

# Data Governance Working Group Report

November 2020 - GPAI Montréal Summit



**GPAI**

THE GLOBAL PARTNERSHIP  
ON ARTIFICIAL INTELLIGENCE

*Please note that this report was developed by experts of the Global Partnership on Artificial Intelligence's Working Group on the Responsible Development, Use and Governance of AI. The report reflects the personal opinions of GPAI experts and does not necessarily reflect the views of the experts' organizations, GPAI, the OECD or their respective members.*



	<b>Co-Chairs' Welcome</b>	<b>4</b>
i)	<b>Introducing the Working Group</b>	<b>5</b>
	<b>Membership of GPAI's Data Governance Working Group</b>	<b>6</b>
ii)	<b>Our Mandate</b>	<b>7</b>
iii)	<b>Work process</b>	<b>7</b>
	<b>Working Group Timeline</b>	<b>8</b>
iv)	<b>Preliminary recommendations and outputs for the Summit</b>	<b>9</b>
v)	<b>Priorities for H1 2021</b>	<b>13</b>
vi)	<b>Longer term vision</b>	<b>14</b>
	<b>Annex 1</b>	<b>15</b>
	<b>Annex 2</b>	<b>17</b>

## Co-Chairs' Welcome



**Dr. Jeni Tennison**  
Vice-President and Chief Strategy Adviser  
Open Data Institute



**Dr. Maja Bogataj Jančič**  
Founder and Head  
Intellectual Property Institute

Good data governance—collected, used and shared in responsible and trustworthy ways—is one of the fundamental challenges in the development of AI. Data governance must be consistent with human rights, inclusion, diversity, innovation, economic growth, and societal benefit, in congruence with the UN Sustainable Development Goals.

In 2020, the importance of good data governance has been both tested and demonstrated. From effective contact tracing systems to securing access to relevant chemical and drug data for AI assisted drug discovery and repurposing, 2020 has challenged those working in data governance - and all of us of course on a personal level. Among the bright spots for us though this year has been establishing this Working Group and getting to know this group of experts: an energetic, passionate and collaborative group that has brought so much energy and global perspective to this enterprise.

We are hugely grateful to the commitment they have shown. Collectively we have a strongly held, shared belief in the value and importance of GPAI's mission, and a desire to see its amazing potential realised. Data governance is likely to be a foundational element across a breadth of GPAI's projects and we are lucky to have a group of experts that are ready to support these. Our Mandate closely and deliberately reflects that of GPAI's overall mission: that we will “collate evidence, shape research, undertake applied AI projects and provide expertise on data governance, to promote data for AI<sup>1</sup> being collected, used, shared, archived and deleted in ways that are consistent with human rights, inclusion, diversity, innovation, economic growth, and societal benefit, in congruence with the UN Sustainable Development Goals.”

This report sets out our initial work in the first six months - a Framework that sets the scope and terms of the Working Group and an investigation we commissioned into the Role of Data in AI with some initial recommendations on further international collaboration - and outlines how we intend to work with other Working Groups, and focus our efforts on supporting action-based projects that will help advance our Mandate. We look forward to discussing the next chapter of GPAI's work.

---

<sup>1</sup> The Mandate draws upon the definitions set out within the [OECD Recommendation on Artificial Intelligence](#) for this purpose

## i. Introducing the Working Group

Our Working Group consists of 27 experts, including two observers, from 17 countries with experience in technical, legal and institutional aspects of data governance. True to the overall ambition of the Global Partnership on AI, they combine cross-sectoral insights from the scientific community, industry, civil society and international organizations. They bring perspectives ranging from those developing COVID-19 datasets to others working with indigenous communities to improve their representation. This has provided a rich discourse in the Working Group's deliberations; we are fortunate to have a highly energetic, passionate and collaborative group bringing a wealth of perspectives beyond any single country.

We are keen to build on that discourse and, in particular, add to the diversity of the group. We are therefore looking forward to the expansion of GPAI membership in 2021 and welcoming experts from new GPAI members.



# Membership of GPAI's Data Governance Working Group

## Working Group members

Jeni Tennison (Co-Chair) – Open Data Institute (UK)  
Maja Bogataj Jančič (Co-Chair) – Intellectual Property Institute (Slovenia)  
Alejandro Pisanty Baruch – National Autonomous University (Mexico)  
Alison Gillwald – Research ICT Africa (South Africa / UNESCO)  
Bertrand Monthubert – Occitanie Data (France)  
Carlo Casonato – University of Trento (Italy)  
Carole Piovesan – INQ Data Law (Canada)  
Christiane Wendehorst – European Law Institute / University of Vienna (EU)  
Dewey Murdick – Center for Security and Emerging Technology (USA)  
Hiroshi Mano – Data Trading Alliance (Japan)  
Iris Plöger – Federation of German Industries (Germany)  
Jeremy Achin – DataRobot (USA)  
Josef Drexl – Max Planck Institute (Germany)  
Kim McGrail – University of British Columbia (Canada)  
Matija Damjan – University of Ljubljana (Slovenia)  
Neil Lawrence – University of Cambridge (UK)  
Nicolas Miaillhe – The Future Society (France)  
Oreste Pollicino – University of Bocconi (Italy)  
Paola Villerreal – National Council for Science and Technology (Mexico)  
Paul Dalby – Australian Institute of Machine Learning (Australia)  
P. J. Narayanan – International Institute of Technology, Hyderabad (India)  
Shameek Kundu – Standard Chartered Bank (Singapore)  
Takashi Kai – Hitachi (Japan)  
Teki Akuetteh Falconer – Africa Digital Rights Hub (Ghana / UNESCO)  
Te Taka Keegan – University of Waikato (New Zealand)  
V. Kamakoti – International Institute of Technology, Madras (India)  
Yeong Zee Kin – Infocomm Media Development Authority (Singapore)

## Observers

Elettra Ronchi – OECD  
Jaco Du Toit – UNESCO



## ii) Our Mandate

Our mandate as a group aligns closely with GPAI's overall mission. Our Working Group 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.”*

Clearly, there are interactions between data governance and the remits of the other Working Groups – particularly on Responsible AI and Commercialisation and Innovation – so we are aiming to work with them on areas of overlap. The Working Group also has the chance to help coordinate GPAI's applied AI ambitions, shape projects carried out by or funded by GPAI's members and across its wider partnerships, and to influence the policy recommendations created by the OECD through its work. Our goal is that our work is also useful more widely, amongst those researching, thinking about and implementing data governance practices in AI.

## iii) Work process

In our approach to our work, we committed early on and in [public](#) that we would place the values of openness, transparency, collaboration and diversity at the heart of our process. As our timeline demonstrates below, those values have shaped our work in some very practical ways, and we will keep adding to our diversity and building on our collaboration in 2021. In particular, we look forward to new partnerships, cross-Working Group collaborations and welcoming experts from new countries as GPAI expands.



## Working Group Timeline

### JULY

- **Co-Chairs' introduced** (2nd)
- **First Working Group meeting** (24th): introductions and agreement on mandate and Summit deliverables (*the Data Governance Framework* and *the Role of Data in AI*)
- **Background materials** on data governance approaches gathered from Working Group experts.

### AUGUST

- **Project Steering Committees established** - meeting weekly from week commencing August 23rd
- **Christiane Wendehorst agreed as Project Lead** for the Data Governance Framework
- **Introductory blog on the Working Group published on the OECD including a request for proposals** and terms of reference published on *the Role of Data in AI*

### SEPTEMBER

- **Second meeting of the Working Group** (2nd) - Kim McGrail agreed as Pandemic Working Group link; breakout sessions on the Data Governance Framework (on the data lifecycle, data ecosystem/actor models, and data governance framework models)
- **Round 1 evaluation of proposals for the Role of Data in AI** by a Working Group Evaluation Panel (week commencing September 7th)
- **Round 2 evaluation of proposals and selection of the Digital Curation Centre, Edinburgh University's School of Informatics, and Trilateral Research** as consultancy partner on the Role of Data in AI (week commencing September 14th)
- **First draft shared of the Data Governance Framework** among the Working Group with survey launched (September 22nd)
- **Third meeting of the Working Group** (28th) - introduction to the Role of Data in AI project team; presentation on the first draft of the Data Governance Framework with plenary discussion on survey questions followed by breakouts on the Framework's roadmap; discussion on outreach.

### OCTOBER

- **Catalogue of Global South experts** initiated with Working Group member recommendations for invitation to open events.
- **Meeting between all Co-Chairs** to compare progress and discuss potential synergies (October 16th)
- **First Role of Data in AI workshop** with additional guest participants and speakers on 'The Role of Data in developing human language technologies for under-resourced languages' (October 19th)
- **Fourth meeting of the Working Group** - presentation of 'beta' Data Governance Framework and first draft and discussion of the Role of Data in AI, breakout sessions on availability/accessibility of data for AI; socio-economic, environmental and legal impact; and our target audience for recommendations (October 28th)



## NOVEMBER

- **Second Role of Data in AI workshop** with additional guest participants and speakers on data-driven justice systems, examining the social, legal and ethical implications (November 2nd)
- **'Beta' version of the Data Governance Framework** published on the OECD *AI Work* with a two-week window for comments from the wider community (November 3rd)
- **Fifth meeting of the Working Group** (19th) - short presentation on the 2nd draft of the Role of Data in AI; jam board session on H1 2021 plans (developing a long list of concept notes in line with the Framework's Roadmap)
- **Final Role of Data in AI workshop** on Responsible and trustworthy open and FAIR data sources for AI (November 23rd)
- **Meeting between all Co-Chairs** prior to the Summit (27th)

## DECEMBER

- **Presentation of finalized outputs and open workshop on next projects at the Summit.**

## iv) Preliminary recommendations and outputs for the Summit

In preparation for the Summit, we 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; and
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 Framework has been led by Christiane Wendehorst, Professor of Civil Law at the University of Vienna and President of the European Law Institute, with support from two research assistants Nina Thomic and Yannic Duller, and—as section iii shows—developed with the full collaboration of the wider Working Group over the course of many workshops, surveys, and drafts over the past few months.

At headline level, the Framework covers four areas:

1. **The role of data in the AI context:** including data for AI development & deployment and data lifecycle
2. **Why data governance matters:** including case studies that illustrate the necessity of good data governance, the role and responsibility of different actors, and principles for Data Governance
3. **Parameters of data governance:** including categories of data, data ecosystems, and rights with regard to data
4. **A roadmap for the Working Group's future work that outlines how the Working Group will focus on three types of approaches to data governance:** (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)



It is important to note that the Framework should be considered as a 'living' document - it will evolve as it needs to in order to keep pace with developments in the field.

For the Role of Data in AI investigation, we commissioned [a consortium](#) led by the University of Edinburgh, combining a breadth of technical and legal expertise made up of the School of Informatics, the Digital Curation Centre and Trilateral Research. It digs more deeply into the issues raised within the Framework, and identifies areas where GPAI could make an impact in deepening international collaboration. It covers the following areas:

- AI development and the role of data at each step;
- Data types used in AI development;
- Data characteristics that influence the process or outcome of AI development;
- Socio-ethical, economic and environmental impacts of data in AI;
- Law and transparency as modifiers to impacts of data in AI;
- Availability of accessibility to data for AI development; data quality and challenges in three fields (pandemic response, human language technologies for under-resourced languages, and AI applications in the criminal justice system); and recommendations on where GPAI could enhance international collaboration on data governance.

We plan to publish the *Role of Data in AI* immediately after the Summit.

Both projects have been guided by a Steering Committee made up of Working Group experts, listed under Annex 2.

#### *Preliminary recommendations*

The Framework provides an agreed scope, structure and vocabulary for the Working Group's work. We present a common understanding of data governance within the context of AI (including the agreed set of conceptual terms and frameworks we will use), and introduce a Roadmap for how GPAI should structure its Data Governance work.

The roadmap outlines how the Working Group will focus on three types of approaches to data governance:

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).

In applying this focus, the roadmap establishes a horizontal lens to the Working Group's approach. This reflects the foundational nature of data governance and suits our expertise, alongside maintaining the flexibility and broader use of our work.

The *Role of Data in AI* investigation includes five recommendations that could enhance international cooperation on data governance, providing a more project-based direction that complements the Roadmap and will inform our next steps as a Working Group. It includes references to existing initiatives on each that GPAI could build upon and support:

Recommendation 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.

Actionable steps include:

Action 1a: Create guidelines around data management for AI projects and systems, which take all steps of the AI development process into account, from data creation and collection through to preservation and deletion. The WG should also work towards creating a data management plan template for AI projects and systems, which will allow for the capturing of information necessary for supporting discoverability, documentation, characterisation, trust and transparency (see recommendations 1b-1e), all of which will drive enhanced and informed re-use of data for AI.

Action 1b: Support good practices around deposition and cataloguing of AI data sources so that they are better discoverable and accessible. This work should include a focus on:

- Conducting a feasibility study around different options for enhancing data access for AI projects and systems. Options may include e.g., setting up a specific AI data repository or a metadata catalogue, or creating a network of existing repositories and a single discovery and access point.
- Working with initiatives that are driving the adoption of the FAIR principles, as well as the Open Science movement, and ensure that AI has input on any issues that are specifically relevant to specific data practices within the field.
- Working with the Pandemic WG on implementing their recommendation for a Central Pandemic Response Portal. Lessons learned from this collaboration can then be carried forward and applied to other domains.

Action 1c: Develop guidelines for dataset documentation and metadata for AI projects and AI systems. This work should include a focus on:

- Defining a minimum information standard for source description of AI data, drawing on good practices in data documentation.
- Develop guidance on how to best incorporate data provenance and lineage in metadata to improve traceability of datasets. Review work of initiatives in this field and collaborate on defining good practices and standards for this information.
- Define how IPR and licencing issues relevant to the data are presented in the documentation.

Action 1d: Develop data characterisation documentation guidelines and suggestions for alignment for each project or system. These guidelines would include a guidance on:

- How to define a desired data use case for the project/system, i.e. what data is needed to reach the aims of the project/system to ensure that data selected is fit for use.
- How to identify data sensitivities, to include legal and regulatory issues relative to the use case and work to mitigate these.
- How to assess existing data for completeness (for re-users) and ensure the completeness of data that is created.
- How to undertake data improvements and manage data generated by the AI system.

Action 1e: Develop guidelines for data creators regarding the provision of transparency for data users around the creation and contents of the dataset, to enhance trust in these data resources and their use. This recommendation is closely related to recommendations 1c and 1d but this work will specifically focus on how to instil data users' trust in datasets they intend to use for their AI projects and systems. This work will include a focus on:

- Data representativeness and coverage. Clarify whether there are issues with representativeness and coverage in the dataset, and if relevant list the steps that have been taken to eliminate bias in the dataset.
- Data accuracy and relevance. Clarify the actions that have been undertaken to verify the accuracy of the data.
- Define the legal and ethical issues that have been identified relating to the data and how have these been resolved.
- Develop trusted mechanisms (e.g., certification badges) for displaying that datasets have undergone processes that incorporate the above checks.

Recommendation 2: 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 focus should be on underpinning the creation of accessible and good quality data sources, according to best data governance practices. Steps should be outlined to work with governments, “AI for Social Good” initiatives, and relevant stakeholders to underpin and establish reliable data sources in priority areas. 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. As part of this work, it is important that the study also identifies gaps in dataset creation from disparate sources of data for the understanding of complex problems. One example of this is the pandemic response, where there is a lack of data sets that include socio-economic data, health record data, and genomic data leading to great risks for the public health.

Recommendation 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
- Data sovereignty

In addition, 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.

Recommendation 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. We suggest priority fields to be:

- Indigenous Data Sovereignty and potential friction in relation to implementation of the FAIR principles and data openness.
- Bias in data and its impacts on society and individual rights. How to ensure inclusivity in AI data so that benefits can be more broadly realised and harms avoided.
- Environmental harms arising from data processing and storage, and how to mitigate these.



Strengthening data capabilities in the Global South through international collaborations and networks specifically working to build soft and hard infrastructure in the region.

We thank the Digital Curation Centre, School of Informatics and Bayes Centre for these recommendations, and the contribution they will now make in informing 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.

The concept notes will include:

A description of the problem and background on what's known so far

- The intended impact of the project long term
- The more specific outcomes of the project
- The activities involved
- Their outputs
- Potentially, a description of the resources required for the project

We may phrase challenges in terms of "we don't know..." (research)

- How to do something
- The impact of something
- How something is currently working

Or in terms of "we don't have..." (development)

- A platform we need (eg for dataset search)
- A dataset or datasets we need
- A set of guidance we need

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*.

The Working Group used its final meeting before the Summit to develop its short list of potential challenges. However, we want to continue as we have started in working with the wider community in developing these challenges, and so the Summit provides an excellent opportunity to open this next phase of work. The Working Group will bring a shortlist of its initial ideas to discuss and test with the wider plenary, and plans to prioritise challenges using the following criteria:

- Which particularly help with data governance in an AI context?
- Which would help make progress towards the SDGs ie have a public good benefit?
- Which require international collaboration?
- Which require collaboration across governments, business, academia, and the third sector?

We look forward to beginning this next chapter of our work.



## vi) Longer term vision

The Working Group will be guided in its longer term vision by the realisation of GPAI's overall mission<sup>2</sup>. 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. This then marks the opening of a project lifecycle for future years.

The 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.

Our Framework notes that the mandates of the Data Governance Working Group and the Responsible AI Working Group in particular are closely related and overlap to a certain degree.

Generally speaking, the Responsible AI Working Group will be looking more into how to model AI development and how to employ which datasets, in order for AI to be shaped and to function in a responsible manner (e.g. without any undue bias). The Data Governance Working Group will therefore focus on how to collect and manage the data responsibly in the first place, 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 (e.g. data subjects and those belonging to communities about which data is collected).

However, there are many other links across the Working Groups and we look forward to offering foundational assistance in their future projects.

---

<sup>2</sup> GPAI's mission as set out in the Terms of Reference is "to support and guide the responsible adoption of AI that is grounded in human rights, inclusion, diversity, innovation, economic growth, and societal benefit, while seeking to address the UN Sustainable Development Goals. GPAI will facilitate international project-oriented collaboration in a multistakeholder manner with the scientific community, industry, civil society, international organizations, and countries, taking into particular account the interests and contributions from emerging and developing countries. It will also monitor and draw on work being done domestically and internationally to identify gaps, maximize coordination and facilitate international collaboration on AI"

## Annex 1

### GPAl Data Governance Working Group Mandate

#### Scope of the Working Group

- The Data Governance Working Group will **collate evidence, shape research, undertake applied AI projects and provide expertise** on data governance, to promote data for AI<sup>3</sup> being collected, used, shared, archived and deleted in ways that are consistent with **human rights, inclusion, diversity, innovation, economic growth, and societal benefit, in congruence with the UN Sustainable Development Goals.**
- Co-chairs will coordinate with their counterparts on the **Responsible AI Working Group** to align on sector-specific use cases and applied AI projects, and with the **Commercialisation and Innovation Working Group** on intellectual property issues. The **International Centre of Expertise in Montréal for the Advancement of Artificial Intelligence / Centre d'expertise de classe mondiale pour l'avancement de l'IA** ("CEIMIA") will coordinate with **the OECD** to consolidate outputs.

#### Deliverables to be presented at the Multistakeholder Experts Group Plenary

- **A brief framework defining data governance<sup>4</sup>**, including a literature review of other data governance research, and breaking down the topic into areas such as technologies, laws/policies, and organisations/institutions. This should also highlight the need to address both opening data up and closing data down, and the need to talk about both personal data, non-personal data (including the implications of intellectual property protections) and collective group data (e.g. for an indigenous community such as the Maori). This should cover links with the other Working Groups (such as, for example, on any necessary exceptions in intellectual property and copyright law concerning data). The goal of this is to provide some general scoping and structure to the Working Group's work.
- **A description of the role of data in the development and use of AI grounded in human-centred values**, including the identification of particular (types of) datasets (eg facial recognition datasets, datasets supporting the development of autonomous vehicles) that particularly support AI innovation. This research should review the literature on the economic and social benefits and risks that arise from better access to and reuse of data. The goal of this research is to situate the importance of data to AI development and to identify specific (open) datasets that could be worthy of national support or international collaboration. This should include guidelines on how to make this data open and reusable for development of AI (licensing, and/or exceptions in the law

---

<sup>3</sup> The Mandate draws upon the definitions set out within the [OECD Recommendation on Artificial Intelligence](#) for this purpose

<sup>4</sup> See for example: <https://royalsociety.org/topics-policy/projects/data-governance/>



## Deliverables to be advanced in the Medium-Term (preliminary list discussed in July 2020)

Three reviews as described below, each identifying:

- International examples;
- Existing good practices and recommendations for government;
- Areas that provide opportunities for further collaboration through cutting-edge research and experimentation through pilot projects;
- Areas that require a deeper investigation by the Working Group;
- Cross-cutting areas of dependency and complementarity between the three reviews.

(1) A review of the **state of the art in technical approaches** to data governance, covering, at a high level:

- Machine-readability of data and metadata, including data about provenance and dataset audit cards<sup>5</sup>;
- Privacy-enhancing technologies, including pseudonymisation and anonymisation techniques, federated machine learning, differential privacy, and the creation of synthetic data
- Bias detection and correction techniques ;
- Technologies that support data access, controls and consent management (e.g individual data wallets), logging and auditing.

(2) A review of the **state of the art in legal approaches** to data governance, covering, at a high level:

- Intellectual property law as it applies to data, including collecting / generating / gathering data, deriving datasets, using and sharing data;
- Data protection law, in particular its application in the creation of AI;
- Legal and regulatory measures that enforce access to data and reuse of data, including freedom of information, reuse of public sector information, access by statistics agencies, city access to private data and data portability;
- The use of voluntary mechanisms, certification, audit, codes of practice etc applied to data.

(3) A review of the **state of the art in organisational and institutional approaches** to data governance, including approaches that focus on:

- Individual data sovereignty and empowerment, such as personal data stores, representatives, trusts and cooperatives;
- Community data sovereignty and empowerment such as civic data trusts;
- Data access for research, innovation, and value creation, such as organisational data trusts, clubs
- Collaborative maintenance of common assets.
- Further outputs are focused on deeper reviews and the development of recommendations that could target specific outcomes described above. For example :
- Scoping and piloting the creation of specific representative, open, datasets that support the development of AI systems;
- A review of the literacy, skills and training required for those working that supports data governance;

---

<sup>5</sup> See for example: <https://arxiv.org/abs/2006.16923>

- Focusing on particular practical, sector-specific use cases and demonstrating an end-to-end data governance process for those use cases, identifying specific examples, good practices and recommendations for government, including areas for potential harmonisation.

## Annex 2

### Project Steering Committees

#### *Role of Data in AI*

Jeni Tennison (Co-Chair), Maja Bogataj Jančič (Co-Chair), Takashi Kai, Dewey Murdick, Shameek Kundu, Alejandro Pisanty Baruch, P J Narayanan

#### *Data Governance Framework*

Christiane Wendehorst, Jeni Tennison, Maja Bogataj Jančič, Bertrand Monthubert, Takashi Kai, Shameek Kundu, Alejandro Pisanty Baruch, Te Taka Keegan, Kim McGrail, Josef Drexl