Data for Good, What Is It Good For?: Challenges, Opportunities, and Data Innovation in Service of Refugees

Nicole Smith¹ Muhammed Idris^{2,3} Friederike Schüür^{2,4} Rita Ko¹

¹The Hive, United States of America for United Nations High Commissioner for Refugees, District of Columbia, United States of America,

²Data Advisory Board, The Hive, United States of America for United Nations High Commissioner for Refugees, District of Columbia, United States of America,

³Morehouse School of Medicine, Atlanta, Georgia, United States of America,

⁴Jain Family Institute, New York City, New York, United States of America

Published on: Jan 27, 2022

DOI: <https://doi.org/10.1162/99608f92.a6dbaef3>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

With 82.4 million forcibly displaced people, we need new approaches to the global refugee crisis. The Hive, the innovation lab at USA for UNHCR, uses data, machine learning (ML), and other emerging technologies to improve lives for refugees in coordination and collaboration with UNHCR (United Nations High Commissioner for Refugees), known as the UN Refugee Agency. We outline five challenges in successfully leveraging data and emerging technologies in the humanitarian space that tend to be overlooked and share the Hive's approach and evolution to tackling these challenges. From assembling the right team and finding the right partners to inclusive and impactful data innovation, the Hive has worked to apply industry techniques to the nonprofit sector since 2015. We hope that our insights can help guide data innovation efforts at other organizations in the humanitarian space.

Keywords: data science, innovation, refugees, humanitarian, nonprofit

Media Summary

USA for UNHCR envisions a world without refugees. With 82.4 million forcibly displaced people, addressing the global refugee crisis will require unique and novel tools. The Hive, the innovation lab at USA for UNHCR, uses data, machine learning (ML), and other emerging technologies to improve lives for refugees, working closely with UNHCR (United Nations High Commissioner for Refugees), known as the UN Refugee Agency. The Hive has worked to apply industry techniques to the nonprofit sector since 2015. We outline five challenges in successfully leveraging data and emerging technologies in the humanitarian space that tend to be overlooked and share the Hive's approach and evolution to tackling these challenges, specifically (1) investing in data infrastructure, systems, and analytics; (2) creating and leveraging a network of peers and experts; (3) prioritizing reciprocal partnerships with stakeholders and beneficiaries; (4) accessing relevant and vital data; and (5) evaluating change beyond traditional monitoring and evaluation methodologies. We hope that our insights can help guide data innovation efforts at other organizations in the humanitarian space.

1. Introduction

The sheer scale of humanitarian crises around the world is daunting. According to the UNHCR (United Nations High Commissioner for Refugees), known as the UN Refugee Agency, there are over 82.4 million refugees, asylum seekers, and internally displaced persons of concern around the world, of which an estimated 42% are children (UNHCR, the UN Refugee Agency, 2021b). Current trends suggest that the number of those displaced by violent conflicts, natural disasters, and economic crises will continue to increase with the rise of protracted refugee situations (USA for UNHCR, 2020). On the other hand, voluntary funding for refugee assistance has not kept up. UNHCR has had a budget shortfall of on average \$3.5 billion since 2016 (UNHCR,

the UN Refugee Agency, 2021e), and despite increased exposure of the plight of those displaced, the Refugee Investment Network has found that only 1% of available grant-based philanthropy has been allocated to the refugee space (Kluge et al., 2018). The stark reality is that there are 17% more refugees than 5 years ago, and funding has not kept pace; as such there is a need for innovative, scalable solutions. In this article, we present the work of the Hive—the innovation lab responsible for bringing data science, machine learning, and new technologies to USA for UNHCR—as a case study and draw upon our experiences to illustrate five challenge areas that we believe are often overlooked.

The magnitude of the refugee crisis has continued to increase steadily over the last decade, with nearly 1 out of every 97 people displaced. A refugee journey typically contains four consecutive stages, each of which presents unique challenges and opportunities for innovative solutions as illustrated in Figure 1. The first stage includes both trying to live in the midst of a crisis and planning to flee. Families are often forced to decide where and when to flee with little time and limited resources. The second stage involves the journey heading to a known or unknown destination, where individuals and families may be displaced internally or seeking refuge in neighboring countries. This journey in motion can be as dangerous, with many opportunities to be taken advantage of physically and financially. In the third stage, those who were able to flee may reside in refugee camps, informal settlements, or urban areas. These refugees are often in protracted situations, where mass displacement affects a country for more than 5 years, and spend decades living in refugee camps. The final stage involves resettlement, but only a fraction of those displaced find a permanent solution, with a handful returning to their country of origin and less than 1% resettled by nation states (UNHCR, the UN Refugee Agency, 2021c). UNHCR supports the displaced at every point in their journey, and the Hive is looking to prototype solutions that improve the livelihoods of refugees both domestically and globally.

Figure 1. The four stages of the refugee journey.

Over the last decade, we’ve seen a proliferation of interest in the use of data and emerging technologies, including machine learning (ML) and artificial intelligence (AI) in the humanitarian sector broadly. Within the context of refugees, we’ve seen many practical applications of these technologies across all stages of the refugee journey, from fleeing violent situations to resettlement. For example, we’ve seen potential in the use of cutting-edge technologies, including blockchain technologies for [global digital identity solutions](#), processing [cash transfers](#), and [satellite imagery to map refugee camps and settlements](#). We’ve also seen innovative solutions for [translation](#) and [service delivery](#) that leverage machine learning to connect refugees with relevant, tailored resources and services, while providing humanitarian organizations with technology platforms to scale their service offerings. Some data innovations are limited in scope; for example, algorithms to aid policymakers with resettlement by [matching refugees to the optimal city for employment](#). Others stand to benefit refugees throughout every stage of their journey, such as supporting fundraising through predictive analytics (Genius Awards, 2019).

Though we are far from having solved even just a mere fraction of the challenges refugees face, we’ve also learned quite a bit about what it takes to move from ideation to practical, real-world applications of data science and machine learning that assist refugees. While the allure of data science and machine learning is strong, key challenges around data literacy, access, and quality remain. In what follows, we outline five challenges in successfully leveraging data and emerging technologies in the humanitarian space that tend to be

overlooked and share the Hive’s approach and evolution to tackling these challenges. Our goal is to shed light on the not-so-glamorous aspects that are fundamental to realizing the promise of data science and machine learning for the benefit of the lives of refugees. That is, “data science doesn’t have to be sexy to be impactful” (Lewis, 2020).

2. Challenges in Data Science

2.1. Investing in Data Infrastructure, Systems, and Analytics

Succeeding with data requires solid foundations. There is a lot of excitement about ML and AI, and organizations ask: ‘How may we use AI and ML in service of refugees?’ However, it takes work to get ready for and benefit from AI and ML, as laid out by Rogati in her (now famous) Data Science Hierarchy of Needs (Rogati, 2017) and further expanded by Harrison & O’Neill (2017). Investment in data and engineering and data infrastructure is required to ensure data access and to provide data professionals with the environment and tools to succeed. The application of AI and ML requires investment in data analytics, a discipline often eclipsed by data science. Data analysts are the ‘first line of defense’ when it comes to data quality, identifying issues and working to improve data quality over time. Data analysts are able to derive insight from data quickly and can therefore illustrate the power of data to stakeholders. Their reliance on descriptive statistics, and the transparency of these methods, enables stakeholders to build trust in data-derived insights. As such, the work of data analysts provides the foundation for data literacy and helps foster an organization’s data culture, data-centric decision-making, and quantitative thinking.

It has been challenging for organizations to translate the lessons encapsulated in Rogati’s (2017) AI Hierarchy of Needs into practice. At times, in a hammer-looking-for-nails fashion, AI-based approaches are assumed to provide solutions without sufficient examination of alternatives, for example, analytics-based approaches. Organizations still fail to sufficiently invest in the foundational work. It can be challenging to make the business case to invest in data foundations because foundational work is less visible to data end users. Foundational work is also less marketable, a key consideration in the nonprofit space when funds are scarce.

Early work at the Hive was focused on data science. The team used data science to apply segmented microtargeting to reach new audiences in the United States. Across 32 fundraising campaigns in 2015, which segmented potential donors on idiosyncratic characteristics such as demographics and interests, the Hive brought in nearly 200,000 new supporters to the refugee cause. However, the capabilities of the customer relationship management (CRM) and the existing team limited the scope and scale of this work in predictive modeling. The Hive had not yet invested in the foundation to succeed with data more broadly.

In early 2018, the Hive partnered with a data science, ML, and AI consultancy. Following their advice, the team expanded to a multidisciplinary team—composed of a data scientist, data analyst, data visualization designer, and data engineer—to support the data work more broadly at USA for UNHCR. In addition to

building the right team, the Hive invested in data foundations and implemented a data warehouse to be able to link multiple data sources across the organization, including supporter behavioral and financial data, providing a more comprehensive view of how supporters could engage with the refugee cause.

The Hive started developing data products, such as descriptive analytics, via dashboards and exploratory data analyses, to support fundraising teams across the organization. The initial focus on analytics support was to create a culture of trust across the organization and establish practices around how to measure success. By building a practice of analyses focused on donor behaviors, USA for UNHCR was able to recognize trends and anomalies specific to our supporter population. For example, the Hive consistently provides analysis on #GivingTuesday—one of the largest days of fundraising in the United States—which empowers fundraising teams to evaluate success and project targets for future years. Although descriptive statistics may seem simple when considering the larger landscape of data science, this type of approach can help to establish trust in the data and also create awareness for the potential of data applications.

Involvement in building these foundations allowed the Hive to ensure that data management solutions can support data science, ML, and AI in addition to analytics and operational use cases. However, this work was not without challenges, as the Hive struggled to overcome data and knowledge gaps. Prior to the Hive's focus on internal analytics, each fundraising channel—direct mail, telemarketing, digital marketing, and major giving—had a different vendor providing insights and helping to distill information from data. Some teams were evaluating performance outside of the traditional CRM; for example, the digital marketing team was utilizing an external platform to understand email engagement, which resulted in a disparate and inconsistent understanding of the supporter journey. This degraded trust in the data, as the quality of data collection varied, and duplicates were created as supporters engaged in multiple ways. Because the teams were siloed, there were often gaps when trying to synthesize the full journey of USA for UNHCR's supporters due to missing or inconsistent data as well as knowledge about how the data was generated, which was partially attributable to changing staff and vendors supporting each channel's data management. By prioritizing internal capabilities for analytics, the Hive was able to support standardized and consistent data collection across different data providers and vendors.

The Hive also supported a cultural shift, empowering fundraising teams to access the data warehouse by holding trainings on querying and accessing data across the organization and encouraging teams to help define key metrics used to support the broader analytics practice. Building data foundations is not the work of a single team and requires support from across the organization and from leadership to adequately invest and prioritize a single source of consolidated, trusted data. The pursuit of data science and machine learning requires an intentional focus on building data systems and teams to support that practice.

2.2. Creating and Leveraging a Network of Peers and Experts

Successful data innovation in service of refugees requires a strong network of partners. Funding partners provide resources. Frontline or implementing partners ensure that data-derived insights and AI/ML-based solutions are not only theoretically but actually useful (as data innovation is often one, or several, steps removed from the frontline). Outside experts can advise in-house data teams during solution development to ensure teams stay up-to-date in a quickly evolving field and to identify opportunities or accelerate delivery. In-kind contributions (e.g., data scientists ‘on loan’) can supplement often scarce human resources and technical talent on the side of the nonprofit organization.

Assembling and managing the right network of partners is no small feat. Funding, implementing, and expert partners may differ in terms of incentives and expectations. One party may have an interest in developing technically advanced solutions, another in developing a certain skill set, while yet another may prioritize impact to the front line. It requires significant in-house resources to define the purpose of partnerships, to align partners, and to put them in a position to meaningfully contribute to work. Without alignment and context, well-meaning efforts by partners can cause distraction from organizational priorities, lack of focus, and slow delivery.

At the Hive, we know we can only succeed with the right partners on board. We work to leverage experts in nonprofit, government, and the private sector supported by leadership to identify and develop novel approaches to solving the global refugee crisis. The Hive engages with the broader data for good community in three ways:

1. **Idea incubators:** The Hive leverages two key groups, hackathons through #HackABetterWorld and student groups with CornellTech and Stanford, to lightly scope and test ‘moonshot’ ideas. In the last 3 years, the Hive has hosted five hackathons to both create awareness for the refugee crisis among new communities and scope ideas that can be leveraged in other partnerships. The nature of hackathons encourages an atmosphere of creativity but also supports a structure to learn from failed ideas.
2. **Grants and pilot partnerships:** Using ideas generated and lightly scoped from hackathons and student partnerships, the Hive pursues partnerships with corporate social responsibility teams, government, and academics. These partnerships enable us to add resources to the Hive’s team of five and work with the broader community to create change for refugees. For example, the Hive worked with two professors from Harvard Business School to design a multi-arm experiment that increased retention among monthly donors. The Hive was also selected as the first nonprofit to participate in the U.S. Census Innovation Lab’s The Opportunity Project, where we worked to create a tool that assists resettled refugees in accessing resources available to them, working closely with USAHello.
3. **Counsel from experts:** Bringing together industry experts from government, nonprofit organizations, and the private sector, the Hive established a Data Advisory Board to influence the structure of the team and data projects internally. The Data Advisory Board has supported the Hive in creating a culture of data, in identifying and leveraging new partnerships, and in advising on technical approaches and infrastructure in

support of data science. The Data Advisory Board supplies decades of experience from organizations such as Airbnb, Deloitte, Google Creative Labs, Facebook, Netflix, and other organizations focused on data and technology. Additionally, we have previously worked with [92Y Giving Tuesday Data Commons](#) to understand data trends and insights across the broader nonprofit sector. The Hive also works with an informal collaboration among data science teams at nonprofits, such as [DonorsChoose](#) and [GlobalGiving](#), to troubleshoot how to do data science at a nonprofit organization.

This structure of partnerships and networking has supported the Hive in the pursuit of using data science to automate tent mapping in refugee camps. In 2017 at a DataKind Data Dive with the Bill and Melinda Gates Foundation and 92Y, the Hive challenged participants to build a minimum viable product that uses satellite imagery, a ‘moonshot’ idea that could inspire creativity without being stymied by reality. Following the hackathon, the Hive partnered with DigitalGlobe (Maxar), Tomnod, and Stanford AI and Sustainability Lab to prototype neural networks on satellite imagery of over 200 refugee camps in three regions. During the prototyping phase, we learned that although we had adequately tagged structures in refugee camps, we lacked tags on the negative space where structures did not exist, resulting in a less-than-ideal ML model. The project was still in the incubation phase, and the Hive needed to identify a new partner to update imagery and train a new machine learning model. In 2020, we submitted the idea of using imagery to update base layer maps and identify new features of Kakuma refugee camp and Kalobeyi settlement in northwestern Kenya to Microsoft’s AI for Humanitarian Action open call for proposals. The initial idea has been refined to support the needs of refugees, specifically in a location where the constant flux of refugees and few updates on their residences can affect the ability of UNHCR and partners to provide services and resources. Kakuma refugee camp was initiated in 1992 in response to the “Lost Boys of Sudan,” where more than 20,000 boys were orphaned, displaced, or forced to flee conscription during the Second Sudanese Civil War. By 2014, the camp had surpassed its capacity by nearly 60,000 (UNHCR, the UN Refugee Agency, 2021a). The camp has continued to expand, reaching a population of 218,380 in August 2021 (UNHCR, the UN Refugee Agency, 2021d). The Hive is now working with Humanitarian OpenStreetMaps to tag and validate drone imagery and Microsoft’s AI for Humanitarian Action data science team to tag features throughout the camp. Many refugee camps may not have initially planned for protracted displacement and are constantly adjusting to the ever-changing refugee crisis. By providing base layer maps of Kakuma refugee camp and Kalobeyi settlement, we can breach a critical data gap on infrastructure and help guide government agencies and organizations in the design and implementation of future interventions in response to refugee crises in East Africa.

Our experiences developing and leveraging appropriate partnerships illustrates the importance of having a diverse set of stakeholders involved in the various stages of data innovation. Our approaches to partnerships for idea generation, pilots for funding, and expert collaborators are independently effective strategies that can easily be adapted to any problem within the refugee sector and are generalizable to other domains for social good. The challenges that refugees and organizations supporting them face are not unlike the challenges other vulnerable underserved communities face (e.g., homelessness, physical safety, little access to food, electricity,

toilets, and other basic necessities like health care services). Like others, we also recognize that not all grants and partnerships are successful. The Hive takes feedback from failed partnerships, such as the difficulties in sharing refugee data, to improve subsequent applications. Not all partnerships can bring outcomes for refugees, and the Hive avoids partnerships that cannot deliver improved livelihood outcomes of refugees. Throughout our partnerships' work, we seek to create an ecosystem that can capitalize on a variety of expertise across industry and government, including direct funding and in-kind contributions, in order to create the most impact for refugees.

2.3. Prioritizing Reciprocal Partnerships With End Users and Beneficiaries

Impactful solutions require input from affected populations throughout all stages of the solutions development process. Stakeholder consultations are all the more important when developing data products for refugees, a population whose lived experience tends to be different from the lived experience of most data professionals. Ideally, stakeholders are consulted throughout the entire lifecycle of project development to foster a *two-way* exchange of information and skills development. Technology-solutionist 'solutions' tend to emerge when relationships are framed as one-way exchanges of information about needs—from refugee to data professional—and solutions—from data professional to refugee. Consultation and co-development pave the road toward meaningful, sustainable empowerment through data innovation, and it starts with the first step: identifying a problem worth solving.

But, access to stakeholders can be challenging in the refugee space, primarily because refugees are a legally protected group of individuals, and the Hive constantly seeks to improve processes that facilitate direct feedback in a responsible manner. To foster a culture of work 'from the perspective of the user,' the Hive prioritizes partnerships in the refugee space, notably working with resettled refugees directly via [Refugee Congress](#) and collecting feedback from programmatic nonprofits in the United States, such as [USAHello](#), [Refugee Investment Network](#), and [Kiva](#). We have also adopted strategies from our partnerships. [The Opportunity Project](#) centers solutions to challenges around user advocates and has created formal systems for feedback from end users, from the brainstorming to prototyping phase. We are working to emulate this structure to include both refugee experiences and feedback throughout our work.

Although we rarely have the opportunity to hear directly from one of the 82.4 million people who are currently displaced, we can leverage similar voices in the broader refugee community, notably resettled refugees, to better understand general challenges in addressing the global refugee crisis. We continue to seek to build connections in this area and are continuing to establish credibility with refugees and other stakeholders through our partnerships work. As the Hive continues to work at the intersection of data science and refugee issues, our focus has shifted to building solutions that are not simply 'for refugees' but 'with refugees.'

When we first began pursuing novel ideas, we did not have a structure in place to identify the most pressing issues among refugees, which resulted in some projects that failed to get off the ground simply because they

did not meet the needs of refugees. This lesson learned has greatly altered how we identify problems to solve in collaboration with our partners, by working with refugees resettled in the United States through organizations like Refugee Congress and HelloNeighbor to collaborating with UNHCR field officers in refugee camps, such as Kakuma camp in Kenya and Dzaleka camp in Malawi, to leverage local context in prototyping solutions. It is critical to root partnerships and networks in realities that can directly improve the lives of refugees, prioritizing inclusion of stakeholders and end users.

2.4. Accessing Relevant and Vital Data

A foundation for data innovation is access to relevant, high-quality data, but getting access to data is challenging in the refugee space. Some data relevant to address the needs of refugees may never have been collected; for example, in hard to reach locations or due to lack of internet access and cell phone connectivity. A challenge with missing data is their interpretation. Lack of data on refugees or their needs may lead to an underinvestment in resources and support if missing data are incorrectly interpreted. Similarly, there may be an underrepresentation of vulnerable groups, for example, women and children, in the data due to differences in access to the internet, cellphones, and the use of social media, which, if not accounted for, can lead to biased insights and algorithms.

Lack of access to data is exacerbated by the fact that much relevant data reside inside government and nongovernmental organizations as well as nonprofit organizations and private corporations where they are hard to access. Data cannot be made fully open either; the privacy of individuals and groups needs to be protected as refugees face persecution and other risks and potential harms. Data access needs to be balanced with protection of individual and group privacy (which is challenging, as [recent experience shows](#)). Lack of resources contribute to lack of access to much relevant data—technical resources to prepare and share data and legal resources to define the terms of data partnerships, in particular, in today’s fragmented data protection and data privacy landscape. A particular challenge in the nonprofit space is getting access to data beyond single-project-based access. Data are [massively reusable](#). Access to data can aid during project discovery, and yet, data access is often project-bound. While there are legal reasons for project-based data access, concerns about data privacy and protection, project-based funding structures contribute to this challenge.

The Hive does not have access to the warehouses of private refugee data, even though USA for UNHCR (and therefore the Hive) is a national partner of the UN Refugee Agency. We apply creative approaches in order to gain access to data. First, we use open source data platforms via UNHCR’s [data portal](#), the World Bank’s [Joint Data Center on Forced Displacement](#), and the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) [Humanitarian Data Exchange](#) to support ideation and prototyping. One of the challenges we face in working with open data is limited information on data collection protocols and other metadata—data about the data—to identify what data may be missing or what groups might be underrepresented in the data. There is a need to develop better data governance protocols and metadata management to reduce the risk of potentially harmful bias in insights and algorithms. Second, we develop our own data sources, leveraging web

scrapers to analyze data from online news sources on refugee and displacement issues and working with partners to create tagged imagery of refugee camps.¹ Third, we develop proxy variables where direct measures are difficult to access, for example, the degree to which a community is welcoming to refugees by identifying refugee-friendly local policies. Of course, the degree to which a proxy (refugee-friendly local policy) tracks the measure of interest (welcoming community) needs to be carefully evaluated over time as it may otherwise become a biased, potentially misleading measure. Where possible, the Hive seeks to make new data sources open source to foster an open and collaborative environment. The challenge in serving as an innovation team is the call to create novel solutions with the same data. The Hive is looking to pursue new data partnerships that will help sustain our existing work in the refugee space.

To support USA for UNHCR's fundraising efforts, the Hive put effort into democratizing internal data at USA for UNHCR. The Hive guided the development of a data lake and data warehouse, including the migration of an internal CRM, and developed analytics dashboards and reports to make data accessible and intelligible to a range of audiences. In the process, the Hive standardized metrics to streamline and foster consistent decision-making grounded in data at USA for UNHCR. As such, the Hive made data more accessible, made supporter management easier across fundraising teams, and in the process fostered a data-oriented culture at the organization.

One particular challenge in the space of fundraising analytics is that the tools needed to manage the operations of fundraising analytics and the tools to derive insights from data are different (a CRM is not a database), and yet you need a consistent, single source of data—known to data professionals as the ‘single source of truth’—for data-derived insights to support fundraising operations (insights derived from distinct, unconsolidated data sources may be inconsistent, decreasing trust in data, insights, and any resulting recommendations making it less likely that these will be acted upon). The different uses of data can pull data teams into different, sometimes conflicting, directions. Optimizing across different needs is a challenge, in particular, as you seek to experiment and innovate while maintaining sufficient continuity in fundraising operations and in the supporter experience.

As the Hive continues to build solutions in the refugee space, we are looking to adhere to and further develop an ethical approach in our practice. We are grounded by USA for UNHCR's vision of “a world without refugees,” which orientates the team and facilitates ethical reasoning among staff and partners. We are guided by the UN Refugee Agency, which, with a mission to safeguard the rights and well-being of refugees, limits the sharing of refugee data in the name of protection. Having good intentions is not good enough (and doing something is not always better than doing nothing), and we evaluate project proposals against our mission prior to development. In the past, this process has led us to abandon proposals. For example, at a planning event for our first #HackABetterWorld hackathon, we lightly scoped the idea of measuring the economic contributions of refugees resettled in the United States. Although we think this information could be powerful in influencing policy to accept more refugees, we feared that it would perpetuate a narrative that there are ‘good’ and ‘bad’

refugees, depending on their economic contributions, and that refugee status needs to be earned rather than given on grounds of human rights and human dignity. Ultimately, we decided the risk involved with perpetuating a narrative that we did not agree with was too great. As the challenge was highlighted by a Data Advisory Board member, we are now leveraging this group to further codify our approach to an ethical data science practice at the Hive.

2.5. Evaluating Change Beyond Traditional Monitoring and Evaluation Methodologies

Successful data innovation in the refugee space should bring meaningful change, making a difference in the lives of refugees. But assessing the consequences of one's efforts is challenging. First, it relies on close cooperation with frontline and other implementing partners for days and weeks, if not months and years, after the end of the development phase of a project; meaningful change can be slow to materialize. Second, it is challenging to identify and measure meaningful indicators of change. Process and outcome measures can help determine whether a project succeeds in delivering desired outcomes (i.e., impact) and can serve as important guideposts to assess necessary adjustments for success. Third, impact may differ from the actual change brought about by novel solutions, positive and negative unintended and unanticipated consequences. Access to stakeholders and other affected parties and the opportunity to gather quantitative and qualitative data on change is the foundation for an understanding of the assessment of meaningful change, which, in turn, allows for closing the loop on the iterative process of solutions development in search for better solutions over time.

When the Hive was looking to build its reputation in the refugee space and project portfolio, it was eager to accept projects. Recently, the Hive put in place a prioritization framework for projects guided by a project's potential to introduce meaningful change. To date, the Hive is utilizing a qualitative approach to understand change, a starting point in the absence of well-developed and recognized quantitative measures. This prioritization framework surfaced projects that may not, at first glance, look like the most impactful projects in the space. For example, the Hive worked closely with UNHCR's Division of International Protection to develop a prototype to automate the manual annotations of legal documents, particularly with respect to case laws, using machine learning. The outcome of this project is that these annotations help enhance navigation and searchability between documents to ultimately assist legal aid workers, lawyers, judges, and UNHCR to quickly find relevant information helpful in building legal arguments in protecting people of concern. Although this work may not be as glamorous as an 'app for refugees'—which is often not impactful—there is immense potential for direct impact on the lives of refugees.

One of the challenges that the Hive is facing is that not all projects are amenable to experimentation even in fundraising analytics. Therefore it is difficult to measure what would have happened if we had not enacted an intervention. For example, the Hive automated a daily data extract of USA for UNHCR supporters, segmented by demographics and engagement behavior, to upload to Facebook for targeted ads, creating both a lift in a catered donor journey and a decrease in the manual effort. We have seen an increase in fundraising via

Facebook ads, but we cannot say with certainty that the data extract caused this measurable change. Currently, the Hive is working on a feedback loop on how data was utilized to influence fundraising goals. In doing so, the Hive focuses on how processes at the organization have changed—sophistication of segmentation—rather than the quantifiable outcomes of each project.

3. Conclusion

Incredible opportunities are emerging to leverage data to make a difference in the lives of the 82.4 million refugees, asylum seekers, and internally displaced people around the world. However, realizing these opportunities in practice is a challenge. Although not exhaustive of the challenges one faces looking to innovate using data in the refugee space, we highlighted five common challenge areas:

- Investing in data infrastructure, systems, and analytics;
- Creating and leveraging a network of peers and experts;
- Prioritizing reciprocal partnerships with end users and beneficiaries;
- Accessing relevant and vital data; and
- Evaluating change beyond traditional monitoring and evaluation methodologies.

The Hive has had successes in tackling these challenges and hopes to serve as an inspiration guidepost to other organizations in the space. Of course, the Hive has work left to do! As data foundations mature, the Hive is looking to use gradually more advanced methods to deliver sophisticated solutions in the interest of refugees. The Hive is looking to partner with a diverse set of organizations, in particular, to continually strengthen the voice of refugees during solution development. The Hive is looking to increase access to data, internally and externally, through curation, collaboration, and increased, yet responsible, data sharing. The Hive is looking to develop a more comprehensive framework to systematically assess change and the consequences of projects, both qualitatively and quantitatively, to guide its efforts toward maximizing attainment of meaningful change.

Although the challenges are pervasive, with 82.4 million individuals displaced globally, USA for UNHCR and the Hive are committed to using the tools at its disposal—analytics, data science, and emerging technologies—to create a ‘world without refugees.’

If you are interested in potentially collaborating with the Hive in our data for good work, please reach out at hive@unrefugees.org.

Disclosure Statement

Nicole Smith, Muhammed Idris, Friederike Schüür, and Rita Ko have no financial or non-financial disclosures to share for this article.

References

- Genius Awards. (2019). *USA for UNHCR*. <https://geniusawards.secure-platform.com/a/page/winners/finalists-2019/unhcr>
- Harrison, N., & O'Neill, D. (2017, June 7). If your company isn't good at analytics, it's not ready for AI. *Harvard Business Review*. <https://hbr.org/2017/06/if-your-company-isnt-good-at-analytics-its-not-ready-for-ai>
- Kluge, J., Docking, T., & Ke Edelman, J. (2018). *Paradigm shift: How investment can unlock the potential of refugees*. Refugee Investment Network. https://refugeeinvestments.org/wp-content/uploads/2019/06/RIN_Paradigm_Shift.pdf
- Lewis, L. (2020, October 19). *Data science doesn't have to be sexy to be impactful*. Towards Data Science. <https://towardsdatascience.com/data-science-doesnt-have-to-be-sexy-to-be-impactful-7c20f3d12b41>
- Rogati, M. (2017, June 12). *The AI hierarchy of needs*. Hackernoon. <https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007>
- UNHCR, the UN Refugee Agency. (2021a, September 28). *Kakuma Refugee Camp and Kalobeyei Integrated Settlement*. <https://www.unhcr.org/ke/kakuma-refugee-camp>
- UNHCR, the UN Refugee Agency. (2021b, June). *Refugee data finder*. <https://www.unhcr.org/refugee-statistics/>
- UNHCR, the UN Refugee Agency. (2021c, September 28). *Resettlement*. <https://www.unhcr.org/en-us/resettlement.html>
- UNHCR, the UN Refugee Agency. (2021d, August 31). *Total refugees and asylum-seekers in Kenya*. <https://data2.unhcr.org/en/country/ken>
- UNHCR, the UN Refugee Agency. (2021e). *Financials*. <https://reporting.unhcr.org/financial>
- USA for UNHCR. (2020, January). *Protracted refugee situations explained*. [https://www.unrefugees.org/news/protracted-refugee-situations-explained/#What is a protracted refugee situation](https://www.unrefugees.org/news/protracted-refugee-situations-explained/#What%20is%20a%20protracted%20refugee%20situation)

©2022 Nicole Smith, Muhammed Idris, Friederike Schüür, and Rita Ko. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. Scraped data is used internally for analysis of how the public is conversing about the refugee crisis. It is not used for profit or shared outside of the organization. [↵](#)