

WORKINGPAPER

November 2023

Fair and Safe Datafication

Isabel Rocha de Siqueira, PhD¹

This paper was presented at a joint expert event co-convened by UNU-CPR and the International Chamber of Commerce on 8 November 2023 in New York.

UNU-CPR Working Papers are research papers that have not been peer-reviewed or undergone a thorough editing and publication process. Written by subject matter experts, they offer unique insights and perspectives in response to current debates on issues of strategic interest to UNU-CPR audiences.

¹ Director of the International Relations Institute (IRI), Pontifical Catholic University of Rio de Janeiro.

Introduction

Between 2010 and 2020, the data market grew by almost 5,000 per cent.¹ A report produced in 2014, by the UN Independent Expert Advisory Group Secretariat, *A World that Counts*, estimated that 90 per cent of all data in the world had been produced between 2012 and 2014 alone. The report was elaborated precisely to discuss the notion of a ‘data revolution’ signifying “[a]n explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data.”² Furthermore, as noted in the report of the UN High-Level Advisory Board on Effective Multilateralism, *A Breakthrough for People and Planet: Effective and Inclusive Global Governance for Today and the Future*, “[d]ata flows are expected to reach triple the current volume by 2026.”³ These numbers testify to the urgent need to assess and act upon the opportunities and dangers of datafication.

Datafication involves the practices of systematization and analysis of data, so that data can be instrumentally used for prediction.⁴ This means that the more datafication is effectively practiced, ideally, the more prepared the world is for coming challenges. Nevertheless, as with most goods produced, the access to data and established capacity for datafication in the digital world is not egalitarian, neither within nor between countries, and this is a truth well-acknowledged, if yet to be systematically addressed.⁵ The fact is that “while 90 per cent of people use the Internet in developed countries, only 57 per cent of the population of developing countries is online,” and the cost of access is higher in developing countries.⁶ There are multiple challenges posed by these inequalities, some of which require financing, but with most requiring effective political will.

Nurturing and achieving that political will has been made more difficult as multilateralism itself has been weakened by different factors in the past few years. Its somewhat precarious current functioning means that any initiatives to address the need for fair and safe datafication worldwide bumps into the significant challenge of strengthening the world’s multilateral institutions, but also reforming them, so that they can better address the urgent digital needs of our time.

¹ Gill Press, *54 Predictions About The State Of Data In 2021* (Forbes, 2021). Available at <https://www.forbes.com/sites/gilpress/2021/12/30/54-predictions-about-the-state-of-data-in-2021/?sh=65c3bd3a397d>. Access: 01 Nov 2023.

² Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG), *A World that Counts* (United Nations, 2014), p. 6.

³ High-Level Advisory Board on Effective Multilateralism (HLAB), *A Breakthrough for People and Planet: Effective and Inclusive Global Governance for Today and the Future* (New York: United Nations University, 2023), p. 40.

⁴ Viktor Mayer-Schönberger and Kenneth Cukier, *Big data: A revolution that will transform how we live, work, and think* (Boston: Houghton Mifflin Harcourt, 2013) p. 78.

⁵ United Nations General Assembly, *Information and communications technologies for sustainable development*, A/RES/77/150, 20 December 2022.

⁶ *Ibid.*, p. 9.

In order for datafication and multilateral initiatives to be more inclusive, a few basic questions will need to be brought to the center of our discussions:

- **What are the issues we produce data *about*?**
- **What do we produce data *for*?**
- **Does the data produced lead to improved decision-making?**

These questions aim to guide critical and innovative thinking, *not* for the sake of data production, but **for the benefit of people and planet**, i.e development, peace, and equality. When we consider the sheer number of reports being produced, and the fact that interoperability of data systems is still an immense challenge – compounded by increasing use of non-traditional data,⁷ should resources be used for producing data whose users are not yet known to producers? Moreover, if data capacity is high, with new information and communication technologies (ICTs) generating new types of data at top speed,⁸ do we need to be producing data that is not of clear usability? What other actions could we be taking with those resources and, most importantly, what if un-earmarked data ends up doing more harm than good?

These questions do not intend to dispute the relevance of data, but instead aim to foreground it within **ethical basic commitments**, thereby leading us to the third question about *how data can impact decision-making*. It is well known by researchers that the path from information to decision-making is never straight-forward, as one needs to factor in political will. Furthermore, as seen within the past few years, people also have their own tendencies when it comes to accepting or weighing in new information.⁹ Other elements are often required to lead to the expected impacts in terms of dissemination and decision-making, such as context, language, visualization, and narrative.¹⁰ Bearing this in mind, it is important to investigate and elaborate the **many possibilities through which local participation can complement, correct, and better direct the effective, safe, and fair global use of data** .

The above questions should be taken into consideration at the organizational level, as guidance for avoiding digital harms, but also for positively developing new methods, mechanisms, and spaces for dialogue and meaningful participation. Indeed, **attending to the ways we can make datafication more inclusive may also lead to more inclusive multilateral institutions and vice-versa, as multi-stakeholder partnerships rely on institutional design and practices**. ‘Meaningful partnerships’ require participation that is qualitatively different from mere counting and are instead based on meaningful participation and accountability.¹¹

⁷ Luis González Morales and Tom Orrell, *Data interoperability: A practitioner's guide to joining up data in the development sector* (Dubai: United Nations World Data Forum, 2018).

⁸ C. Vijayakumar, *The 2030 Agenda's data challenge. Approaches to alternative and digital data collection and use* (GIZ, 2020). Accessible at: https://www.data4sdgs.org/sites/default/files/le_uploads/2030AgendasDataChallenge.pdf.

⁹ S. Jasanoff, “Virtual, visible, and actionable: Data assemblages and the sightlines of justice,” *Big Data & Society* Vol. 4 No. 2 (2017).

¹⁰ P. Dourish and E. Gómez Cruz, “Datafication and datafiction: Narrating data and narrating with data,” *Big Data & Society* Vol. 5 No. 2 (2018); I. Rocha de Siqueira and L. Ramalho, “Participatory methodologies and caring about numbers in the 2030 Sustainable Development Goals Agenda,” *Policy and Society* Vol. 41 No. 4 (2022), 486-497; I. Rocha de Siqueira and L. Ramalho, “The Challenges of Governance in a Datascape: Theorizing the Role of Non-Extractive Methodologies in the 2030 Agenda,” *Global Digital Data Governance Polycentric Perspectives* eds. C. Aguerre, M. Campbell-Verduyn and J. Aart Scholte (New York: Routledge, 2024).

¹¹ E. Sondermann and C. Ulbert, “Transformation through ‘meaningful’ partnership? SDG 17 as meta-governance norm and its global health implementation,” *Politics and Governance* Vol. 9, No. 1 (2021), p. 152–163.

The following recommendations are set against the background of discussions regarding a **shift in digital and data governance**. Mostly, the recommendations are located in the intersection between a core commitment to protecting human rights, and the need to conceive a mutual commitment framework between public and private sectors in data sharing in general, and especially in advance of crises.

The Context of an International Decade For Data and a Potential Global Data Compact

The call for an **International Decade for Data**, as envisioned in *A Breakthrough for People and Planet*, is the kind of initiative that is not only symbolic in an important way, but can be effective in building up momentum and securing resources for monitoring and accountability of data production and circulation. Potentially leading up to a **Global Data Compact**, the UN-led initiative can be most useful in making sure diverse stakeholders remain engaged with issues of *rights, safety, and fairness*, which might otherwise be lost as technologies advance at an ever-quicker pace.

Global Compacts are long-established UN instruments. They are examples of international public-private partnerships that serve the purpose of advancing the exchange of good corporate practices, provide mutual learning, and help establish common principles and standards.¹² For a Global Data Compact to advance concerns with a fair and safe datafication, it is vital that a diversity of stakeholders are able to participate in a significant way, and that enough knowledge is produced as the frontiers of datafication keep being pushed. With that in mind, the main recommendation presented here is for the International Decade for Data to establish a commitment to **data literacy** through an **education track**.

Establishing an Education Track Within the International Decade for Data

The International Decade for Data can help address the many problems of fragmented, polycentric data governance, offering spaces for collaborative action. However, in order for such initiatives to also move datafication towards a fair and safe scenario, marginalized groups and other relevant stakeholders within civil societies worldwide need to have a seat at the table. This should be done the ‘conventional’ way, through making sure effective multilateralism takes place via providing opportunities for all stakeholders to be truly and properly listened to in relevant fora. Nevertheless, considering all recommendations regarding effective multilateralism as we approach the *Summit of the Future* in September 2024, it is only fitting that significant partnerships within the International Decade for Data and the Global Data Compact are guided by a concern with the future sustainability of such efforts. This can be done through considerable investments in education, especially **data literacy**.

¹² Diane Stone, Leslie A. Pal and Osmany Porto de Oliveira, “Private consultants and policy advisory organizations: a blind spot on policy transfer research,” *Handbook of Policy Transfer, Diffusion and Circulation* ed. Osmany Porto de Oliveira (Cheltenham: Edward Elgar, 2021), p. 184.

1. Investing in Data Literacy

Such investments would attest to a true commitment to mutual learning and safeguarding human rights and the concerns of the most vulnerable groups, whose problems are often invisible to global dynamics.

The recommendation is for **data literacy initiatives that can nurture what some have called “bilinguals,”**¹³ i.e. people well versed both in data science *and* the urgent social and environmental problems of our times. That knowledge cannot be dissociated from some of the main advances in critical thinking taking place on the ground, which are often the only sources of information about people’s urgent problems.

It is also key that leading local NGOs and other civil society groups are taken on board, not only as target publics for these education efforts, but as legitimate partners within their communities. As will be shown in some examples below, such organizations are often already advancing data production through innovative methods, contributing to shedding light upon a variety of issues, but struggling to have access to predictable funding and, therefore, to keep a constant flux of training programmes. These organizations count alongside professionals who are experts in their territories and frequently come to be ‘bilingual’ through their own efforts. Learning and partnering with them can be a powerful accelerator in advancing digital and data governance.

Moreover, an investment in an education track would go a long way in demonstrating public-private partnerships’ concerns with remaining open to constant accountability: advancing quality, timely education oriented towards datafication would guarantee generations of data stewards, capable of not only following the pace of changes, but of leading them in a fair and safe way.

2. Partnering With Academia

Beyond the importance of including academia among the many stakeholders with whom any initiatives for a fair and safe datafication should engage, there must be a serious effort to cultivate open channels of collaboration with universities, especially those with narrower access to funding and which are often tasked with important social and environmental projects within their communities. The International Decade for Data and the Global Data Compact should help harness such **collaborations**, for instance, incentivizing private companies to provide safe access to datasets that can be further scrutinized to support research on key social and environmental issues. Considering corporations have many more resources than most universities, especially in developing countries, sharing certain datasets can help advance scientific innovation, including in terms of social technologies, for the benefit of many.

Commitments such as this and the overall investment in data literacy might not lead to agreements on regulatory practices or guaranteeing ‘Data Free Flow with Trust’ (DFFT)¹⁴, but their value resides in forging a safe and fair environment of accountability. By nurturing future

¹³ “The 100 Questions Initiative,” GovLab, last accessed 29 October 2023, <https://thegovlab.org/project/project-the-100-questions-initiative>.

¹⁴ World Economic Forum, “Data Free Flow with Trust: Overcoming Barriers to Cross-Border Data Flows,” *Briefing Paper* (World Economic Forum, 2023).

scientists and scientific advances beyond their immediate action areas, corporations will be offering a valuable contribution to digital and data governance.

The remaining section of this paper will offer some insights into what kind of education could and should be nurtured within the initiatives suggested above, showcasing the social impacts they can have and what contribution they have made so far that can serve as **learning packages** in the recommended education track.

Learning Packages For Inclusive Datafication: Educating For Rights

Activism on the issue of digital data has produced important innovative thinking and practice. Often, academia and social movements walk hand in hand in advancing these propositions, qualifying research and political action. In the past decades, the global agendas have evolved to an expressed consensus that complex social issues need multifaceted, holistic and multi-stakeholder approaches. However, when it comes to practice, these understandings are also often limited and harmed by **inequalities and discrimination**. **A rights-centered approach to data literacy will make sure it does not reproduce exclusions**. Some contributions from academia and other civil society groups are key examples of how an education track is vital for an inclusive agenda on data and datafication, just as they also show the importance of learning with local experience in that front.

Learning Package 1: Localization - ‘All Data Is Local’

When one says that ‘all data is local’, it is because data always impacts local communities, be it producers, users or non-users. Therefore, it is in the local in the sense that we see both practice and exception, and that one should revert to accountability and fairness.¹⁵ To use an example: the ‘localization’ of the 2030 Agenda¹⁶ in Rio de Janeiro, Brazil, has led a key NGO, Casa Fluminense, to elaborate and convene courses on Public Policy for major grassroots organizations’ teams and leaderships.¹⁷ This initiative has created a ripple effect whereby these agents have set up communitarian groups to think through the localization of the Sustainable Development Goals (SDGs) in their territory. For that reason, they have established local priorities, methodologies, and even new indicators.¹⁸ There are many limitations to what has

¹⁵ Y. A. Loukissas, *All data are local: Thinking critically in a data-driven society* (Boston: The MIT Press, 2019).

¹⁶ Global Task Force of Local and Regional Governments, *Roadmap for Localizing the SDGs: Implementation and Monitoring at Subnational Level* (2015).

¹⁷ Casa Fluminense, *Relatório de Monitoramento: Agenda Rio 2030* (Rio de Janeiro, 2021);

Casa Fluminense, *Agenda Rio 2030: Propostas por justiça econômica, racial de gênero e climática* (Rio de Janeiro, 2022).

¹⁸ - AGENDA SÃO GONÇALO 2030 // @ressuscitasaogoncalo

- AGENDA VILA KENNEDY 2030 // @casadeayavk

- AGENDA CAXIAS 2030 // @movimentacaxias

- AGENDA BELFORD ROXO 2030 // @simeusoudomeio

- AGENDA NOSSA MERITI // @nossameriti

- AGENDA MAGÉ 2030 // @naocorte.plante

- AGENDA REALENGO 2030 // @agendarealengo2030

- AGENDA ITABORAÍ 2030 // @observatorioitaborai

- CARTA DA MARÉ 2030 // @cocozap @redesdamare @data_labe

- PLANO SANTA CRUZ 2030 // @santa.cruz2030

- AGENDA QUEIMADOS 2030 // @ampara.amigosdoparaíso @golfinhosdabaixada @pedalaqueimados @visaocoop

been called Citizen-Generated Data (CGD), among them the difficulty in collecting certain data, the challenge of keeping funding coming for both training new people and the work itself, and not least, the lack of legitimacy when facing other parties.¹⁹ Nevertheless, nothing perhaps has disseminated the SDGs more effectively in marginalized territories, as localized initiatives such as these. Moreover, the data and narratives produced by agents practicing CGD offers lessons on being inclusive and avoiding doing harm. They tend to focus on issues that are ignored, such as the mental health of families who have lost loved ones as victims of police violence.²⁰

To think that ‘all data is local’ is less about equaling the methods and results of initiatives such as CGD to the best practices of international organizations and corporations, and more about considering how data literacy can contribute to inclusive thinking when defining questions that data is supposed to answer.

Learning Package 2: Fighting Inequalities With Data Feminism

Data feminism is “a way of thinking about data, both their uses and their limits, that is informed by direct experience, by a commitment to action, and by intersectional feminist thought.”²¹ Therefore, it can be perceived as a pillar of the argument that ‘all data is local.’ If we are serious about inclusion, fairness, and doing no harm, learning from and respecting direct experience is important. In Latin America, there have been outstanding projects that we can say contribute to advancing data feminism. They have much to teach to multilateral international organizations and both their methods and their experience should support global advancements, such as the International Decade for Data.

- AGENDA JAPERI 2030 // @mobilizajaperi @sepe.japeri @grupocodigo @fppjaperi

¹⁹ S. Jameson, D. Lämmerhirt, and E. Prasetyo, *Acting locally, monitoring globally? How to link citizen-generated data to SDG monitoring. How to Link Citizen-Generated Data to SDG Monitoring* (2018).

²⁰ G. Roza, “Violència mental,” *data_labe*, 15 January 2021, <https://datalabe.org/violencia-mental/>.

²¹ Catherine D’Ignazio and Lauren F. Klein, *Data feminism* (Cambridge, Massachusetts: The MIT Press, 2020), p. 8.

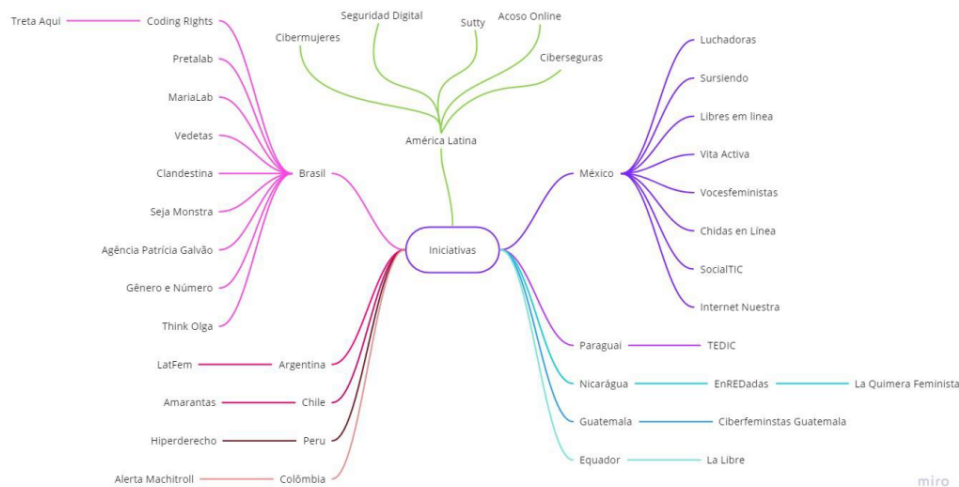


Figure 1: Mapping feminist data projects in Latin America²²

Data feminism has a twofold purpose: to identify how discriminatory practices might be incorporated into data science, and to use data science to fight asymmetries and violence.²³ There have been various advances in terms of questioning the ability of science to include all people and distribute equally the benefits of scientific inquiry, and data science should indeed be a part of this movement.

Learning Package 3: Fighting Data Colonialism

Another key contribution for an agenda on fair and safe datafication is the prolific research that has been denouncing data colonialism as yet another manifestation of North-South divide. Entering an International Decade for Data and advancing a Global Data Compact should not, in any way, further this divide.²⁴ That is, these initiatives cannot privilege economic gains in data flows if these mean increasing inequalities.

On the one hand, it is common for lower-income countries to have larger administrative (government) data gaps; on the other hand, it is also more common that in these countries there are no regulations or infrastructure to face the growing demands of the global data market. Thus, large corporations have taken the lead in ‘counting, categorizing and visualizing’ development issues in low- and middle-income countries, offering their own or user-produced data to donors and investors, something that has become possible with the more widespread use of information and communication technologies (ICTs).²⁵ In face of this reality of ‘deputization’ of data production, multilateral initiatives such as the Global Data Compact can be crucial to create an

²² J. Monteiro Fernandes, *Violência e política: uma cartografia das ciborgues desde a América Latina*. Dissertation presented as partial requirement for obtaining MSc certificate from the International Relations Postgraduate Program of the International Relations Institute (IRI), PUC-Rio, in 2021, p. 93.

²³ D’Ignazio and Klein, *Data feminism*, p. 9.

²⁴ United Nations General Assembly, *Information and communications technologies for sustainable development*, A/RES/77/150, 20 December 2022, p. 9.

²⁵ L. Taylor and D. Broeders, “In the name of Development: Power, profit and the datafication of the Global South,” *Geoforum* Vol. 64 (2015), pp. 229–237.

environment of accountability where public-private partnerships are guided by a rights-based approach to data collection.

To provide another example from Brazil, the project *Desvelar* ('Unveil'), led by activist and scholar Tarcízio Silva, seeks to map the impacts of algorithmic systems and artificial intelligence.²⁶ The cases illustrated have taken place all over the world, and are shown in a timeline which has the impact of making clear the fact that racism has been identified for years now, and is still modelled into some of the technologies used to collect data. Racism is pointed out as the basis of many systems of exploitation and practices of violence, thus, at the basis of data colonialism and 'technological colonialism,' which tends to approach people as 'human natural resources,' that is, as a source for information extraction.²⁷ A rights-based data literacy geared towards a fair and safe digital and data governance would be much aware of these risks and constantly conceive of ways to impede discrimination and other forms of harm.

The Way Forward

The ideas shared in this working paper provide brief recommendations as the world moves towards a Pact for the Future and an International Decade for Data, raising awareness about the need for fair and safe datafication, based on practices that do no harm and are more inclusive and accountable. Great contributions have been produced in communities all over the world, especially in the Global South, and there is much that can and should guide thinking and action in the next few years.

²⁶ See *Desvelar*: <https://desvelar.org/casos-de-discriminacao-algoritmica/>.

²⁷ A. Birhade, "Colonização Algorítmica da África," *Comunidades, algoritmos e ativismos digitais: Olhares afrodiaspóricos* ed. T. Silva (ebook: Editora LiteraRUA, 2020).

United Nations University - Centre for Policy Research

767 3rd Ave Floor 35, New York, NY 10017

This Working Paper is an output of the UNU-CPR initiative, “A Breakthrough for People and Planet: Building Momentum for the Summit of the Future and Beyond”.

