

# **Exploring the Potential for Data Stewardship in the Migration Space**

---

**June 2022**

## **DIALOGUE ON TECH AND MIGRATION**

**A project of the Migration Strategy  
Group on International Cooperation  
and Development**

## Authors

Astha Kapoor, Suha Mohamed, Shefali Girish, Aapti Institute

*The views expressed in this publication are the views of the author alone and do not necessarily reflect those of the partner institutions.*

## About DoT.Mig

The **DoT.Mig In Brief** paper series is part of the The Dialogue on Tech and Migration, DoT.Mig.

DoT.Mig provides a learning platform to connect the dots between digital technologies and their use and impact on migration policy, as well as connecting relevant stakeholders. The **DoT.Mig In Brief** paper series highlights debates and concepts relevant to navigate the emerging field of Tech and Migration.

DoT.Mig is a forum by the Migration Strategy Group on International Cooperation and Development (MSG). The MSG is an initiative by the German Marshall Fund of the United States, the Bertelsmann Foundation, and the Robert Bosch Stiftung.

# Key Takeaways

- 1.** In the migration and refugee space, traditional data resources are lacking yet there is overcollection of personal data. This presents a double edged-sword, making the data landscape fairly complex.
- 2.** Whether data is lacking or abundant, migrants and refugees need greater participation in decision-making with respect to their data rights.
- 3.** Data can be regarded as both a tool of exploitation and empowerment. While increasingly deployed to surveil and control migrant and refugee populations, it also has positive uses, examples of which are coming up globally.
- 4.** New models of data stewardship support bottom-up, community-driven approaches to data governance that can bring migrant and refugee communities together to deliberate on data-related decisions, prioritize their needs and concerns, and negotiate with authorities as needed.
- 5.** Stewards can play the role of intermediaries and be equipped to spread awareness and safeguard digital rights while also creating opportunities for migrant and refugee communities to mobilize and have collective negotiating power around data.
- 6.** Stewards can be designed and structured in different models, such as data cooperatives, data collaboratives, or data exchanges.
- 7.** Civil society organizations that work with migrants and refugees should consider ways in which they can build capacities of the community to evolve models of data stewardship as an extension of their other rights-based work.
- 8.** Data stewardship is an evolving space where political will, policy environment, and international agreements will have to align in order to achieve an enabling environment for data stewardship for refugees and migrants at scale.

# Introduction

Available and accessible data on migrant and refugee populations is critical to providing timely support and helping policymakers, civil society organizations, and community groups design refugee sensitive policies and interventions. This data can come from a variety of sources and be characterized as either personal (relating to individual/community attributes such as name, location, address, ID) or non-personal (de-identified personal/anonymized data or data that does not relate to an individual). Captured from both traditional (official statistics and surveys) and non-traditional sources (social media data, call detail records, satellite imagery), it can be used to improve international migration governance.

Most data on refugee and migrant groups is collected passively through the use of social platforms and through digital traces left behind when using phones and other technologies. It is also collected by governments through official channels when refugees or migrants provide their data in order to register for protection, asylum, or a visa. At either of these junctures, migrants have limited to no visibility on how corporations or states use this data. Of particular concern is how this data can be used covertly by governments to paint migrants or refugees as 'security threats' or to build algorithmic systems — which may render autonomous decisions on asylum

applications and offer applicants no recourse or mechanisms for accountability. Therefore, while big data has been lauded as a silver bullet when used to leverage de-identified digital traces on migrant and refugee populations, it also presents a host of ethical and privacy concerns. Without governance mechanisms, there remain possibilities for misuse of this data by both state and non-state actors, threatening the safety, health, and wellbeing of migrant populations.

In this context, data can be regarded as a tool of both exploitation and empowerment as it is increasingly deployed to surveil and control migrant and refugee populations globally. However, new models of data stewardship can empower communities by providing the advice, guidance, and support necessary to navigate the data economy and use data meaningfully to create value and prevent harm. Stewards play the role of intermediaries and are equipped to both spread awareness and safeguard digital rights. Additionally, they can create opportunities for migrant and refugee communities to mobilize and have collective negotiating power around data. The following text unpacks how stewardship can be considered as a pathway for more responsible data collection and governance in ways that empower and enhance the agency of migrants and refugees.

# 1. What does data stewardship mean?

Data stewardship is a broad paradigm that explores the potential legal, technical, and social structures that can unlock data for societal value while providing individuals and communities with greater control, transparency, and ability to make informed decisions using their data. The current data economy is characterized by the many inequities present in the offline world — key among these is that data is siloed by companies and governments. This status quo offers individuals and companies little insight into how they can derive value from this data and, furthermore, leaves them powerless with little to no bargaining power in the face of data harm.

Stewards can be structured with varying levels of community participation based on different types of data. For instance, **data cooperatives** enable members to deliberate and decide on questions of collection, access, and use of data, often through direct voting. **Data trusts** empower a board of trustees with a fiduciary responsibility to represent the best interests of the community in making data decisions. **Data collaboratives** involve different stakeholders that co-define rules for data. Finally, **personal data stores** provide individuals with granular control and consent dashboard-like interfaces. These structures serve as intermediaries, enabling refugees and migrant communities to build trust in institutions to safeguard their data rights.

# 2. What data-related challenges exist in the current migration data landscape?

As previously mentioned, a lack of data sources on refugees and migrants makes it challenging to develop adequate policies. For example, UNHCR has age information for only 56 percent of the refugee population. This renders invisible the risks and vulnerabilities refugees and

migrants face from decision-makers who, in turn, are unable to design responsive policies. Incomplete information not only impacts offline decisions but, increasingly, AI-driven ones which are linked to existing data sets and have the potential for serious harm if the quantity

and quality of data is suspect. For instance, [7,000 students](#) were deported from the United Kingdom because of a faulty English language testing program. While [not all data driven decision-making is problematic](#), there is a need to address lack of data which can result in ad-hoc decision-making. Poorly timed responses can lead to migrants and refugees being left excluded and increasingly vulnerable.

Simultaneously, refugees and migrants are often subject to data extractive relationships with governments and private companies that knowingly or unknowingly follow a [techno-solutionist approach](#) and collect significant personal, biometric, and mobility data about them. Along with this lack of information, the inherent power imbalance between migrants and refugees and the authorities makes them unable to question or resist this demand for data, even though they may realize that not all the data collected is required or helpful. Of course, states are empowered to collect data on non-citizens in order to grant visa applications or during border checks, but they must also respect basic human rights and adopt a [right-based approach to data](#), especially since the [right to life is intrinsically connected to the right to privacy](#) in many jurisdictions.

This is especially important to understand as there are examples of more insidious deployment of technology through firms like [Palantir](#), which mine data without explicit consent and share it with immigration authorities. In all of this, refugees and migrants have little to no say on what data is being collected and how it is being used. In situations like these, stewards could play two roles to start to rebalance power. First, they could help sensitize and articulate data harms to refugee and migrant communities. Second, stewards who already possess data on migrants, particularly civil society organizations, can act

as bulwarks against extractive data practices and practice more representative or delegated consent processes when sharing of this data has to take place. *Where active decision-making around data usage and sharing may be burdensome for refugees or migrant communities to take part in, stewards that are structured to have a duty of care can represent these interests.* Through pre-negotiated governance mechanisms and consultations with migrant groups, a steward will be best placed to exercise the responsibility for conditionally sharing data for a specified purpose and with technical safeguards in place.

Considering the double-edged sword of poor data resources on one hand and overcollection and surveillance on the other, the data landscape for refugees and migrants remains fairly complex with harms at both the individual and collective level — there is a need to reimagine refugee and migrant data rights such that individuals and communities are protected and empowered.

#### **How does data stewardship solve these challenges?**

Whether data is lacking or abundant, it is clear that migrants and refugees need greater participation in decision-making with respect to their data rights. Data stewardship offers mechanisms for bottom-up, community-driven approaches to data governance that can bring migrant and refugee communities together to deliberate on data-related decisions, prioritize their needs and concerns, and negotiate with authorities as needed. Implementing stewardship can take a variety of forms and requires the support of top-down regulation or legislation and engagement at the grassroots level with data-holding organizations that work closely with refugees and migrants to explore the potential for instituting a range of mechanisms.

## **3. How do stewards ensure a participatory approach and enhance agency for migrants?**

Models of stewardship, when structured to be community led, are by design participatory — engaging communities on decision-making around data collection, processing, and sharing. They also enable migrants to collectivize their data journeys and negotiate with technology platforms and, going forward, with governments as needed. The negotiation powers of the data steward are anchored in being a representative of the refugee and migrant community and being legally and otherwise empowered to represent community interests. The level of participation in each model will vary depending on the needs and choice of the community and how much they would like to engage — for instance, data cooperatives may have members vote on every decision, whereas data trusts provide members with the option of delegation to and empowerment of trustees.

The governance of stewardship models is anchored in the idea of fiduciary responsibility. Whether through cooperatives, trusts, or collaboratives, this responsibility is enforced through organizational structure or contractual obligation. For instance, if an entity is registered as a cooperative, it is codified to have a certain responsibility to its members. Similarly, trust law gives trustees a fiduciary responsibility to its members. In other cases, the terms of responsibility can be defined through contracts. These mechanisms ensure trust in the process of stewardship

and incentivize people to participate by demonstrating the value of data and process.

There is a broader question of incentives for all stakeholders. Why would migrants and refugees want to sign on to yet another intermediary and how would these institutions garner the trust of the community? Why would governments be willing to negotiate with data stewards? Why would private sector companies that provide services to migrant and refugee populations engage with data stewards? These are evolving questions but we are seeing the use of data stewards such as data cooperatives. As an example, [Drivers Seat](#) works with gig workers, to collect data to enhance their own income, to negotiate with app companies such as Uber, Lyft, and Grab. Simultaneously, driver data is made available to municipalities to facilitate evidence-based decision-making. This model demonstrates ways to build incentives for all stakeholders and can be applied to refugees and migrants as well. *That said, it is important to acknowledge that this is an evolving space where political will, policy environment, and international agreements will have to align to create an enabling environment for data stewardship for refugees and migrants at scale.*

Further, there is no one size fits all approach when it comes to data stewardship for refugees and migrants. There may be circumstances, such as the ongoing [Ukraine crisis](#), where the deployment of data has been shown to

be helpful to refugees — data wallets are being used to store IDs and other sensitive information while granular data made available to the government enables action. This is a ripe case for data stewardship mechanisms, which should be made available before the crisis. [Elsewhere, the United Nations refugee agency](#) shared improperly collected data on Rohingyas with the Bangladesh government which, in turn, shared it with Myanmar to verify people for

possible repatriation — mechanisms of data stewardship could prevent this kind of misuse and resultant harm. Both situations described above need models of stewardship, on one hand to facilitate the movement of refugees and allow governments to better understand the ongoing situation and respond accordingly, on the other to safeguard the privacy of refugees and migrants and ensure that personal data is not revealed without their consent.

## 4. What are the pilot projects or use cases in this context?

There are several use cases for data cooperatives for migrants and refugees. In the case of [immigrant health](#), data on access to health services is limited, siloed, and under-utilized. A health data cooperative may allow migrants to contribute, store, and manage their health-related information and deploy it for research and access to commercial goods and government support. Similar use cases may be seen in the context of employment and access to credit. The International Labour Organization (ILO) [already recommends](#) a cooperative structure for [refugees and migrants](#) to improve access to markets, livelihoods, and housing — and it is natural to explore the application of these structures to the data economy.

These models may be most relevant to explore in the context of internal migrant communities or long-term refugee groups. In India, for instance, there are periodic waves of migration every six to eight months when

workers travel from rural parts of the country to the city for economic opportunities. In March 2020, at the peak of the Covid-19 pandemic, statewide lockdowns and nationwide transport restrictions imposed a significant burden on this community. [The lack of available data](#) from the public sector exacerbated the situation and further hindered interventions required to plan transport services and provide basic relief. In response to this gap, civil society organization Jan Sahas developed the [Migrants Resilience Collaborative \(MRC\)](#) in collaboration with other non-profit, philanthropic, and private sector actors. Jan Sahas plays the role of a steward in this instance and makes available aggregated data from over 10 million migrant families for timely interventions. Decision-making at the collaborative takes place through a steering committee model and includes representation from former migrant workers along with other philanthropic partners and research organizations. The MRC model is demonstrative of how stewardship may be conceived and in

which scenarios it may be most practical to implement.

In terms of models, given the legal and jurisdictional complexities of the refugee and migrant landscape — as well as the new regulations with regard to data protection — models of data stewardship are

still evolving. There are efforts, such as [Big Data For Migration](#), which seek to accelerate the ethical use of data to inform migration policies and programs and focus on plugging data gaps for migrant and refugee issues — but they do not necessarily focus on participative data governance - a possible next step.

## 5. What debates about data stewardship are relevant to the migration space?

As mentioned above, data stewardship is, in general, a new concept and especially novel in its application to migrants and refugees. Therefore, some debates need to be resolved in this context. First is the form of data stewardship relevant for migrants and refugees and the specific use cases for each model. Debates include whether collaboratives are more effective in unlocking public value of data; whether cooperatives should be deployed to help migrants negotiate better at the collective level; and whether personal data stores safeguard individual data rights. There are also questions about how to demonstrate value to migrants and refugees such that they are willing to engage with data stewards, so that managing and safeguarding data rights doesn't become another point of confusion in an already overwhelmed and precarious community. Another big question which plagues data stewardship across sectors is that of financial sustainability, and the ways in which it can be achieved — it is clear data stewards should be structured to prevent data harms, but are there models for financial

sustainability which can be deployed to ensure stewards remain true to their purpose? More and more states are using big data analysis to make sense of, control, and manage the movement of people. However, the design, testing, and deployment of AI tools for [migration often view people in isolation of their rights](#) — this belief can insert dangerous bias in the use of AI and harm migrants and refugees. There is a lack of regulation as most countries' AI policies are still evolving — therefore, the use of AI currently exists without adequate oversight. In this context, the use of AI-driven technologies requires greater [collaboration with refugees and migrants](#) to build ethical models that safeguard refugee and migrant interests — this will help bring the conversation out of mere legal compliance and structure it for the well-being of the community.

Models of data stewardship can facilitate this but need to be tested with regard to the specific issues of migrants and refugees and answer the complex questions of incentives, value, sustainability, and scale.

## 6. What are the insights for policy stakeholders to advance the discussion on data stewards for refugees and migrants?

To implement data stewardship, multiple stakeholders need to come together, including international cooperation agencies, policymakers, and civil society organizations.

Inter-agency collaboratives need to consider how data rights can be mainstreamed in the conversation on refugees and migrants. Efforts such as Global Migration Group are invested in [‘data and research’](#) but these efforts need to move beyond insights derived from migration data to the data rights of migrants. It is imperative to advocate for a special category of migrant data rights to be inserted into refugee policy and implementation documents. Alexander Beck, the UNHCR’s senior data protection officer highlights [“Data protection is part and parcel of refugee protection.”](#) The Office of the United Nations High Commissioner for Human Rights (OHCHR) has also issued guidelines for [a human rights approach to data](#). These efforts need to be amplified and at the forefront of every discussion on migrant and refugee rights.

With policy makers, the task of inserting data rights of migrants is more expansive. Given that data is not a vertical issue but something that runs across asylum seeking, social media, healthcare, and employment — it needs to be embedded across different agendas pertaining to refugee and migrant issues. To

this end, policymakers must develop policies and regulations that enable data stewardship such as data portability, interoperability, and the ability to delegate consent. They should also recognize models of stewardship such as cooperatives, trusts, and collaboratives such that they can be implemented and regulated.

**Civil society organizations working with migrants and refugees should begin to consider ways in which they can build capacities of the community to evolve models of data stewardship as an extension of other rights oriented work.**

For this, civil society organizations will have to build a certain data consciousness, engage with data experts to understand points of collection and use along with related harms and opportunity, and communicate these to migrant and refugee communities. Thereafter, it’s critical to consult with and co-design ideal mechanisms of data stewardship that can help achieve data goals (such as minimization) for the community. However, there are outstanding challenges of funding, stakeholder buy-in, and incentives which need to be resolved through a multi-stakeholder approach that embeds data stewardship across organizations working on refugee and migrant issues.

In summary, data stewardship is evolving — there is a need to understand its use and implications in the migration and

refugee space which can happen only through investment in pilots, policies, and partnerships.

