

## DATA GOVERNANCE WORKING GROUP REPORT GPAI – MONTREAL SUMMIT 2020

### EXECUTIVE SUMMARY

#### i. Introducing the Working Group

The Working Group consists of 27 experts (plus two observers) from 17 countries with expertise in technical, legal and institutional aspects of data governance. It is led by Co-Chairs Jeni Tennison of the Open Data Institute (UK), and Dr. Maja Bogataj Jančič of the Intellectual Property Institute (Slovenia).

#### ii. Mandate

Our mandate as a group aligns closely with GPAI's overall mission, with an aim 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.”*

#### iii. Work process

As the report outlines, the Working Group has placed openness, transparency, collaboration and diversity at the heart of its processes. The Working Group has organized breakouts, Workshops, consultations, jam board sessions, surveys, blogs and project steering groups in the process of pursuing its mandate.

#### iv. Preliminary recommendations and outputs for the Summit

In preparation for the Summit, the Working Group agreed to develop two headline outputs:

1. **A Framework for GPAI's work on Data Governance** - setting the stage for all future Working Group projects, serving as an overview over the most relevant terms and defining the understanding of the Working Group of data governance in the context of AI.
2. **An investigation into the Role of Data in AI** - to complement and dig into topics in the Framework in more depth, this situates the importance of data to AI development and identifies areas both where more data would be useful - such as specific, open, datasets that could be worthy of national support or international collaboration - and where harms arise due to the collection of, use of or access to data. The Working Group commissioned a consortia led by the University of Edinburgh (School of Informatics and the Digital Curation Centre) and Trilateral Research for this work - the report will be published shortly after the Summit.

The Framework provides a roadmap that outlines how the Working Group will focus on three types of approaches to data governance to support GPAI:

1. **Technical approaches** (e.g. privacy-enhancing technologies, bias detection and correction techniques)
2. **Legal approaches** (e.g. taking into account IP law, data protection law), and
3. **Organisational/institutional approaches** (e.g. data representatives or trusts, common data spaces).

The *Role of Data in AI* investigation includes recommendations to enhance international cooperation on data, providing a more project-based direction that complements the Roadmap. It references existing initiatives on each that GPAI could build upon and support and steps to do so:

1. The Data Governance WG should work to **shape best practices and standards for data governance** with the aim to drive access to good quality data for AI projects and systems. A potential area of collaboration with the Pandemic Response Working Group is highlighted on supporting good practices around deposition and cataloguing of AI data sources so they are more discoverable and accessible.
2. **GPAI could underpin the creation of good quality and accessible data sources to fill data gaps in priority fields, in line with the UN Sustainable Development Goals**, through targeted research and collaboration with initiatives in this field. The WG should explore those areas in particular where investment is unlikely to happen, and work with other WGs and GPAI to push for action and make the data available for global benefit.
3. **Undertake research into how to improve cross border data sharing** and write guidelines for organisations on how to address current barriers, such as intellectual property rights, privacy and data protection legislation, and data sovereignty. The WG could identify the main areas where more certainty around AI is required and can develop concrete guidelines on how this could be achieved at a national and an international level. It could also suggest a set of best practices in preventing fragmentation among jurisdictions that can hinder data flow and availability for AI. Linked to this, the WG should explore how to best support technological developments, such as **federated learning technologies** and **privacy-enhancing technologies** for data sharing as potential mitigation of legal challenges, especially around personal data, and support their development and uptake where possible
4. **Undertake targeted research into the broad topic of data injustice and harms** that arise from data practices around the world and identify pathways to counteract current problems. Analysis should be carried out of potential mechanisms that can overcome the challenges identified. The WG should seek out initiatives that work in this field and support them in creating concrete mechanisms to redress the harmful impacts of data in AI. A number of priority fields are suggested.

These recommendations, produced by the Edinburgh University / Trilateral Research consortium, will now inform the next phase of the Working Group's work in H1 2021.

#### **(v) Priorities for H1 2021**

The Working Group is now drawing from its two opening outputs to develop a set of *concept notes* in H1 2021 for projects and programmes of work that could advance GPAI's mission, and could be funded by GPAI's members and in partnership with others. The intention is that this will provide a way of getting to action by framing a set of challenges in terms of activity. They will draw on the recommendations made in the *Role of Data in AI*, and be structured and framed as a set of challenges in line with the *Framework*.

#### **(vi) Longer term vision**

The Working Group will be guided in its longer-term vision by the realisation of GPAI's overall mission. The concept notes that the Working Group will develop in H1 2021 will outline a set of practical outcomes that GPAI could help achieve, and specify the means to deliver them. Through collaboration and partnership, those outcomes will then become the focus of the Working Group over the next 2 years.

Recognising that Data Governance will be a foundational element of many GPAI projects and its horizontal focus, the Data Governance Working Group will also seek to collaborate with other Working Groups - either directly on our own projects, or by 'lending' our experts to other Working Groups to advise on the data governance aspects of their applied AI projects. There are many other links across the Working Groups, and we look forward to offering foundational assistance in their future projects.