Advancing Data Justice Research and Practice Project

Final Report

ITS

Instituto
de Tecnologia
& Sociedade
do Rio

SUMMARY

INTRODUCTION

This document presents the Institute of Technology and Society (ITS Rio) general assessment and findings from the *Advancing Data Justice Research and Practice* project. This project is a partnership with Alan Turing Institute (ATI) and aims to widen the lens of current thinking around data justice and provide actionable resources that will "help policymakers, practitioners and impacted communities gain a broader understanding of what equitable, freedom-promoting, and rights-sustaining data collection, governance, and use should look like in increasingly dynamic and global data innovation ecosystem".

In this final report, ITS Rio provides their reflections, findings and recommendations from project activities, namely interviews and workshop conducted with policymakers. Our goals throughout the project were to *(i)* understand Brazil's standpoint on the meaning of data justice and how stakeholders perceive the concept in practice; *(ii)* explore efficacy and usability of the data justice guidance for policymakers; *(iii)* understand existing systematic, organisational and participant-level barriers that could obstruct the achievability of the guidance; *(iv)* evaluate the role of law and regulation as a positive or negative determinant for the realisation of data justice; *(v)* assess prospective institutional and community-mobilising approaches that could optimise the advancement of data justice for policymakers.

An important note for understanding ITS Rio's activities over the project is clarifying which target audience the research was directed at. The institute focused its efforts on engaging with policymakers for the project outputs. The definition of policymakers used incorporates a wide range of players key to the policymaking process. This means considering scenarios where different stakeholders participate in formulating policy. As in Brazil the institutional arrangements are multi stakeholder in nature, the Institute follows the same understanding proposing collaborations not only with politicians and high level officials, but to collaborate with actors from different groups and sectors, such as, private sector, academical and technical communities and civil society.

At the end, our expected results are to provide local and innovative contributions to support the possible preparation of a preliminary guide for policymakers on how to promote data justice. Thus, in addition to the critical overview of the research stages, we will bring recommendations based on our findings on how to promote positionality awareness, intercultural learning, and higher levels of policy reflexivity among policymakers.

METHODOLOGY

The methodology chosen to prepare this report was shaped by the need to arrive at a comprehensive understanding of a highly complex and dynamic ecosystem. This process leveraged the combined expertise of the key stakeholders engaged during a workshop and through semi-structured interviews, combined with detailed and extensive literature research. Below, we detail the specific methodology behind the interviews and the workshop conducted.

Stakeholder interviews

Semi-structured interviews were held across a broad range of stakeholders to gain insights into their views on the meaning of data justice and how they perceived the pillars in practice. Great care was taken to ensure inclusiveness and diversity, with a geographically diverse array of policymakers representing academia, civil society, governments and other key public institutions and in terms of race and gender.

Considering the increase in COVID-19 and influenza cases in Brazil between 2021-2022, that even impacted ITS Rio personnel, interviews were conducted via digital platforms, in order to follow health precautions. In fact, performing the interviews online made it possible for a wider range of policymakers to participate. **In total, 11 interviews were carried out during the course of the project.**

Each interview lasted from 30 to 45 minutes on average, depending on the interviewee availability. The interviews were recorded, transcribed, translated, and de-identificated, per agreed in the Project Services Agreement. Still, detailed notes were collected and structured in an organized manner that enabled cross-referencing results and facilitated the analysis. Policymakers interviewed were able to express their views freely on different topics related to data justice.

Interview structure was divided in two classes of questions: *(i)* understanding the participant's overall perception of data justice; and *(ii)* gathering a specific perspective on the pillars from the interviewee's professional expertise. To that end, the second class of questions varied from interviewee to interviewee, while the first one was common to all participants. The strategy captured a general overview of how they perceive each of the six pillars, exploring potential alignments.

The contributions made by policymakers interviewed were vital and have been used throughout the report to highlight particularly important observations. All input from the interviews has been used without any personal attribution.

Stakeholder workshop

Another method for obtaining policymakers input was the conduction of a workshop where a variety of topics relevant to data justice were brought up and discussed. As with the interviews, ITS Rio identified participants representing different policymaker groups, including academia, civil

society, different branches of government and private sector. **In total, the workshop had 32 participants**.

The strategy behind the workshop focused on: *(i)* understanding potential **connections** noted by the participants; *(ii)* establishing potential **correlations** between stakeholders experiences and the different pillars; and *(iii)* seeking to unveil **common understandings** of what do the different pillars mean for the policymakers and how they tackle the challenges, considering the Global South political and social context.

In similar fashion to the interviews, the workshop took place via a digital platform and enabled policymakers from the most varied regions of the country to participate. It lasted approximately 2 (two) hours and was structured around three phases: icebreaker activities, group discussion and final open discussion, as described below.

- ■ **Icebreaker**. The initial activity focused on setting the stage for further discussion. Participants introduced themselves briefly and expressed their thoughts on what data fairness means in "up to two words". To ensure that this activity is energetic, the ideas were expressed as a "tweet". Words were captured in a platform to form a "word cloud" summarizing their perspective on data justice.

- ■ **Discussions in groups**. Participants were divided into three groups to cover two pillars each, inspired by the "World Café Method". Members of ITS Rio team were responsible for moderating the discussion within each group. During each round, participants were encouraged to answer two questions on each pillar, they were: "*How to understand this pillar?*" and "*What are the opportunities and challenges to implement it?*". After the two pillars of the specific room were addressed, participants moved on to the next room, where they discussed other two pillars. The dynamic ended when all participants circulated in the 3 rooms and reflected on all six pillars.

- ■ **Open discussion**. Afterwards, participants returned to the main room. The moderators (from ITS Rio) presented a brief summary on what was debated regarding each pillar. Then, participants were able to share their insights or reflections. This final phase of the dynamic was essential in finding common ground and to assess how the dynamic impacted their views on data justice.

The workshop was recorded, transcribed, translated, and de-identified, per agreed in the Project Services Agreement. Anyhow, notes were collected during the event and the interactive materials provided solid observations that facilitated ITS Rio analysis. Outputs from the workshop were vital and have been used in the report. All input from the workshop has been used without any personal attribution.

I. INTERNAL ASSESSMENT

The analysis of the empirical material of the research, after carrying out the interviews and workshop with key policymaker stakeholders and considering the existing literature on the subject, enabled a broader understanding of the Brazilian ecosystem in terms of data justice. The policymakers raised not only ethical and value-base concerns around data governance and the data innovation ecosystems, but also issues related to the basic underlying infrastructure in order to be able to benefit from the current data-driven environment. In other words, material conditions related, for example, to the communities' connectivity to the internet, the availability of data or their quality seem to be a chief concern as there may be an impact on the possibility to create, research, develop, and implement data-driven solutions. Therefore, it appears that while in the Global North the chief focus is on ethical ramifications of data-driven tools and solutions, in the Global South, particularly in Brazil, there is a prior concern in terms of having the underlying infrastructure conditions.

Additionally, the research showcased that the concepts of fairness, equity and non-discrimination may not cover certain aspects of inclusion and diversity that are of significance for the local understanding of data justice. The term inclusion proved to be more appropriate given that it goes beyond merely guaranteeing a balance between different social groups, and seeking historical reparation. This is because the challenges mentioned by participants involving the fairness, identity and participation pillars are understood by policymakers as encompassing not necessarily discriminatory policies, but a wider mix of circumstances, with overlapping socioeconomic elements. Thus, the demand is for more inclusion of both social groups and individuals.

In several instances, what leads to a potential situation of exclusion is the application of a data-driven tool in a context where certain material circumstances are present that create an imbalance, usually as the result of a lack of infrastructure. Two examples may provide an illustration: (i) an artificial intelligence (AI) tool used to optimize the delivery of services to a population may be less than optimal if certain areas of a city are not yep mapped; or (ii) the development of an AI solution for image diagnostics that does not work or is not accurate because it depends on certain protocols or certain specific tools that are not in place or available. This is less a function of a discriminatory or unfair policy, as it is a lack of an inclusive mentality that take into consideration local circumstances.

Moreover, the research indicates that there is a need for an approach towards diversity that moves beyond group diversity, understood within a spectrum of different experiences both local and individual. That is, a diversity that enhances and centralizes the native voice, so that participation occurs in an effective and unmediated fashion.

Thus, through the description and analysis of the material collected during the research, the report concludes that the discussion on data justice in Brazil is an ongoing process, and stakeholders are seeking to consolidate the ecosystem and achieve a higher level of maturity. Furthermore, Brazilian policymakers are concerned with the practical uses and real impact on society of data-driven technologies, particularly artificial intelligence.

II. INTERVIEWS

a. Knowing the voices: Interviewees' demographic information

Understanding video interviews as an opportunity for stakeholders to voice concerns more openly, 11 semi-structured interviews were conducted with Brazilian policymakers. As previously mentioned (see Methodology, stakeholders interviews, above), the selection of policymakers was based on a multistakeholder approach. Thereby, in an attempt to embrace these multiple stakeholders involved in the policymaking process, the interviews were conducted with: three members of the Executive, two members of the Judiciary, one of the Legislative, and five stakeholders including civil society, academia, technical community and the private sector.

The aim was to provide a broad and inclusive perspective, which as a matter of understanding the local dynamic appears fundamental when discussing data justice. Although there were intensive efforts to achieve gender, racial and regional diversity within the selection of participants, due to several factors – mainly structural inequalities that historically make it challenging for some demographic groups to act as policymakers and the diminute number of participants – the attempts may fall short to optimal. Therefore, the group of interviewees was composed of: 7 women and 6 men; 4 non-white women and 8 white people. In addition, considering the Brazilian macro regions[1], the participants were divided into: 5 participants from the Southeast region, 2 from the Midwest region, 2 from the Northeast, 1 from the North and 1 participant from the South region. On education level attainment, 6 participants hold advanced degrees (Masters, PhD, or equivalent formal qualifications) and 5 completed post-secondary school. On a more subjective side, regarding the extent of their familiarity with data and algorithm-related technologies, including formal or informal background/training/education in this area, and experience or interactions with data driven systems: 4 participants responded moderately familiar, 3 feel very familiar and 4 are extremely familiar. All

---

[1] The regions of Brazil are the groupings of federation units elaborated by the Brazilian Institute of Geography and Statistics (IBGE), which created five regions: Midwest, Northeast, North, Southeast and South. This division fulfils the purpose of helping statistical interpretations, implementing public management systems and guiding the application of public policies by the federal and state governments. Information available at:https://www.ibge.gov.br/geociencias/organizacao-do-territorio/divisao-regional/15778-divisoes-regionais-do-brasil.html?=&t=o-que-e

participants stated having unlimited access to the internet, digital applications, and computing resources.

b. Six pillars, multiple visions: an analysis based on policymakers' responses

In order to start the debate and pave the way for the specific questions, the interviews started by asking participants to say what data justice means to them. The answers, despite having a common core related to the proper use of data, went in multiple directions, evidencing an ecosystem in the process of maturation. The term "data justice" is not widely used in Brazil, thus, it may be a factor that contributes to this broad view of the matter. However, it became evident that data justice depends a lot on what they apprehend as a parameter of justice.

Therefore, in order to clarify the participants' understanding of each pillar, this report section will be dedicated to elucidating key points that were discussed during the interviews, as well as major cases brought by participants that illustrate the different approaches.

i. Power

When discussing power in the context of data justice, participants highlighted the role of data in guiding and creating public policies. In this sense, before worrying about challenging power inequalities or empowering people, the stakeholders stressed that in the Brazilian scenario it is first necessary to extract data, even to locate where the potential power inequalities are. Therefore, public sector policymakers stressed that care must be taken not to derail or hamper research capacity in an attempt to bring in regulations to promote data justice. One of the participants even made the following remarks:

> "*Data guides how we fight a problem in society. What often happens is that we do not produce data on various issues that are a problem for society as a whole. An example of this is the participation of black people, of black women in the field of technology [there are no official datasets].*"

A participant brought up the case of "Bolsa Família"[2], a Brazilian government's income transfer program. According to the interviewee, the program needs to be fed with a consistent database, as this base greatly impacts the final goal of the public policy. Therefore, if the data does not reflect the reality and/or if there is lack of quality data, the final goal of the policy is affected – which in this case results in vulnerable parcels of the population not receiving welfare support. In addition, the scope of the policy may be impacted as well by the absence of relevant data or data that may not

---

[2] Bolsa Família is a federal program of the National Secretariat for Citizenship Income (Senarc), created in 2003, which contributes to the fight against poverty and inequality in Brazil. Information available at: https://www.gov.br/cidadania/pt-br/acoes-e-programas/outros/bolsa-familia

be cross-checked. Hence, power imbalances are seen as embedded not only in the selection of the relevant categories of data, but also in the collection, treatment and quality of the data available.

ii. Equity

From a linguistic point of view, the term "equity", as with "data justice", encompasses elements that may not necessarily be well encapsulated in the Portuguese translation. Additionally, elements of the local culture may lead to a narrower view of the term. However, based on the idea of fairness, the responses lead to a more cohesive understanding of the pillar. In this regard, by taking the concept of fairness as a starting point, most respondents associated the equity pillar with "participation". Thereby, solutions were aimed at the possibility of vulnerable actors participating in the elaboration of policies and technologies, to create more equitable environments. On the other hand, some participants highlighted that if vulnerable citizens do not have access to more basic things such as decent housing and basic sanitation, it is difficult to think that they would have equity in acceding to data justice. Therefore, what was asked is: how can people with such vulnerability be able to counter possible injustices, if they often cannot understand the benefits and harms of a particular artificial intelligence (AI) system? In this scenario, the role of policymakers is evident, one noted the following:

> "*This brings much greater responsibility to policymakers. Because policymakers need to be even more aware that they will be working with public policies that will impact the lives of millions of people. And there are millions of people who will hardly be able to challenge the effectiveness of these policies*."

iii. Access

The access pillar was reflected in the responses mainly through the idea of transparency. Public sector stakeholders highlighted that there is already a legal theoretical framework in the Brazilian public sector that ensures access and transparency in relation to information. The challenge would be to transport this debate to the application and use of technology by the government.

Despite the regulatory discussion, many responses highlighted a preliminary concern, which is for people to be aware of the existence and implementation of data-driven (usually AI) instruments. In this way, the interviewees indicated that policies need to consider who are the individuals involved, as well as their vulnerabilities and challenges, taking into account all possible communities impacted. Thus, access in the Global South goes through issues of literacy as well as digital literacy, so it is necessary to consider alternative forms of providing access. In addition, the effectiveness of access policies that do not discuss historical reparation for, for instance, racism suffered by the black population in Brazil were questioned.

To illustrate the difficulties involving access, a participant mentioned a social benefit called "Emergency Aid"[3] that the Federal Government offered to vulnerable populations during the COVID-19 pandemic. The program was designed to be digital and automated, so the process was done primarily without human interaction. However, the app provided lacked space to enter identification data other than the "CPF"[4] – a sort of Brazilian social security number –, thus, excluding scores of people from the aid, as there are still instances in which citizens already do not have such identification number. In addition, many were unable to access the Emergency Aid app for lack of connectivity or devices that would support it, ending up without the benefit when they needed the most. To try to reach those people who were unable to register, the government allowed it to also be done by mail, but according to the interviewee, the policy only lasted a month, was not disclosed and impacted few people. This case exemplifies how much the issue of access to data ends up impacting the population.

iv.   Identity

Identity was the pillar that the interviewees discussed the most in their responses, mainly because the debate on facial recognition is on the rise in Brazil and this has been understood in many ways through the lenses of identity and group politics, often focused on police action. Several answers indicated that there are issues of discrimination in different phases of AI projects in general and specially in terms of facial recognition. A significant portion also signalled diversity as one of the main ways to solve identity problems. For public sector respondents, identity involves chiefly access to data, as there is a need for quality data to ensure there are no biases and or to mitigate potential imbalances and impacts in certain groups and populations.

One of the interviewees cited a case that took place in January 2022, where a photo of the American actor Michael B. Jordan appeared in a database of the Ceará Civil Police[5] – used for facial recognition – as a suspect murderer. The presentation of the photo of the actor resulted in the apprehension of a 17-year-old suspect, which further showcases the potential inaccuracy of the tool vis-à-vis the Brazilian population. Other interviewees highlighted as well the risks of facial recognition systems have to the Brazilian population, especially to the black community.

---

[3] For more information on the Emergency Aid: https://www.gov.br/cidadania/pt-br/servicos/auxilio-emergencial.
[4] The "Cadastro de Pessoas Físicas"(Individual Registration) is the taxpayer registry maintained by the Federal Revenue Service of Brazil.
[5] Information available at:
https://cultura.uol.com.br/noticias/45551_ator-michael-b-jordan-aparece-em-foto-de-reconhecimento-de-suspeitos-da-policia-civil-do-ceara.html.

v.   Participation

The pillar of participation was closely understood in terms of inclusion and often discussed in tandem with the pillar of equity. Interviewees noted the importance of an individual not only being the object of a data treatment, but also the agent behind the decision of what is being done with one's own data. The answers indicated a particular concern with this pillar, participants noted that there were two dynamics of participation that were relevant: an international, in terms of where most of the data-driven technologies are being researched and developed; and national, as a matter of ensuring participation in the decision-making process.

Numerous challenges were cited to establish mechanisms that guarantee effective participation and safeguard against detrimental impacts of lack thereof may have. An important difficulty noted is the fact that topics and technical terms related to such technologies (AI mostly amongst them) are not accessible to the general population, resulting in a limited cohort of interested actors getting involved. Thus, education appeared as one of the main tools to ensure participation, as one participant noted: "*I think that a path, without a doubt, is education, it is information, but an education with a language that communicates, that speaks to other people.*" Another also highlighted:

> "*There is no participation without transparency, without access to information. And, on the other hand, as it is an incipient topic and we still have huge gaps in society, in terms of knowledge about it, training processes are very important so that people can participate on an equal footing with one another.*"

Policymakers also brought up some practical dilemmas involving participation. Allowing open-ended participation in the decision-making process may lead to a very large amount of data provided by those involved. There is a difficulty in navigating not only the volume of information, but sometimes also the quality of the data and the actual representativity of those involved. Thus, there are questions on how to listen effectively, how to ensure a plurality of voices, how to empower people who are unable to express themselves in a qualified fashion, as well as to guarantee governance structures. Participants from the public sector were significantly sensitive to such a challenge.

Another problem mentioned, related to participation, is that of local research and development, as there are enormous barriers to economic entry in the universe of data-driven technologies. When policymakers and developers focus only on foreign technologies, the opportunity to build technologies with a local cosmovision may be lost, as noted by one of the participants:

> "*Data sovereignty depends a lot on technological structures and infrastructures that need to be shared better and must necessarily also be controlled at a regional and local level. Not only for economic reasons, not only for creative reasons, but because whenever I just focus on intelligence, on Californian technologies, I fail to create other worldviews, other world techniques, other possibilities with my culture that is not the Californian one.*"

vi.   Knowledge

The pillar of knowledge also benefited with the lenses of inclusion and fairness. The policymakers interviewed understand that technological innovation is a concept under construction, and little has been done to guarantee the production of knowledge in an equitable manner.

For some participants, the concern is that either developers or the government *assume* what communities want, and do not effectively communicate and listen to them, thus failing to evaluate impacts and actual usage.

Knowledge was understood from the standpoint of the possibility of building new arrangements that ensure opportunities for knowledge exchanges, which sometimes may not be a priority from the researcher or developer (often from the Global North), but which may interest local communities. Interviewees noted that to build these new sociotechnical arrangements, there is a need to be creative particularly in terms of distributing information and bringing in not only new questions and problems to solve, but also new avenues of solution under a more diverse and inclusive fashion.

One interviewee cited a local project called "CocôZap" by data_labe[6], which maps through citizen participation the situation on basic sanitation in "*favelas*" (informal irregular settlements). The project takes place in the Maré complex in Rio de Janeiro[7]. CocôZap provides a cell phone number that receives, through WhatsApp, photos, videos and narratives about rubbish and sewage in order to locate and illustrate the challenges of a daily life of inequalities in terms of access to public services. The project is producing a new database with the intention of building diagnostics complementary to the official datasets available. The participant highlighted the project as an example of creativity, in which the accessible language and service – which is showcased by the name of the app ("PoopZap") and the chosen platform, WhatsApp, used by virtually all cellphone users in the country - facilitates the inclusion of vulnerable communities in the generation of knowledge.

c.   Understanding the ecosystem and solutions: Suggestions for the future

Through the policymakers' narrative, one can perceive the level of maturity of the Brazilian "data justice ecosystem". It is possible to identify that the local debate is still focused chiefly on understanding the opportunities and impacts data-driven technologies may bring to the country. Hence, it seems to suggest an environment that is still developing with a space to mature.

In terms of suggestions directed towards promoting a positive agenda in the data justice ecosystem, they steam from three fronts.

---

[6] More information available at: https://cocozap.datalabe.org/sobre/
[7] The Maré Complex is a neighborhood in Rio de Janeiro that comprises a set of 17 communities where around 140,000 people live. Information available at:
https://wikifavelas.com.br/index.php/Complexo_da_Mar%C3%A9

The first seeks to develop the context in which the new data-driven technologies should be researched, developed and implemented, in other words, it understands that data justice is dependent on a broader discussion of the country's socio-economic challenges. In this way, policymakers suggested that there should be an investment in education and digital literacy in order to enable citizens to share in the benefits and mitigate risks of the new technologies. This solution starts from the understanding that the debate about data in Brazil is limited to a small group of participants, even though injustices may be felt by vulnerable individuals outside it. Thus, participants seem to suggest the necessity of making more accessible the language of the debate, translating complex issues to a language that is directed to those affected. The aforementioned "CocôZap" project seems to be an important illustration.

The second category of solutions seem to focus on effectively guaranteeing inclusion and diversity. All participants proposed inclusion as a tool to promote data justice, in line with the concept of social justice They diverge, however, on how to achieve such inclusion. For some participants, diversity and inclusion should start with those that are researching and developing ways to use data. Thus, it would be important to actually be aware of, for instance, how many black women or members of indigenous groups are technologists, programmers, CEOs of data-driven technology companies. The participation of a wide variety of people should be seen as a priority. This includes capacitating and incentivizing people to engage in the ecosystem and assume decision-making positions. In the interviewees' views, diversity at the research, development, besides the decision-making and consulting level are fundamental. Additionally, the importance of the participation of the scientific community in the formulation of public policies was also mentioned, as academics could warn about possible risks, biases and propose solutions.

Furthermore, interviewees suggested that diversity and human rights should be incorporated "by design", from the offset, as there are different forms of discrimination and injustice that can be prevented if identified at an early stage. Therefore, it is necessary to monitor levels of participation and collect indicators on how inclusion is being carried out at all stages.

One of the participants from the academic sector proposed using the concept of "data sovereignty", that is, a sovereignty that is not focused on the state, but rather that ensures that the local community has the final decision on their data. This proposition focuses, above all, on the pillars of power, participation and knowledge, recommending that data be used and democratized from shared technological infrastructures at a regional and local level, not only for economic reasons, but also for innovation.

Some participants questioned whether participation would be sufficient to ensure data fairness. They noted that in the country there are still high levels of poverty and structural racism and discrimination, thus potentially having an impact on exercising such participation effectively. It was

pointed out that policies such as affirmative action, reservation of vacancies for admission to institutions, universities, public tenders, are still minimal. In their view, it was necessary to advance on such debates, otherwise the problem of data justice will not be addressed in any significant way neither, or worse, structural inequality may be reproduced (and automatized) with the use of data-driven technologies.

The third scope of solutions proposed by the interviewees were those aimed at changing the regulatory landscape. Policymakers noted the need for governance structures within both the public and private sectors. Some participants stated that the Brazilian data protection laws are robust, but there still is a need to implement it and to develop the institutional ecosystem in order to promote better governance. In addition, interviewees emphasized that laws may not be full proof guarantees, as it is necessary both a culture that supports and institutional arrangements to enforce them.

One suggestion was to propose institutions that allowed for biases, discrimination, unfair treatment to be identified early on. Thus, it was proposed to have a permanent "evaluation system" which should be available to assess the data-driven technologies and probe for any biases, discriminatory or non-inclusive practices. For the private sector, the suggestion was that corporate structures should be nudged towards carrying out impact assessments and auditing.

In this sense, in order to identify and rectify any injustice, it would be necessary to have constant interaction with any affected population. Some stakeholders suggested that this should be done through easily accessible feedback channels and complaints mechanisms. When technologies are used for criminal prosecution and public security the risks seem to be higher and some interviewees emphasized that there should be constant monitoring and a feedback loop so that assessment and improvement are also permanent. Furthermore, interviewees proposed that the public sector should create oversight mechanisms, that is, establish control bodies that monitor the research, development and implementation of data-driven technologies.

For some stakeholders, in order to discuss any potential regulation, there should be an assessment of context (local and domestic) as "one-size fits all" solutions may not be adequate.

III. WORKSHOP

a. Knowing the voices: Workshop's demographic information

The Workshop adopted a flexible, qualitative research design that allowed the pillars to be explored in depth while reaching a variety of stakeholders across different regions of the country and different demographics. As with the interviews, great care was taken to ensure inclusiveness and diversity, with a wide array of stakeholders. In an attempt to embrace the multiple stakeholders involved in the policymaking process, the following were participants in the workshop: six members of the Executive, seven members of the Judiciary, three of the Legislative, eleven participants who

work in civil society organizations, two academics, and three stakeholders from the private sector. Among these, twenty women and twelve men. In addition, regarding geographic regions, the participants were divided into nineteen participants from the Southeast region, five from the Midwest region, three from the Northeast, one from the north and four participants from the South region.

b. In a tweet, what does data justice mean to policymakers?

As mentioned above, during the ice-breaker dynamics, participants were asked to express their thoughts on what data justice means in "two words", as in a "tweet". Then, words were captured to illustrate a word-cloud. After analyzing this world-cloud, it became clear that the concept of "data justice" is still in a process to be developed among Brazilian policymakers involved with the data ecosystem. The words mentioned by participants, despite being in the same context, provide an open-ended definition of the term. Words such as transparency, accessibility, inclusion, ethics, equality, empowerment, governance and protection, seem to show that there is an aspiration to achieve data justice in a way aligned in principles, but without not necessarily establishing a well-defined vision.

WHAT DOES DATA JUSTICE MEANS TO YOU?

justice gap social justice repair codefinition power protection respect
technical complexity - moral challenge governance empowerment data empowerment
citizenship, equity, free childhoods digital divide combat international disparities
informational self-determination **transparency**
accessibility and knowledge
redistribution **ethics** accessibility
digital-based democracy **equity**
equality inclusive datification **representation**
diversity citizenship authenticity
non-discrimination efficiency
legal security ethics of use equal treatment

c. The pillars' many facets: opportunities, challenges and solutions

This report section focuses on the pillars and how the participants interacted with them during the workshop. The goal was to focus on: *(i)* how the collective **understanding of each pillar** took place, considering the Brazilian ecosystem; *(ii)* the **opportunities and challenges** identified by the participants for each pillar; *(iii)* what **creative and local solutions** the participants collectively built throughout the dynamic.

i. Power

There were multiple concerns in the discussion on power, the general view, however, was that power in the context of data justice relates to the capacity to influence the decision-making process, particularly in terms of partaking in the governance and regulatory regimes. In this sense, the challenges refer to power asymmetries that are reflected not only in the structure of databases and the data selected, but also in the institutional arrangement that selects them. Most of the participants were in agreement that power and equity are directly linked, since the way in which power over data is exercised has a direct impact on the fairness of the use of data.

In addition, public sector policymakers stressed that it is necessary to differentiate when data is used for policy decisions – aimed at the well-being of citizens – and when the private sector uses data for business purposes, with no disclosure on how personal data is being handled. In this context, many defended open data so that the public power can formulate better policies driven by data, in addition to analyzing possible asymmetries of power through them.

ii. Equity

In the context of the workshop, the discussion on equity emphasized transparency as the key to achieving fairness in the data ecosystem. Participants understood that power and equity are linked, and transparency serves a "rope that ties the two pillars together". By understanding the different uses of data, individuals and groups may identify discrepancies, inequalities and act against unfair decisions. Therefore, transparency translates into an opportunity to make data inequality visible and allow for accountability. In this sense, transparency is very close to communication, awareness and empowerment, communicating to those who are affected and those who benefit from certain data systems.

One of the propositions refers to the need to have means of communication, open channels, as well as awareness-raising practices. In terms of raising the population's awareness, participants noted that this should be a two-way practice, as actors of the data ecosystem should themselves also be aware of the needs of the affected population.

iii. Access

Policymakers understood that in order to talk about access to data, it is first necessary to face the digital gap, the "haves" and "have-nots" in terms of connectivity, access to basic infrastructure and devices (e.g. computers, cell phones, etc.). Additionally, the participants remarked that access is also a product of education and digital literacy. Hence, the main challenge cited involving access, was actually the underlying social inequality and lack of opportunities.

Another issue mentioned was the challenge of how to provide digital services to citizens that are not familiar or do not want to depend on digital technology. Stakeholders noted that sometimes aiming to achieve more efficient services – through digitalization – the population may be left without access to the very same services they actually need – be it because they may not have the means to connect digitally, such as broadband, be it because the system does not accommodate for one or other need (it is not geared towards the visually impaired, depends on a biometric that someone may lack, e.g. fingerprints may disappear, etc.).

The great opportunity that participants noted in terms of access refers to connecting this pillar with participation, allowing people that benefit or are impacted by data-driven technologies to provide inputs necessary for the services to function properly.

iv. Identity

Participants interpreted the pillar of identity based on the idea of representativeness and non-discrimination in all stages of projects involving data. Under this interpretation, a major discussion related to Brazilian structural racism and its impacts on data-driven technologies.

As with the interviews, the discussion about the use of facial recognition technologies took up a large part of the dynamics. Many participants advocated for the total banning of these technologies in Brazil, given the high incidence of police violence against the black and vulnerable population, and numerous cases reporting systematic errors in these systems. On the other hand, some stakeholders were more moderate, recognizing the potential risks of the technology, but understanding that with an effective impact assessment, and efforts to mitigate risks, under certain contexts the technology may play a role.

v. Participation

From the dynamics, several participants understood this pillar in terms of the concept of "*participatory design*", that is, conditions of participation that are part of the design of processes and/or products. In this way, participation should empower the user to exercise his/her rights, engaging as an active subject in solving community problems. Therefore, the concept involved embedding the perception of affected/interested groups since the conception of the product.

To this end, a significant portion of the policymakers stressed that there is no participation without knowledge, that is, there is no possibility of inclusion in the decision-making processes if there is no minimum "technical" knowledge to do so. Policymakers highlighted that participation may lead to transparency, to minimize the asymmetries of powers, and even a higher degree of equity and fairness. Thus, participants noted the centrality of participation to achieving the goals of other pillars.

Equally relevant was the notion advanced that to ensure effective participation, diversity and inclusion should act as a guideline. To this end, the participants listed challenges of structural inequality that may impact society in its various spheres, such as sexism, racism, ableism and lgbtqphobia, which end up making it difficult for vulnerable groups to participate in the data ecosystem.

vi.   Knowledge

As well as the access pillar, participants understood "knowledge" through the opportunity to create an accessible language, especially on technical issues, to increase participation in the production of knowledge. In addition, they understood knowledge from the recognition of other cosmologies, that is, other forms of knowledge than institutionalized ones.

One opportunity that stakeholders suggested expanding this pillar would be drawing on inspiration from successful community experiences, such as Wikipedia, and social movements. In addition, they mentioned funding policies for projects that consider minorities as a possibility.

IV.   CONCLUSIONS

The research revealed an ecosystem in the process of maturation. The challenges, questions and solutions brought by policymakers are prior to some issues debated particularly in the Global North. That is, when thinking about data justice, Brazilian stakeholders are concerned not only with ethical ramifications, but above all with how to develop and implement data systems within a context whose infrastructural and socioeconomic situation imposes additional challenges. That is, to be able to think about data access, for example, it is necessary to digitize and structure data, which in most cases represent and even bigger challenge. More than that, structural lack of access to basic rights and services, such as basic sanitation, and sociotechnical infrastructure, such as access to electricity, connectivity, signify a major barrier for researching, developing and implementing data-driven technologies and may even lead to further impacts to all pillars.

Therefore, the construction of the six proposed pillars seems to be on quicksand, as the foundations of what is being discussed are not stable enough to reach the specific questions about the pillars. The overlapping views in terms of their conceptualization of the different pillars seem to be further evidence of the level of maturity of the ecosystem. This can be illustrated by the discussion on the pillar of knowledge. Although the pillar was designed to, among other things, embrace pluralism of knowledge and interrogate the ways in which certain forms of knowledge are prioritized within decision-making relating to data, policymakers were discussing difficulties related to digital literacy – and even literacy itself.

The discussion about equity, identity and participation exposed the Brazilian context and its socioeconomic and political dynamics. Policymakers, by looking at these three pillars through the lenses of the concept of inclusion, reveal the complexity of researching, developing and implementing data driven technologies in a country that still is developing several aspects of its infrastructure and has a significantly unequal society. On the other hand, it also creates an opportunity as it allows for local and creative solutions to appear that may respond to the various layers of adversity. The example of "CocôZap" (mentioned in more detail above) provides a great illustration as, despite countless infrastructural and socioeconomic difficulties, it is possible for data driven projects that promote positive inclusion to emerge. By creating a communication and reporting channel on WhatsApp – a platform widely used – data_labe developed a project that brings people closer to technology. In addition to the database, monthly meetings are held with residents, schools, health centers and residents' associations, in order to disseminate the complaint channel and build a permanent debate on sanitary issues in the neighborhood. Accordingly, it showcases a horizon of possibilities for the creation of inclusive and even more legitimate policies and solutions, based on the narratives of those who experience the day-to-day difficulties that the technology aims to solve.

Therefore, the different infrastructural problems do not represent a barrier to the discussion of data justice in Brazil. They represent an opportunity to explore new solutions suited to the local context. Based on inputs from Brazilian stakeholders, the guide intended to support policymakers should add a lense in terms of inclusion. Moreover, the question of infrastructure shortcomings should also be considered as they may impact all pillars of data justice. The combination of these elements – inclusion and concern with infrastructure – results in the possibility of emerging new questions, and above all, better solutions to combat possible injustices in the data ecosystem.