Final Report
Digital Rights Foundation
March 13, 2022

Internal Assessment

The team noted some issues with methodology, particularly that both interviews and workshops were structured around the 6 pillars of data justice it took a lot of time explaining each pillar and meant that for large chunks of the interviews the interviewees were talking. Interestingly, the term data justice was not familiar to many who were interviewed or participants in the workshops which made approaching the discussion more complicated than conducting a workshop on more popularised terms such as privacy, data protection or security. It was obvious to us that many, particularly during expert interviews, did not fully grasp the concept of data justice as a distinct concept.

The participants of workshops felt that the format of asking questions and conducting exercises in the absence of answers was not satisfactory. As facilitators who work in an organisation that does provide services such as our 'cyber harassment helpline' and digital security trainings we were able to pivot these queries by providing resources. The feedback, communicated during and at the end of the workshops, made us reflect on terms of engagement with marginalised communities who were generous with their time. While we want to avoid making the terms of engagement with these communities transactional, however thought should be given to ensure that this interaction is mutually beneficial for both parties involved. As a way of ensuring this, DRF made a commitment to share the findings of the research in an accessible manner—in the form of a short written summary or a small video/podcast that will allow the participants to understand how their insights were used. This is perhaps something that should be implemented project-wide by other partners as well.

Another major hurdle was the fact that the project—including the primary literature and preparatory material—had been conceptualised in English which resulted in translation of concepts rather than re-conceptualisation in other languages. We felt that this was a challenge for two reasons:

1. Language informs thought and conceptualisation of data justice, and can often limit thought as well. We used terms such as "انصاف" (loosely translating as justice in Urdu) however there was no corresponding word for data. A larger discussion on language is warranted as the project moves on. Perhaps there is wisdom in dedicating some time exclusively to language in the future.
2. These observations have been noted in the Contextual Takeaways section as well but it bears pointing out that these challenges inhere when projects

conceptualised elsewhere are *adopted* in a local context. Given the short-term nature of the research there wasn't enough room to *reconceptualise* but simply translate. These issues were not immediately obvious to us as well, and was never directly brought up during the expert interviews, but became glaringly obvious during the workshops.

The observations about language point to limitations, however the process of arriving at this conclusion was valuable in and of itself. This is something we would want to work on, not simply building a vocabulary but rather drawing on existing language that can be applied to work on technologies and data justice.

Another takeaway from the methodology was that conducting research on a concept that isn't popularised or readily understood meant that as interviewers we had to take explain the concepts a little which at times did influence the answers that people gave. Most interviewees tended to agree when pillars were put to them, without much pushback, and only elaborate when nudged by the interviewers that they had the option to disagree.

Demographic Questionnaire

Interviews

Interviews were conducted of 6 women, 1 transgender woman, 1 transgender man, and 2 men. These interviews were conducted over Zoom.

Gender	Category	Country	Education	Familiarity	Internet Access
Woman	Developer	Pakistan	Completed post-secondary school	Extremely familiar	Unlimited access
Woman	Policymaker	Pakistan	Completed post-secondary school	Moderately familiar	Unlimited access
Man	Policy expert	Pakistan	Completed post-secondary school	Moderately familiar	Unlimited access
Trans man	Policy expert	Pakistan	Completed post-	Moderately familiar	Unlimited access

			secondary school		
Trans woman	Policy expert	Pakistan	Completed post-secondary school	Very familiar	Unlimited access
Man	Policymaker/ Developer	Pakistan	Completed post-secondary school	Extremely familiar	Unlimited access
Woman	Policy expert	Pakistan	Completed post-secondary school	Extremely familiar	Unlimited access
Woman	Policymaker/ Developer	Pakistan	Completed post-secondary school	Extremely familiar	Unlimited access
Woman	Policy expert/ Developer	Pakistan	Completed post-secondary school	Moderately familiar	Unlimited access
Woman	Policy expert	Pakistan	Completed post-secondary school	Moderately familiar	Unlimited access

Workshops

For this research, DRF conducted two workshops. The first workshop consisted of members of the transgender community and had 11 participants. The second workshop had 19 participants consisting of a mixed group of men and women. Both workshops were conducted over Zoom due to the spread of the Covid-19 pandemic.

Workshop 1

#	Gender	Category	Country	Education	Familiarity	Internet Access
1	Trans woman	Impacted individual	Pakistan	Some post-secondary school	Slightly familiar	Moderate access
2	Trans woman	Impacted individual	Pakistan	Completed post-	Moderately familiar	Unlimited access

				secondary school		
3	Trans woman	Impacted individual	Pakistan	Completed post-secondary school	Moderately familiar	Very limited access
4	Trans woman	Impacted individual	Pakistan	Completed post-secondary school	Slightly familiar	Unlimited access
5	Trans woman	Impacted individual	Pakistan	Completed post-secondary school	Very familiar	Unlimited access
6	Trans woman	Impacted individual	Pakistan	Some secondary school	Not familiar at all	Very limited access
7	Trans man	Impacted individual	Pakistan	Completed post-secondary school	Slightly familiar	Moderate access
8	Trans man	Impacted individual	Pakistan	Some secondary school	Not familiar at all	Very limited access
9	Trans woman	Impacted individual	Pakistan	Completed primary school	Not familiar at all	Moderate access
10	Trans woman	Impacted individual	Pakistan	Completed post-secondary school	Moderately familiar	Unlimited access
11	Trans woman	Impacted individual	Pakistan	Completed post-secondary school	Very familiar	Unlimited access

Workshop 2

#	Gender	Category	Country	Education	Familiarity	Internet Access
1	Woman	Impacted individual	Pakistan	Completed post-secondary school	Moderately familiar	Moderate access

2	Woman	Impacted individual	Pakistan	Completed post-secondary school	Moderately familiar	Unlimited access
3	Man	Impacted individual	Pakistan	Some post-secondary school	Not familiar at all	Moderate limited
4	Woman	Impacted individual	Pakistan	Completed post-secondary school	Slightly familiar	Moderate access
5	Woman	Impacted individual	Pakistan	Completed post-secondary school	Not familiar at all	Unlimited access
6	Woman	Impacted individual	Pakistan	Completed post-secondary school	Not familiar at all	Unlimited access
7	Woman	Impacted individual	Pakistan	Completed post-secondary school	Moderately familiar	Moderate access
8	Woman	Impacted individual	Pakistan	Completed post-secondary school	Slightly familiar	Unlimited access
9	Man	Impacted individual	Pakistan	Completed post-secondary school	Moderately familiar	Unlimited access
10	Woman	Impacted individual	Pakistan	Completed post-secondary school	Slightly familiar	Moderate access
11	Woman	Impacted individual	Pakistan	Completed post-secondary school	Moderately familiar	Unlimited access
12	Woman	Impacted individual	Pakistan	Completed secondary school	Slightly familiar	Unlimited access
13	Woman	Impacted individual	Pakistan	Completed post-secondary school	Not familiar at all	Unlimited access

14	Woman	Impacted individual	Pakistan	Completed post-secondary school	Very familiar	Unlimited access
15	Woman	Impacted individual	Pakistan	Completed post-secondary school	Not familiar at all	Unlimited access
16	Woman	Impacted individual	Pakistan	Completed post-secondary school	Moderately familiar	Unlimited access
17	Man	Developer	Pakistan	Completed post-secondary school	Very familiar	Unlimited access
18	Woman	Impacted individual	Pakistan	Completed post-secondary school	Very familiar	Moderate access
19	Woman	Impacted individual	Pakistan	Completed post-secondary school	Moderately familiar	Unlimited access

Contextual Takeaways

Almost everyone engaged for the research said they were unfamiliar with the term data justice or did not understand where it stood in relation to concepts such as privacy and data security. There were many reasons for this and we wanted to ground these terms in the local context. Based on conversations for this project as well as our own understanding, we have translated key terms in Urdu alongside notes on how the translations were at times inadequate.

English	Urdu	Comments
Justice	انصاف (insaf)	The word insaf resonated with many and was useful in explaining a wider understanding data justice beyond just data protection legislation and laws. Insaf is as much as about a social understanding of justice as it is about formal machinery of the state in order to attain justice.
Data	ڈیٹا (data)	This was the most tricky to translate and we ended up on data as the word we would use in our conversations. Urdu is a fluid language

	اعداد و شمار (Aadad-o-shumar)	(historically known as a collection of languages such as Arabic and Farsi) and does contain a lot of English words. The phrase aadad-o-shumar as a possible translation to data was suggested by many but it deemed to work more in the context of statistics or a collection of data which would be restrictive in our context.
Data Justice	ڈیٹا انصاف (data insaf)	Albeit somewhat awkward in phrasing, we decided to use the phrase "data insaf" in conjunction with data justice in our discussions to connect the concept of justice to our conversations on data.
Data self	ڈیٹا خاکہ (data kakha)	In order to explain, particularly to impacted communities, the importance of data and its impact on their lives, the concept of the data self was used. It was quite effective in explaining how the data self can be distinct from their self-identity.
Data exploitation	ڈیٹا استحصال (data istisaal)	The translation was effective in communicating the meaning.
Power	طاقت (taqat)	The translation was effective in communicating the meaning.
Equity		There seemed to be no corresponding work for equity.
Access	رسائی (rasai)	While useful in explaining access individuals and institutions had to data, it was not as instructive in highlighting the bottom-up approach. That had to be explained in other ways, particularly in relation to participation.
Identity	شناخت (shenakht)	The translation was effective in communicating the meaning.
Participation	شرکت (shirkat)	The translation was effective in communicating the meaning.
Knowledge	علم (ilm) معلومات (maloomaat) شعور (shahoor)	The last pillar was difficult to explain in this context, there seemed to be no exact translation but a combination of these words were used according to the context.
Accountability	احتساب (ehtisab)	The translation was effective in communicating the meaning.
Autonomous	خود مختار (khudmuktar)	The translation was effective in communicating the meaning.

Many of these words such as insaf/justice and accountability/ehtisab already exist in the political milieu, but have not been used in the context of data and data justice. The ruling political party at the time of writing this report is the Pakistan Tehrek-e-Insaf, with justice/insaf being its primary ideology. Thus to invoke these concepts in the context of data and technologies was subversive and entirely new for many.

Lastly, terms such as Machine Learning (ML), algorithms, artificial intelligence (AI) did not have any direct obvious Urdu translations. Industry experts interviewed for this project informed us that work within the industry happens in English, and that they were taught subjects on technologies and data in English as well. The literature used in classes was often written outside of Pakistan and the medium of instruction was mixed, with text in English and bilingual verbal instructions.

Interviews

Overall, data justice was not a term that most interviewees responded to and it was only when broken down in terms of pillars was a more in-depth discussion possible. Initially at least, many related data justice directly to privacy and data protection. There was a lot of discussion about data protection laws and how the legal system can play a part in ensuring data justice. Those experts more familiar with research and use of data for research were able to conceptualise data justice beyond the narrow paradigm of simply data privacy.

Those working on policy looked towards Western models, such as the GDPR as benchmarks for talking about data justice. There was very little articulation of data justice concepts beyond these conceptions. For developers, it was pointed out how their work is often in relation to clients located in other countries, so they operate on a more "globalised" understanding of data justice which is predominantly dictated by Western conceptions. This allows for a certain level of abstraction to take place, where interests of external clients can be privileged.

In light of these structural issues, there had been very little work in uncovering an "indigenous" understanding of data justice. These findings made us as researchers reflect on whether it is even useful to search for a purely indigenous understanding in a post-colonial context without providing support to indigenous communities to mobilise around these conceptions. Given the systematic erasure of indigenous identities and top-down classifications over a long period of time, there many marginalised communities themselves aspired to be part of existing data systems and structures by way of inclusion and "mainstreaming".

Technological interventions were not seen as neutral and naturally fair, and there was a fair bit of understanding among all those interviewed that special interventions

need to be made in order to make data more equitable and representative. However much of this labour was done by marginalised communities themselves or civil society. The transgender rights experts interviewed spoke to activists themselves working to make transgender folks part of the ID card system and collecting "missing" data on issues such as gender-based violence. the burden of correcting existing and structural harms was placed on under-resourced and impacted communities themselves.

While there was awareness of power dynamics between impacted communities and policymakers/developers/researchers in data collection and mining for data from vulnerable groups, there was little clarity on how to address this power imbalance. While there were mechanisms such as IRBs in place, they did not focus on data justice per se.

On the question of participatory design and policymaking, all interviewees were agreed that it was important in principle, however there was no clear roadmap for how to ensure meaningful participation. One interviewee astutely pointed out that letting the community decide can often lead to uncomfortable questions about what justice looks like. She gave the example of dowery (Urdu: jahaiz) as a regressive and patriarchal concept, but she pointed out that when designing fintech for communities who practice dowery should developers override the wishes of the community? Overriding the wishes of the community altogether might mean that the community doesn't use the tech altogether. She suggested working with local interlocutors who can work towards shifting attitudes rather than overriding local conceptions of data justice.

Many interviews pointed out that in the context of Pakistan, where data collected by the state is often shrouded under the grab of national security. One interviewee spoke about "red lines" which spoke about the military establishment collecting and using citizen data without any transparency and accountability. This unique context bears highlighting given the political history of the country, and thus any articulation of data justice absent dismantling of these structures would be incomplete.

Lastly, the question of what is seen as data in the first place came up in some interviews. In the context of oral cultures like Pakistan where very few things are achieved and documented, what counts as data is a subjective decision. One interviewee pointed out that institutive data about the city that is not represented in technologies such as Google maps, does that make the data less valid or objective?

Workshops

In the workshops, there was an understanding of technologies and digital data systems as beneficial for marginalised communities. Many pointed out how digital platforms allow for marginalised communities to express themselves and raise issues of social justice. However there was awareness that these platforms often becomes cites of violence and abuse.

In the workshops as well, there was a conflation of data justice with the right to privacy, rather than a larger concept that encompasses privacy. Many expressed the need for robust laws and legislation that could address the harms stemming from data misuse. In the workshop with transgender individuals, lack of trust in institutions was also expressed alongside a desire for these institution to be strengthened in order to provide relief. These articulations, while contradictory at first, speak to the dominance of juridical discourse when it comes to data justice.

Data breaches and lack of accountability of state and private institutions was a major concern for many. The fact that data is collected without adequate safeguards weighed heavy on many and there was little confidence in institutions that collect data to keep its safe. Some participants spoke of data leaks by telecom companies and the harassment that stems for women from this leaking of numbers.

Those working in journalism, research and advocacy spoke about the lack of access to data and information. Data that is available is often incomplete. One participant talked about how data on the issue of forced marriages, primarily targeted towards minor girls from religious minorities, was incomplete or simply not collected.

On the whole, there was articulation of policy demands in the broad sense without granular details, perhaps because these issues are seen as technical issues best left up to policy experts and policymakers.

Lastly, there was strong emphasis, in both workshops, on increasing the awareness of impacted communities regarding data extraction and surveillance practices. Awareness raising, many felt, should also be prescriptive so as to equip members of these communities on how to secure their data and take preventative measures to secure their data.

Feedback on Pillars

POWER:

Developers were much more likely to talk about global power imbalances, whereas marginalised communities spoke about power bested at the local and national level. This in and of itself is interesting as power in the data ecosystem manifests itself in

the form of local actors and institutions, while obscuring the structural and global power dynamics. It also speaks to who gets to interact with these centres of power, developers often get to interact with global tech companies, activists in urban centres are often included as stakeholders (albeit superficially) by global technology companies, and policymakers have the power to be included in these conversations, however marginalised communities rarely get a seat at the table. The intersection of the power pillar with the participation pillar is apparent here.

It was easy for many to understand decision-making power, the fact that they are excluded from decision-making regarding what data is collected, how it is collected, process and used. The powerlessness in this regard was lamented and prescriptive means were articulated in juridical terms. When it came to agenda-setting and ideological power, there was difficulty in generating a discussion. Many were comfortable with language that allowed them to talk about participation within the existing structures of data governance, however could not articulate term beyond that. More imaginary exercises and workshops, which allow participants to understandings beyond that framing are needed. This was done somewhat to ask specific questions

EQUITY:

Historical inequities in terms of representation and uses of data were understood by most participants in the study. What was interesting that many felt that representative data would lead to substantial changes, i.e. inclusion within the system. For instance, inclusion of transgender communities, persons with disabilities and religious/ethnic minorities would lead to more data equity.

When speaking of historical inequities that facilitate data inequities, the link with other social movements aiming to topple these discriminatory institutions and structures became obvious. Data equity and justice are often not linked to social justice questions in Pakistan, rarely on the agenda of even the most progressive of social movements. This is linked to hierarchy of social issues ranked in order of perceived importance given the hierarchy created in terms of issues. Data justice when connected to issues of economic, gender, social and political justice resonates more in terms of articulating a call to action.

ACCESS:

Participants found it difficult to distinguish between the pillars of access and participation. On the issue of process transparency, it was observed that while the principle was seen as necessary by all, there was very little clarity on making the

process of designing technologies and data processing more accessible to marginalised communities.
In order to fully capture the bottom-up ethos of this principle, it was suggested that traditional mechanisms of impact assessments and community-driven approvals to projects should be applied to the development of technologies and data processing. However what this would look like for technologies meant for a global audience, developed in other contexts, was not as clear. Many agreed that there are no incentives present for global tech companies to ask for community consent and input before designing or implementing technology. At most, there can be country-level assessments which are unlikely to trickle-down to the local, community-level.

IDENTITY:

On the issue of measurement equity, it was pointed out in some interviews that the exercise of data collection, primarily by the state, has been a unilateral exercise laden with power imbalances. A dominant method is the national census where the population is categorised into groups. The entire exercise, undertaken by the Pakistan Bureau of Statistics (PBS), often excludes many. The figures regarding the number of transgender folks in the country (estimated to be 10,000, counted for the first time in the 2017 census) is widely understood by gender activists be inaccurate, unable to account for the fluidity of gender identities in the country.

Once understood from a historical data equity lens, this undercounting is not surprising since census operations have their roots in colonial modes of governance predicated on making the local population "countable" as a stepping stone for disciplining. For populations such as the Khawajasira/transgender community who were criminalised under the British colonial regime, it is no wonder that many chose not to participate in these state-led practices. Families often do not declare transgender family members due to taboos around gender. Furthermore, there are significant barriers to capturing gender statistics as many families/households are reluctant to count women, especially when male enumerators are conducting the census due to cultural notions of purdah. The very methodologies of counting data, and determining who counts, is steeped in a patriarchal, colonial history.

Religious minorities are also categorised into arbitrary categories. For instance, those belonging to the Sikh faith are sorted into the "others" category when it comes to data on religion.[1] In 2022, the government is gearing up for what it is calling the country's first-ever "Digital Census" which seeks to collect data through digital tablets instead of a paper-based system. The entire exercise has been budgeted for Rs. 5 billion, however there has been little conversation on issues of data equity.

[1] "Population by Religion," PBS, https://www.pbs.gov.pk/sites/default/files/tables/POPULATION%20BY%20RELIGION.pdf.

Those questioned about this exercise noted that they did not have much hope for progress in course-correcting historical imbalances and under-counting.

PARTICIPATION:
Participation was a fruitful pillar as many workshop participants posited that there must be meaningful participation. Many in interviews expressed the view that "inclusiveness" and "consultations" are now common practices by the government when it comes to drafting legislation and policies, however these consultations do mean that inputs, especially those regarding transparency and accountability by the state, are incorporated. The project speaks of the difference between transformational versus power-preserving participation—the need for the former was apparent in the conversations for this project.
Another interesting trend was the need for participation at the local level, particularly through the local government level. For context, in Pakistan there has been a long Constitutional battle regarding the formation of local governments (which successive governments delaying local government elections), thus the vision for consultations at the local level are part of that political battle as well as an understanding that many of the power imbalances can be upended at the local level as opposed to the provincial or national level.

KNOWLEDGE:
Under this pillar, the oral tradition as opposed to documented forms of knowledge talked about. Modern forms of technologies privilege and process data forms that are documented and digitised. This has caused a bridge between knowledge systems in languages that are likely to be in written form, or are legible in digitised forms.

Some languages do not have much written literature, or are written in forms that cannot be made legible by technologies, which means that knowledge systems of entire races, ethnicities and communities end up being invisibilised in data ecosystems. Furthermore, lack of documentation especially in the Pakistani context where people lack documentation with regards to family origin, land, etc. has implications for which identities are recognised by data and which are not. In the context of the NADRA citizen database, refugees and migrant populations—such as those residents of Afghan-origin or migrants from Bangladesh—often lack documentation and this cannot be part of the database that is essential for accessing basic health, education and welfare services.

Conclusion and Recommendations

- The feedback from communities indicated that there is a dire need to demystify data process and technologies in order foster discussions among marginalised communities that are largely left out of design and policy

discussions. Democratising knowledge regarding technologies through community-driven interventions is crucial.

- While awareness-based interventions was a major demand from communities, most of these interventions focus on individual capacity building which shifts the burden of ensuring data security and justice on individuals and communities themselves. While these interventions are important, particularly in the context of oppressive structures and irresponsive institutions, these interventions should not be the be all and end all of data justice-based actions.
- The role of language in democratising data discourse is important and resources should be extended to developing locally conceptualised and contextually relevant vocabularies.
- As pointed out above, the burden of ensuring representation within data often falls on the shoulders of the impacted communities themselves. Interventions should be developed to build the capacity of communities to not only produce their own data but also ensure control and autonomy over their data.
- Shifting the focus of data justice discussions from the "global" and "national" level to more local structures and nodes of organising where marginalised communities are much more likely to represented.
- Creating specific guidelines for researchers, academics, developers and policymakers to ensure that tangible gains and data justice for impacted communities can be incorporated into the work that these stakeholders are doing.