Data Justice Policy Brief:
Putting data justice into practice

November 2022

GPAI | THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

This report was developed by Experts and Specialists involved in the Global Partnership on Artificial Intelligence's project on Data Justice. The report reflects the personal opinions of the GPAI Experts and Specialists involved and does not necessarily reflect the views of the Experts' organisations, GPAI, or 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..

Acknowledgements

This report was developed in the context of the Data Justice project, with the steering of the project Co-Leads and the guidance of the Project Advisory Group, supported by the GPAI Data Governance Working Group. The GPAI Data Governance Working Group agreed to declassify this report and make it publicly available.

Co-Leads:
Alison Gilwald*, Research ICT Africa/University of Cape Town
Dewey Murdick*, Georgetown University

The report was written by: **Alison Gilwald***, Research ICT Africa; **Kelle Howson**‡, Research ICT Africa; **Parminder Jeet Singh**‡, IT for Change; **Anita Gurumurthy**‡, IT for Change; with the significant contribution of the following GPAI Expert: **Dewey Murdick***, Georgetown University.

GPAI is also grateful to the following individuals for reviewing the work: **Paul Dalby***, Australian Institute for Machine Learning; **Christiane Wendehorst***, European Law Institute; **Yannic Duller**‡, University of Vienna. GPAI would also like to thank the following individuals who provided valuable comments to enrich the report: **Morgan Briggs**†, The Alan Turing Institute, **Shmyla Khan**†, Digital Rights Foundation; **Sandra Aceng**†, WOUGNET. Finally, GPAI would like to acknowledge the efforts of colleagues at the International Centre of Expertise in Montréal on Artificial Intelligence (CEIMIA). We are particularly grateful for the support of **Arnaud Quenneville-Langis** and **Edward Teather** from CEIMIA, and for the dedication of the Working Group Co-Chairs **Jeni Tennison*** and **Maja Bogataj Jančič***.

* Expert of GPAI's Data Governance Working Group
** Observer at GPAI's Data Governance Working Group
† Invited Specialist
‡ Contracted parties by the CEIMIA

Citation

GPAI (2022). Data Governance Working Group – A Framework Paper for GPAI's Work on Data Governance 2.0, Report, November 2022, Global Partnership on AI, Paris.

Contents

- Globally, the collection and use of non-personal data remains almost completely unregulated. Yet, when accumulated at scale, data confers almost unimaginable power to govern peoples' lives and extract value. Without regulation and other interventions, **data has become a site of injustice—exacerbating inequality and marginalisation**.

- The concept of data justice has been advanced by researchers and activists as a way of talking about and challenging data-related harms to people and communities. There is a **need to bridge the gap between theory and practice in the area of data justice**.

- Regulation of data-related harms has largely focused on safeguarding personal data and privacy. This does not necessarily produce just outcomes. To advance data justice, **more expansive approaches are needed to govern data in the collective interest**, or for the common good, while still protecting individual privacy.

- Data collection and data-driven systems are also **exacerbating social, and economic inequalities as they amplify bias, marginalisation and discrimination** against vulnerable or disadvantaged groups (such as women)—this demonstrates the need for rights such as equal representation in data, non-discrimination and the protection of group identities and indigenous knowledge systems.

- Currently, **ownership of and the ability to utilise data is highly concentrated** amongst a small number of platforms primarily in the US and China. This underscores the need for economic regulation to enable greater access to and control over the resources of data and artificial intelligence (AI), as well as the capabilities and infrastructure required to utilise them, to redistribute opportunities and benefits more evenly within and between countries.

- **Change can only be achieved through responsive policy making that is sensitive to the most vulnerable and accountable to all**. This brief provides 10 key policy recommendations for government and other decision makers for realising a more just data future. These coalesce around localising and contextualising data rights and protections; ensuring affected communities can participate meaningfully in data governance; and guarding against the extraction and concentration of data as a resource and means of production.

Introduction

Data[1] has, over a relatively short period, emerged as a defining force of social organisation and of global capitalism, which can be harnessed to influence and govern almost every aspect of our day-to-day lives. Data is a valuable resource, but how it is valued is contested. It is generally through the processing, transmission, storage and combination of individual pieces of data, that value is added. While there are many competing typologies of data, they can be broadly organised into the categories of personal and non-personal data.

In economic terms, data can be understood as a public good in that it is non-rivalrous and can be made non-excludable. Goods are non-rivalrous when their consumption by anyone does not diminish the consumption by others. Like most public goods, it is an impure public good because people can be excluded from using it (unlike an entirely non-excludable public good such as air) for instance when personal data is extracted by platforms for commercial use. The regulation of data as a public good thus entails controlling the degree of excludability—either by preventing access (through data protection) or enabling access (through data sharing or open data).

Personal data has been the focus of regulation to protect the privacy of individuals most notably often referred to in this context as data subjects - most notably in the European Union with the General Data Protection Regulation (GDPR).Yet non-personal data remains almost completely unregulated as does personal data in many parts of the world.

In part as a result of this lack of regulation, the ownership of and ability to derive value from data is concentrated in a small group of powerful actors, and specifically in a few global platforms. Over 90% of the market capitalisation value of the 70 largest platforms is estimated to be concentrated in two countries—China and the United States of America. By contrast, Europe accounts for 4% and Africa and Latin America together account for 1% (UNCTAD, 2019).

Disparities do not only occur at the global level between regions, though. This rapid datafication (or the increasing generation and analysis of digital data in society) is resulting in an uneven distribution of opportunities and harms following historical patterns of social and geographical inequality both within and between countries. Such harms include the appropriation of valuable data from communities, as well as the marginalisation, misrepresentation or erasure of communities in and through data and data-driven systems.

The uneven distribution of opportunities and risks associated with datafication correlates strongly with the levels of human and economic development of countries and the inequalities between and within countries. The ability of countries and regions to counter these trends is dependent on their ability to create an enabling environment for data-driven value creation that is more inclusive and equitable. As a key economic resource, creating data commons equitably available to all, would ensure much better distribution of economic power, and therefore greater economic justice.

[1] In this brief, we use “data” to mean digital data. Digital data refers to any form of information represented digitally—this can be organised or unorganised, qualitative or quantitative. Data represents and abstracts information about the world, and forms the building blocks of knowledge and intelligence (Kitchin, 2014).

Regulation at national and supra-national levels attempts to tackle some of the harms of datafication in some places. While more progressive conceptions of data governance are emerging in Europe, legal and institutional responses have for the most part drawn on a narrow conception of data rights at the individual level, focusing chiefly on personal data protection and privacy, and—more recently with regards to non-personal data—the promotion of economic growth. At the same time, regulation has been formulated in a broad and general way, overlooking contextual specificities and not providing sufficiently narrow implementation, accountability and enforcement frameworks.

These measures largely fail to respond to the harms of datafication which take place at the collective level, as well as the role of data as a resource and a means of production which confers enormous economic power when accumulated in large volumes (Singh & Gurumurthy, 2021). To address this gap, policy approaches are needed to govern data in the collective interest or for common good. These must be guided by overarching principles of equity and justice, however they must also be responsive to local and contextual differences in how datafication and artificial intelligence is experienced.

The concept of data justice has emerged in response to these critical gaps. It has been advanced by a global community of researchers and activists as a way to talk about and challenge the complex, overlapping harms of datafication at multiple scales (from the personal, to the local, to the global) (Kitchin, 2014; Dencik et al., 2016; Taylor, 2017). Emerging work is also beginning to link data justice to economic fairness and rights, and to create comprehensive frameworks for what economic justice and a fair distribution of benefits might look like in a datafied society (Singh & Gurumurthy, 2021). The Global Partnership on Artificial Intelligence (GPAI) commissioned the Alan Turing Institute to develop a state-of-the-art review of the data justice literature, together with an extensive use case repository and three comprehensive guidelines on data justice for policy makers, developers and communities and end-users. (GPAI 2022). This online resource is publicly available for researchers, practitioners and policymakers.

This policy brief provides an introduction to the developing concept of data justice for policymakers, and practical guidance on how it can be institutionalised at various points in the policy lifecycle, to ensure fairer outcomes as datafication progresses.

Social and economic harms of datafication

Data is societally produced, that is, generated by and about people (through both online and offline activities). However those who generate data, and/or those who are represented in data (which encompasses almost everyone, including people who are not online but whose data is collected and stored online), seldom share in its economic value. Digital capitalism is predicated on widespread "data extraction", whereby data is collected from our activities, often without our knowledge (Sadowski, 2020; Zuboff, 2019), and exploited by data collectors or holders.

Large pools of data are accumulated this way, becoming de facto proprietary data assets concentrated amongst a few monopolistic firms in global centres of wealth and power. As big data becomes an increasingly important resource, this leaves countries in the global South, as well as marginalised communities all over the world, especially disadvantaged (Couldry & Mejias, 2020). In response to this, calls have been increasing from activists, academics and policymakers to recognise data as a public good, to be governed through democratic institutions.

These come in light of the fact that those with the capability to collect very large stores of data can use it in powerful data-driven productive processes such as the development of AI systems.[2] AI is increasingly deployed in almost all economic and social spheres including the provision of public services, the management of labour, and the mediation of our social interactions and the information we are exposed to. Across these areas, AI can predict and influence our behaviour and govern our social and economic relationships in complex and often hidden ways, to the benefit of powerful platforms.

The ability of a select few to convert the resource of data into AI systems perpetuates and accentuates historical social injustice and inequality—particularly for women at the intersections of other inequalities such as class, race, sexual, religious, cultural or physical location. The real world impacts of AI systems depend on what data is used to develop these systems, how that data is processed, and who controls these activities. The data used to develop AI, because it is derived from our social activities, reflects existing biases and inequalities. Human decisions made in the design and development of AI can often serve to exacerbate these biases. All this is placed in a context where currently almost all development and deployment of AI is driven by the motive of profit maximisation.

Without regulation and other interventions that are strongly cognisant of and sensitive to these issues, this inevitably leads to particular groups being discriminated against, disadvantaged, or further marginalised as a result of the deployment of AI systems (Birhane, 2018; Noble, 2018). An example of this is in facial recognition software trained on data that is not representative or processed appropriately. If, for instance, black women's faces are underrepresented in the dataset, the software will be less adept at recognising black women's faces, and will have discriminatory social consequences when it is used in the world.

[2] When data is accumulated in large quantities and aggregated, it can be used in modelling systems—which transform the raw material of data into an output through complex computational processes (algorithms). Machines can be trained with large datasets to perform cognitive functions such as classification, prediction and recommendation. This process is often called machine learning, and the resultant systems which perform these functions previously reserved for humans, are called AI.

Another concern around the collective harms of data has emanated from indigenous communities who may require limitations on access to their data and knowledge systems. This is critical to guard against commercial and scientific practices of knowledge extraction which might occur in data collection and data-driven systems. This calls for nuance in advocacy around open data sharing which attempts to respond to the concentration and enclosure of data assets. Without responsible stewardship mechanisms, open data carries risks not only for personal privacy, but for the appropriation of indigenous knowledge systems. Indigenous data sovereignty has been identified as a key concern by indigenous scholars, in advancing a decolonial data policy agenda (Kukutai & Taylor, 2016).

The ways in which data is collected and used can also have the effect of imposing inappropriate categories and labels onto people or erasing distinctive identities in data's representation of the world. Processes of cleaning, sorting and categorising data in order to train AI systems involve human decision making which can result in binary categorisations being imposed onto people, by grouping them together, or omitting or erasing identity characteristics which are valued or claimed by people. This can reinforce the dominance of particular groups, and erase the identities of other groups, miscategorise them, or render them less visible.

Platform corporations are the key sites of data-based power, increasingly replacing public agencies as the main holders of society-wide data. At the same time, they have consolidated their role as monopoly service providers in a number of sectors. These dual and mutually-reinforcing roles of data-repository and service provider gives platforms extraordinary social and economic power. Platforms leverage their data lock-ins to unilaterally set the terms of engagement with users, including business users (such as e-commerce traders) and workers who are economically dependent on the platform (such as ridehail drivers).

Platforms can employ their enormous repositories of user data to manipulate users and put them at an economic disadvantage, such as by advertising an item to us and/or increasing its price when our data indicates that we are more in need of it. The same platforms move data globally to create intense concentration of AI and data power, largely in just two countries, the US and China. This concentration of data and power with platforms, and its use for economic extraction at multiple levels, has been largely unchallenged so far, though there are emerging theoretical and policy efforts to imagine ways to diffuse and redistribute data as a commons or a public good. A key demand has coalesced around the need to establish, protect and democratise data infrastructures, to enable equitable access to the resource of data.

Although much of the theory and practice around data value creation references private sector value, there is considerable value creation potential in the public sector (Mazzucato 2021). Effective public value creation however, requires a coherent transversal approach to understanding data demand and how it can be used to enhance socio-economic efforts and public service delivery through effective data governance.

Another area of data related injustice is in the impact of data and data-driven systems on workers' rights and labour relations. Datafication affects workers in myriad ways, and overall is leading to increased precarisation, casualisation and commodification of labour, as data and data-driven systems are used as tools by powerful actors to exploit workers and extract more value from labour (Srnicek, 2017). This includes through the use of data to develop algorithms to manage the labour process in different sectors. For instance platform workers' data is used to rank and rate them, and workers rely on these metrics for access to future work—this can significantly reduce workers' power and agency, especially when they are unable to port rankings and ratings off their platform.

Workers directly engaged in the development of data-driven systems can experience specific harms. There is a massive amount of human labour involved in data value chains and the process of developing and maintaining AI systems. Activities such as processing, cleaning and annotating data are critical to AI development, and these tasks are overwhelmingly performed by precarious platform workers in the global South (Gray & Suri, 2019). This reveals how the development of AI (as well as its deployment) perpetuates social and geographical inequalities. These workers are subject to platform surveillance and enclosure through the use of algorithmic rating and ranking systems which are not portable to other platforms. Data that they do not create directly through their work activities, but indirectly through their use of platforms (including in idle time) is collected, extracted and monetised by platforms. Alongside the expansion of employment and social protections to platform workers, measures are needed to realise data justice in relation to platform labour—these include by regulating platforms' ability to surveil workers and collect their data without transparency or compensation, as well as enabling workers to port their work profiles, ratings and reviews between platforms.

Some researchers have viewed data as a new and unique kind of value contribution, which should be recognised. In other words, workers produce value for companies through the generation of data (even if this is secondary to their main job). But many parties alongside workers contribute to the (co-)generation of data in different ways, such as by being the subject of the information, through their activities, or by having rights in a product or service that has contributed to the generation of data. However, despite their contribution and legitimate interest in accessing such co-generated data, data holders may deny contributing parties access due to their de facto power. Recognising different parties' roles as data generators should give rise to other kinds of rights and entitlements, including the right to benefit from one's data, and share in its value, as well as the right to participate in its governance (Singh, 2020).

The extraction of and centralisation of data as a resource, alongside the reinforcement and normalisation of dominant groups in data and through data-intensive systems, and the commodification of labour through data systems, mirrors historical processes of colonisation and imperialism. Unless the power imbalances perpetuated in these processes are recognised, understood and redressed by policy frameworks, unequal and unjust outcomes will proliferate.

Current regulatory responses to datafication

As datafication has accelerated, various norms, ethical codes and legal frameworks have emerged to govern data in different places. However, regulation has struggled to keep up with the multiplying harms and risks of data collection and AI, and the cycle of data enclosure and service monopolisation in platforms that excludes new players and intensifies platform power over users.

Amongst legal instruments, the European Union's General Data Protection Regulation (GDPR, 2016) is considered the global standard. GDPR focuses predominantly on codifying and protecting rights related to personal data. Personal data is generally understood as any information which is related to an individual or could be used (on its own or in combination with other information) to identify an individual.

Whilst personal data is a category of data which does require special protection, as it carries particular privacy and security concerns, the emphasis on personal data in GDPR (and frameworks which take their cue from GDPR) doesn't mitigate the collective, social, and economic harms of data systems. Data becomes useful and valuable only in large quantities, and only to those who have the infrastructure and capabilities to exploit it. Individual personal data is unlikely to hold much value, but when data is extracted and aggregated (even in anonymised form), its immense value is realised further down the chain, and in ways that might not even be anticipated at the time it is collected.[3] As such, protecting people's access to and ability to benefit from their personal data does little to stop the process of the extraction of data as a resource from groups and communities.

Although data is regularly generated by various parties, control usually resides exclusively with large and economically powerful players. Based on notions of fairness, the Principles for Data Economy of the American and European Law Institute (ALI & ELI 2021) suggest affording data rights to persons or communities that had a share in generating the data (e.g. by being subject of the information, performing an activity by which the data were generated, or having rights in a product or service that has contributed to the generation of data).

A rights-based approach to data governance centres on the rights of data subjects, both individual and collective. These include the right to:

- access and port one's data, including for third parties of choice/data stewardship;
- the right to have data holders stop using data, correction of data;
- an economic share in profits derived from data;
- benefit from one's data, and avoid economic harm;
- appropriate representation in data, including to invisibility and remaining absent, and;
- participate in governance of data, and of the relevant economic systems based on data.

[3] Most AI-enabled systems become more and more useful based on recognising patterns in aggregate data, which is why aggregate data is valuable. This utility does require corresponding individual data when applied to a particular individual context. But the more effective the system is based on access to aggregate data the less individual data it may need for its application. This is why aggregate data is so valuable, or at least its immense value should not be lost sight of in what is currently an almost exclusive focus on individual data, even in economic regulation.

While contributing to the generation of data can justify the recognition of a data right against the controller of the data, several other factors, such as the scope and nature of the party's contribution or the legitimate interests of other parties, need to be considered as well. The concept of rights in co-generated data has been picked up by, among others, the United Nations Commission on International Trade Law, the German Data Ethics Commission, and the European Commission in its Data Act Proposal.

Despite these calls, regulation to protect people from data-related harms remains primarily oriented towards individual privacy concerns. However, data-related harms to individuals do not only arise from personal identification. Individuals may be harmed when harm is directed at groups and communities they are part of based on collective data. As outlined above harms also arise through the processing of their data as well as their contact with artificial intelligence systems predicated on data. These harms can include erasure and miscategorisation of, or discrimination against vulnerable people and communities. They can also include socio-economic control and manipulation, as our own data may be used by powerful actors to predict and even influence our behaviour. Attempts to mitigate these harms have so far mostly been limited to private or voluntary codes of ethics for AI developers, which have proliferated amongst global forums (e.g UNESCO, 2021) and industry bodies (e.g. IEEE).

While these codes are important in establishing ethical norms and conventions in response to accelerating digitalisation, they have critical limitations in realising just outcomes. Whilst many codes are produced through stakeholder consultation, they lack meaningful democratic governance and accountability to affected communities, such as mechanisms for third-party auditing. Furthermore, consultations are generally limited to "experts" in the field as opposed to genuine community engagement that is built on public participation in policy processes and administrative justice.. Ethical codes are useful to inform and complement enforceable regulation, but in a regulatory vacuum they serve to substitute for political contestation and democratic governance, and can be used as a way for established interests to avoid greater accountability.

Some developments have occurred with respect to regulating data as an economic resource, mostly emerging from the EU. These have focused on breaking up data market concentration and redistribution through data access and portability. GDPR provides the right for individual data subjects to access and port their data from current holders to third-parties, in an attempt to empower people to resist lock-in to dominant platforms. The Digital Markets Act also allows business users of large platforms (like ridehail drivers and e-commerce traders) to access, and port, data related to their activities and customer interactions (subject to privacy protections), including in aggregate form. The draft Data Act allows Internet of Things (IoT) device owners or users to access and port data that may have been collected via their device by a platform or application. Data porting also requires cloud interoperability, which the draft Data Act proposes to mandate.

However, there is little incentive for individuals, business users of platforms, or IoT device owners to access or port their data because there is little they can do with it in isolation, as data is a systems-level resource which requires scaled infrastructure to be of value. In addition, these frameworks do not allow data subjects to benefit from the collection and use of their data once it is aggregated and anonymised. Stores of anonymised aggregated data may be of even more value to established platforms than the kind of personal data to which these frameworks enable data subjects access. It is the identification of data patterns which are particularly useful in data-driven innovation and production. To ensure a fair distribution of data benefits, frameworks need to empower the use of data by collective data subjects—made up of affected groups and communities (for example ridehail drivers of a city, or farmers in a given region).

This kind of collective right of a community to its data has been recommended in the draft report of the Indian Committee of Experts on Non Personal Data Framework (Ministry of Electronics and Information Technology, 2020). Possible legal mechanisms for enabling responsible data sharing which still protects privacy also include data cooperatives or data trusts (Ada Lovelace Institute, 2021). The EU's draft Data Governance Act also responds to this gap, proposing to provide a legal basis for neutral and transparent data governance through special "data intermediaries". The EU's GAIA-X project has aimed to enable sectoral "data spaces", or universally accessible, neutral and regulated data markets/infrastructures. These are governed collaboratively by contributors, who must be EU-headquartered enterprises.

However, new international approaches are also needed to mitigate between-country economic inequalities in data access and use. Developing countries at the World Trade Organisation have been calling for data to be recognised as a national resource similar to natural resources (like minerals or oil), to allow for strategic digital industrialisation. The 2021 Digital Economy Report of the United Nations Conference on Trade and Development called for developing an international framework for data governance. Regulating data flows for economic data justice at the international scale is complex, as it also calls into question the ability of actors such as the EU to enforce their own privacy and security provisions, as well as new economic data rights for business users of platforms, and IoT owners/users. However without such international measures, between-country inequalities in data access and use are certain to perpetuate historical patterns of dispossession and disadvantage.

As countries assert their national sovereign rights over data, it is important to recognise the local political context in which this is done and that calls for data localisation do not enable unproductive protectionism or another way for states to exercise more control over citizens' data without any benefit to citizens.

Data justice advocates have called for the principle of data justice itself to be institutionalised into regulatory frameworks, and for these to be based on expanded conceptions of data rights. The African Union adopted a rights-based Data Policy Framework in 2022 aimed at ensuring Africans benefit socially and economically from their data as well as data systems and infrastructures. The framework explicitly references data justice as a guiding principle: "Member States shall ensure data collection, processing and usage are just and lawful, and data should not be used to discriminate unfairly or infringe peoples' rights." It explains the principle as: "...seeking to ensure that the increasing reliance on data, especially for automated decision-making, does not perpetuate historical injustices and structural inequalities… Data justice also extends beyond notions of political rights and justice to social and economic rights and regulation that is necessary to redress inequities and enable people to exercise their rights" (2022, p. 28).

Much work remains however to further develop and operationalise this framework, including clearer enunciation of collective rights and economic rights to data, and through translating it into an action plan that is locally appropriate, clear and implementable, which is currently underway. Other data treatise and regulation frameworks in Africa, but also elsewhere, including GDPR have been criticised for their lack of practical and contextualised guidelines for enforcement and implementation.

One of the outcomes of sometimes explicit export, sometimes default application of data protection mechanisms of the GDPR where neither the institutions nor data subjects are able to effectively deploy them have meant that many of the countries in the "developed world" enjoy greater data protection as opposed to the rest of the world—creating a hierarchy of data protection and rights afforded to citizens.

Finally, some progress has been made especially in Europe to protect digital platform workers. However, this pertains predominantly to the employment rights of location-based workers such as ridehail and delivery drivers. By contrast, very little work has been done to protect the millions of precarious workers engaged in remote data-processing and AI development. This is largely due to the cross-border nature of this work and the fact that these workers are classified through platforms' terms of service as independent contractors. As they complete work for dozens of clients all over the world in the space of a day, and are contracted by third-party platforms in a different jurisdiction, it is very unclear which jurisdiction is responsible for protecting them, and what the platforms' and clients obligations are. With regard to workers' specific data rights, regulation has not grappled with the extraction and monetisation of platform workers' data without their consent, nor adequately addressed platforms' use of workers' data to lock them into dependent relations, though non-portable ranking, rating and credentialing systems.

As we have outlined, significant policy gaps remain in ensuring data justice across both social and economic dimensions. These include in the areas of mitigating the biased or discriminatory impacts of AI, which can cause harm by exposing, miscategorising or invisiblising groups—harms which are not addressed through personal data protection. Moreover, national, regional and international regulation is needed to affirm and ensure economic data rights including a fairer distribution of control over and benefits from data. In addition, solutions are needed to protect workers from data-driven exploitation. To mitigate both the social and economic harms of data and AI, regulation is needed to ensure collective—in addition to personal—data rights.

Policy recommendations for advancing data justice

The high-level policy recommendations below are intended to be applicable at different levels of government, as well as in non-government and private sector settings. Within specific country and regional contexts and through the participatory processes intrinsic to democratic policy formulation more detailed guidelines for achieving data justice can be developed.

1. Basing data regulation on rights: Data today is so closely intertwined with our social and economic organisation and outcomes, that the need to base data regulation on human and community rights has to be recognised as a key political imperative. This requires developing actionable data rights framworks, which include economic and collective rights to data. A basic data rights framework should include: the right to benefit from one's data, and to not be harmed by data collection and use; the right to access and port one's data; the right to appropriate representation in data, including to remaining invisible; the right to participate in governing one's data, and the data systems based on it; the rights to alternative and collective forms of data stewardship.

2. Democratic participation of affected communities: Data justice requires policymakers to identify the full set of stakeholders who might be impacted by data collection and use, and data-driven activities. Individual and collective data subjects, as well as primary data generators, are essential stakeholders. Their participation must be built-in democratically to the design, development and deployment of data-intensive systems, including AI.

3. Contextualisation and localisation: While larger frameworks of data rights, and data justice—and transversal laws and policies based on them—are necessary and useful, data justice also requires policymakers to move beyond one-size-fits-all solutions. Data justice when put into practice looks different in different places and contexts, and these nuances and differences should be locally developed through appropriate participatory exercises, and in consideration of the political economies and institutional endowments of countries.

4. Equitable access to resources: Material inequality and structural exclusion can prevent marginalised groups from sharing in the benefits of data-driven systems. To mitigate this, policy needs to ensure equitable access to skills development and digital infrastructure including connectivity and computing resources, as well as data assets—especially where communities have contributed to the generation of these data assets.

5. Preventing anti-competitive data practices: Trade and competition and antitrust law should be updated to respond to the role of data and platforms in economic structures. This includes preventing data hoarding, and requires data sharing to address unfair data practices by market gatekeepers.

6. Enabling alternative forms of data sharing/stewardship: Equitable access to data can be achieved through responsible data sharing models such as access to data, or data commons, in managed safe conditions. Vehicles such as data trusts or data cooperatives can be empowered to manage data in the collective interest.

7. National data sovereignty and global governance: Sovereignty as a right of a national community to manage its affairs, including its resources, is enshrined in various international human rights covenants. Data (and the ability to utilise it) is so central a resource to our social and economic organisation today that principles of national sovereignty must refer to data and data skills and infrastructure. This applies at personal, enterprise, community and national levels, but must also be informed by the needs of the global digital society and economy—particularly as some of the most intractable governance issues relating to the globalised and cross-border flows of data can only be addressed through international cooperation and solidarity. It is vital however, that the human rights and democratic precepts underlying multilateral conventions upholding national sovereignty are used to redress inequality and enhance data justice, not abused to further marginalise the dispossessed or control legitimate contestation and dissent.

8. Workers' data rights: Alongside the need for the expansion of employment and social protections to platform workers, regulation is needed to elaborate and advance workers' data rights. Specific data rights for workers should include limitations on the ability of platforms or employers to collect workers' data without their consent, and to monetise workers' data without ensuring workers share in its value, and possibly some participation in governing the corresponding data-based systems. In addition regulation should ensure that platform workers can own and port their work profiles, ratings and reviews off-platform.

9. Transparency in data practices and systems: Those with power in processes of collection use of data and data-driven innovation should be obliged to make information publicly available about what data is collected and how it is used, including information about AI inputs, and algorithms, and to provide this information directly to impacted individuals and communities.

10. Appropriate frameworks for redress: Where data collection and use of data, and data-driven systems result in breaches of individual or collective rights, and injustices, there must be clear, institutionalised processes for individual and collective redress. These should be encoded in local, national and global legal frameworks, and place obligations on powerful data users.

Resources

- GPAI (2022). A Primer on Data and Social Justice, Report November 2022,Global Partnership on AI.

- GPAI (2022). A Primer on Data and Economic Justice, Report November 2022, Global Partnership on AI.

- GPAI (2022). Advancing Data Justice Research and Practice: An Integrated Literature Review, Interim Report, March 2022, Global Partnership on AI. https://gpai.ai/projects/data-governance/advancing-data-justice-research-and-practice-literature-review.pdf

- GPAI. (2022). Advancing Data Justice Research and Practice: An Interim Guide for Policymakers, Interim Report, March 2022, Global Partnership on AI. https://gpai.ai/projects/data-governance/data-justice-in-practice-guide-for-policymakers.pdf

- GPAI. (2022). Advancing Data Justice Research and Practice: An Interim Guide for Developers, Interim Report, March 2022, Global Partnership on AI. https://gpai.ai/projects/data-governance/data-justice-in-practice-guide-for-developers.pdf

- GPAI. (2022). Advancing Data Justice Research and Practice: An Interim Guide for Impacted Communities, Interim Report, March 2022, Global Partnership on AI. https://gpai.ai/projects/data-governance/data-justice-in-practice-guide-for-impacted-communities.pdf.

- GPAI. (2022). Data Justice Stories: Repositories of case studies, Interim Report, March 2022, Global Partnership on AI.https://gpai.ai/projects/data-governance/data-justice-stories-repository-of-case-studies.pdf

- Policy Pilot Partners & Final Reports

  - AfroLeadership — Report
  - The Collaboration of International ICT Policy in East and Southern Africa (CIPESA) — Report
  - The Centre for Intellectual Property and Information Technology Law (CIPIT) — Report
  - Digital Empowerment Foundation — Report
  - Digital Natives Academy — Report
  - Digital Rights Foundation — Report
  - EngageMedia — Report
  - GobLab UAI — Report
  - Internet Bolivia — Report
  - ITS Rio — Report
  - Open Data China — Report
  - Women of Uganda Network (WOUGNET) — Report

References

Ada Lovelace Institute. (2021). Exploring Legal Mechanisms for Data Stewardship. Working Group Final Report. UK AI Council. https://www.adalovelaceinstitute.org/wp-content/uploads/2021/03/Legal-mechanisms-for-data-stewardship_report_Ada_AI-Council-2.pdf

American Law Institute and European Law Institute (2021). Principles for a Data Economy. https://www.principlesforadataeconomy.org/fileadmin/user_upload/p_principlesforadataeconomy/Files/Principles_for_a_Data_Economy_ELI_Final_Council_Draft.pdf

Birhane, A. (2021). Algorithmic injustice: a relational ethics approach. Patterns, 2(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Couldry, N., & Mejias, U. A. (2020). The costs of connection: How data are colonizing human life and appropriating it for capitalism. Stanford University Press.

Gray, M. L., & Suri, S. (2019). Ghost work: How to stop Silicon Valley from building a new global underclass. Eamon Dolan Books.

Institute of Electrical and Electronics Engineers Standards Association. (n.d.). Raising the Standards in Artificial Intelligence Systems (AIS). https://standards.ieee.org/initiatives/artificial-intelligence-systems/

Kukutai, T., & Taylor, J. (2016). Indigenous data sovereignty: Toward an agenda. ANU press.

Mazzucato, M (2019) The Value of Everything: Making and Taking in the Global Economy, Penguin, Random House, London.

Ministry of Electronics and Information Technology. (2020). Report by the Committee of Experts on Non-Personal Data Governance Framework. Government of India. https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. New York University Press.

Sadowski, J. (2020). Too smart: How digital capitalism is extracting data, controlling our lives, and taking over the world. MIT Press.

Singh, P.J. (2020). Economic Rights in a Data-based Society. Public Services International. https://library.fes.de/pdf-files/iez/16034.pdf.

Singh, P. J., Gurumurthy, A., (2021). Economic Governance of Data - Balancing individualist-property approaches with a community rights framework. IT for Change. https://itforchange.net/sites/default/files/1880/Economic-governance-of-data.pdf

Solano, J. L., Martin, A., de Souza, S., & Taylor, L., (2022). Governing data and artificial intelligence for all: Models for sustainable and just data governance. Brussels: European Union. https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729533/EPRS_STU(2022)729533_EN.pdf

Srnicek, N. (2017). Platform capitalism. John Wiley & Sons.

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. Big Data & Society, 4(2), https://doi.org/10.1177%2F2053951717736335

UNESCO. (2021, 24 November). Recommendation on the Ethics of Artificial Intelligence. https://en.unesco.org/artificial-intelligence/ethics#recommendation

United Nations Conference on Trade and Development. (2019). Digital Economy Report 2019: Value and capture: Implications for developing countries. Geneva: United Nations.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile books.