social benefit through Data Institutions

Project Co-Leads

Teki Akuetteh
Founder &
Executive Director
at the Africa Digital
Rights' Hub

Thomas Nkoudou
Researcher in
residence at CEIMIA

Context

Supporting the creation of real-world data trusts that enable data sharing for social benefit requires proper data governance tools that empower individuals and communities to enact their data rights, ensuring that data-sharing activities reflect the diverse interests of all in society.

This year's project focuses on data institutions with a wider lens than data trusts to conduct a field study to provide practical tools for trustworthy data governance. This 2023 phase builds upon important foundational work by the Working Group since 2020, including the first international definition of a data trust, a synthesis of the 'state of the art' in the design and implementation of data trusts, a review of the legal and legislative frameworks that are in place to support the operationalisation of data trusts and the exploration of real-world use cases on climate action.

Activities of 2023

Building on our foundational work on data trusts in climate action the Working Group conducted a field study in Lake Chad Basin in Cameroon[2], a region grappling with a complex humanitarian crisis with over 3.2 million people displaced, largely because the lake has shrunk by 90% (scarce water supplies, food insecurity, degraded farmlands, farmers-herders conflicts). We engaged with local organizations and consulted with affected communities to co-design a framework for trustworthy data exchanges within the climate migration data ecosystem, improving how data is being collected, stewarded, shared and used to better serve their needs and empowering them to play an active role in the data value chain.

From this field study, the Working group decided to build a practical tool - The Trustworthy Data Institutional Framework[3] - that improves the process for good data governance throughout the data lifecycle. The TDIF is structured in four levels: cycle of trust, system of trust, chain of trust and functions of trust. This living tool enables organizations to assess their standing in building trust in their data practices. It includes four key indicators with their variables and five maturity levels.

[2] The 2023 report on the Lake Chad Basin field study in Cameroon is available here.
[3] The 2023 Trustworthy Data Institutional Framework is available here.

Privacy-enhancing and ajdacent technologies

Project Co-Leads

Kim McGrail
Professor
University of
British Columbia

Shameek Kundu
Chief Strategy Officer
Truera

Context

In 2022, the Working Group kicked off a new technically-focused project seeking to explore possible applications of privacy-enhancing technologies (PETs+) in AI-for-social-good contexts, specifically across three priority topics: Better Health, Future of Work, and Climate Action. The Working Group has been supported in this work by the Infocomm Media Development Authority (IMDA), a Singapore government agency. The primary objective was to increase the availability and/or usability of AI systems by providing a means to safely develop and use data sets while preserving privacy, sovereignty, IP rights, and security.

Activities of 2023

This year, the Working Group partnered with Singapore's Infocomm Media Development Authority (IMDA) to conduct a technical demonstration of PETs+ technologies at work, using the "Better Health – Modelling the effects of human movement during a pandemic" project[4]. Singapore's Digital Trust Centre (DTC) acted as the delivery partner and scoped precise data sources and models for the demonstration. The selected model for the demonstration came from Singapore's Ministry of Health (MOH). The model simulates SARS-CoV-2 infection transmission and control in large-scale multi-day events.

Four key learnings were highlighted from this year's demonstration:

1. PET systems must be designed to balance the security and privacy concerns of data owners with the usefulness concerns of the modellers.
2. Efforts to specify and minimise the data fields involved can improve the performance of the PET-enabled pandemic model.
3. Education and early involvement of stakeholders in charge of data ownership, governance and compliance help design a more realistic PET system.
4. There are clear trade-offs between Homomorphic Encryption (HE) and Differential Privacy (DP).

Next steps

Building on the learnings from this demonstration, the Working Group aims to deploy an awareness campaign to disseminate the results from the demonstration project.

[4] The 2023 demonstration report is available here.

Forward look

For 2024, the Working Group aims to continue two of its current projects and start a new joint Working Group collaboration Responsible AI. The Working Group has proposed the following projects for 2023 subject to GPAI Council approval at the New Delhi Summit:

Data Governance Working Group 2024 Projects		
DG #1	From co-generated data to generative AI - implementing new models of sharing and governance in digital ecosystems	Continuing
DG#2	The Role of Government as a Provider of Data for AI	Continuing
RAI #1	Repositories of Public Algorithms (joint project with Data Governance)	New

We're looking forward to starting 2023 with these upcoming projects in the pipeline. We're hopeful that the next months will be productive and that our future research agenda will guide the next steps on opportunities to go further and deeper in advancing research and practice on responsible development, use and governance of AI.

Participation across our Working Group is a big part of what makes these projects true international collaborations. We would like to invite those who are interested to make a personal contribution to these projects by joining our project steering groups to help shape direction, give feedback and review research. You can express your interest to contribute by connecting with the Montreal Centre of Expertise (the CEIMIA) at info@ceimia.org