Participatory Data Collection and Management in Low-Resource Contexts: A Field Trial with Urban Refugees

Ying Xu
College of IST
Penn State
State College, PA USA
ying.xu@psu.edu

Carleen Maitland
College of IST
Penn State
State College, PA USA
cmaitland@ist.psu.edu

ABSTRACT
As access to and control of data becomes increasingly democratized, understanding the potential and constraints for low resource contexts has important implications for system design as well as practice. Our research pushes the bounds of current system deployment by proposing and testing an ICT-based participatory data management system to transform participants from data providers to data consumers. This tool begins with participatory design, which engages participants in deciding which types of data to collect. Then, it involves training them in data collection, analysis and management. This enables participants to gain basic data science skills to make informed decisions. Our study uses mixed methods to explore the feasibility and effects of this system with urban refugees living in Rwanda. The quantitative results indicate refugees’ perceived effectiveness in using the system to build communities is directly influenced by system usability assessments, which in turn are influenced by education levels but not ICT self-efficacy. Qualitative results highlight the community-based interactions experienced by all participants and highlight important differences in pursuing projects with urban versus camp-based refugees.

CCS CONCEPTS
• Human-centered computing → Human computer interaction (HCI); HCI design and evaluation methods;

KEYWORDS
Participatory Design; Humanitarian Data Collection; Data Science; Open Data; Data Management; ICT4D

ACM Reference format:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICTD ’19, January 4-7, 2019, Ahmedabad, India
© 2019 Copyright held by the owner/authors. Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6122-4/19/01. . $15.00
https://doi.org/10.1145/3287098.3287104

1 Introduction
Data is increasingly produced and managed by communities themselves to make informed decisions in community building. However, in low-resource areas, data collection and management efforts are typically provided by other organizations like aid agencies or research institutes. They aim to (1) explore the needs of so-called beneficiaries, (2) make informed decisions, and (3) to provide the basis for monitoring and evaluation of aid programs, and enhancing organizational accountability [13, 38]. Within this realm, various initiatives have been undertaken to enhance data sharing and data exchange across organizations [5, 29]. However, despite such efforts to enhance sharing, collected data are rarely shared back with research subjects and the communities from which they are derived. Moreover, in some humanitarian sites, “research fatigue” can occur given the large number of data collection initiatives [8, 30]. This can deter research subjects from participating as well as affect the quality of collected data.

These limitations can be overcome when data collection and management adheres to open data principles, where data collection processes are designed from the outset to generate shareable data. However, some challenges remain in building such datasets that are open not only to those conducting research but also the researched community as well as the general public. Most importantly, open data requires data plans that strongly protect participants’ privacy while still gathering relevant and useful information. Here, we propose engaging both researchers/service providers and participants in the participatory design of a community data management system is both feasible and will enhance the usefulness of the data as a community asset.

This effort is undertaken in a humanitarian context, where ICTs are increasingly playing a critical role. As mobile phone and Internet skills evolve, the ability to manage data and extract valuable information from such sources are becoming crucial in navigating crises. Over time, as data science tools become more accessible and have higher usability, it is likely basic data science skills can be mastered by diverse user communities. Traditionally, in humanitarian response, control over information and the ability to use that information in decision making has been limited to organizations with sufficient resources. However, as technologies and skills diffuse, even low resource communities are likely to
have the potential to make data-based decisions on their own behalf. Developing these skills is likely best accomplished through training that provides the community access to data that is aligned with their own needs and interests [14].

Similar to many data science education initiatives, training low resource communities in data system design, including collection and basic analytic techniques, is a complex and multifaceted undertaking. Here, we report on an effort to introduce and test a participatory data system design process with urban refugees living in Rwanda. Most of the refugees are newly settled. In order to gain familiarity with their community and to be aware of the resources within it, refugees may benefit from defining, collecting and managing these data themselves: they can determine which data in the community they find interesting. This methodology is implemented in practice by actively engaging participants throughout the whole process of data inventory design, data collection, data analysis, and data management. The goal is to promote participants’ awareness of their community while gaining skills in collecting and using data. More importantly, unlike the majority of technology-based humanitarian efforts, which position participants as passive users of a given technology [12,18], we engage them in co-designing the data management system, including data collection forms, to reflect their interests, as well as how to ethically collect and utilize such data. This study incorporates lessons from a pilot conducted with Za’atari camp refugees in Jordan [48]. Both studies aimed to support community development that focuses on producing empowered, data-aware and data-utilizing community leaders.

2 Related Work

For the related work, we survey current studies in humanitarian data collection, participatory design in low-resource contexts and key metrics including perceived effectiveness usability and ICT self-efficacy.

2.1 Data Collection in Humanitarian Contexts

Traditionally, data collected in humanitarian contexts are used to inform the design of relief and recovery programs, to evaluate their efficiency and efficacy, or to advise on policymaking [22]. This type of data collection is important; however, communities lack access to these data for their own community development.

Among current practices, the most common method to involve participants in data collection is by paying them to collect data. For example, to minimize the cost of data collection and to gain contextual knowledge, researchers pay participants from the research site to issue surveys or conduct structured interviews [18,41]. In these studies, even though they took a more active role, community members were not acting on their own behalf. Participatory data collection can revolutionize such approaches by changing the goal: facilitating participants to act for themselves given the information that is available in their community. It gives a voice to the studied population by allowing them to determine which data are important to solving their own problems.

Humanitarian data collection has benefitted from the introduction of open source tools, such as Open Data Kit (ODK), which is used by a wide variety of organizations across numerous countries [11]. Two main components in ODK include ODK Collect which is a mobile platform that renders complex application logic and supports the manipulation of various data types, and ODK Aggregate which provides an easy way to deploy server to upload, storage and transfer data [21]. ODK has been transforming data collection efforts, particularly where the Internet is not universally available and programming skills are not present [26].

However, many uses of ODK embrace traditional organization-led approaches. Of about 363 articles (retrieved through Google Scholar) published since 2017 using ODK in data collection, only 34 involved the research subjects by using participatory design methods. The topics of these 34 studies cover knowledge co-production such as for fishing communities and for non-literate forest communities [32,45], health related projects such as case identification and monitoring for HIV [15], participatory mapping such for health facilities and roads [10,19], and policy design such as for land and forest resources [31]. Of those 34 articles, some were engaging participants to help design the questions while some were involving participants in collection strategies. However, none of them reported having discussed participants engaged in determining the ethics of data collection, having a say in the technical structures of their systems, and analysis and management of the collected data with their participants. Since the data are the aggregated information from the participatory approaches, a further step is demanded to train participants to be the leaders in both collecting and managing data. Therefore, for our first question, we aim to understand:

**RQ1:** What is the potential for and limitations to training community leaders on design and implementation of a participant-driven data management system in a humanitarian context?

2.2 Participatory Design in Low-Resource Contexts

Participatory design aims to enhance participation by a researched population in technology and process design in order to maximize project impact [4,26,47]. In refugee contexts, participatory design efforts have targeted adoption of specific programs (e.g. health related projects [1]) and promoting community development and social cohesion [2,23]. Refugee-driven data management can overcome several challenges faced by this community, including the lack of permanent status, such as accorded citizens, which hinders their self-determination and voice. They also suffer displacement, having lost knowledge of their local environment and community members [27].

Our previous research training camp-based refugees on data management demonstrated the initial feasibility of the participatory approach [48]. However, camp-based refugees live in confined spaces, where they can readily access services. By
comparison, urban refugees, left to find their own housing, are typically spread across an entire city. Arguably, this would make community knowledge even more valuable but at the same time more challenging for data collection and management.

Thus, for our second research question, we integrate participatory design, and data management to explore urban refugees’ involvement at the intersection of traditional ICTD and Data Science for Social Good. First, we will engage our participants in co-designing a data inventory form, which is composed of questions about community members’ capabilities that our participants care about. Then, we will engage them in designing the collection process and collecting data, and finally managing the inventory data. With this, we hope to propel our participants a step further along the ‘active users’ continuum [46]. Integrating participatory design into ICT-based interventions could be a better approach to engage participants as active contributors. Especially in ICT-related studies, given the participatory nature of mobile phones and the Internet, we can immerse participants in the various stages of research. While there are many studies using participatory methods with refugees and other marginalized communities in the design of technologies in the form of interfaces, tools, mobile applications [3,7,49], here the focus is on data-centric research. In summary, our second research question is:

RQ2: How do the potential and limitations to training community leaders on design and implementation of participant-driven data management vary between camp and urban refugees?

2.3 Effectiveness, Usability, and ICT Self-Efficacy

As the purpose of the participatory data management system is to equip community leaders with the system and skills to build data-driven communities, we are interested in understanding their perceived effectiveness of the system as well as the factors contributing to it.

For perceived effectiveness, there is no universal construct. Researchers usually customize their own metric to reflect their objectives [36]. The construct is important in gaining timely feedback directly from participants. For our participatory data management project, we aim to understand how it is influencing participants’ community building activities. Moreover, we are interested in understanding each procedure’s impact, which includes data collection, analysis and utilization. Example items include: “the collected data are useful in solving problems in life”, “I am confident that I can use the techniques learned from the training to conduct similar projects for my community in the future”, and “helping others using data increases my sense of responsibility in building a better community”.

To be able to use the system, one key question is to understand whether the system is easy to use and whether it makes sense to the users. Usability tests are established methods to get users’ insights through system and behavioral questions [6]. Brooke’s 10-item usability test – SUS - is one of the most adopted ones [9]. Example items include: “I found the various functions in the system were well integrated” and “I would imagine that most people would learn to use the system very quickly”. Other usability testing methods are also commonly used, such as heuristics, interviews, expert review and experiments [20,37]. Very few researchers who used ODK as part of their systems also tested usability of their systems. One study used heuristic evaluations on an mHealth application developed using ODK for midwives in rural Ghana [44]. However, the evaluation was only used to help inform usability of the prototype, and was not used to measure actual users’ perceived ease of use. However, to our knowledge, no studies had tested the ODK-incorporated systems using quantitative methods like Brooke’s scale.

As users have different levels of experience and competency with ICTs, to take individual differences into account is helpful to measure the true effects of the system. We use ICT self-efficacy to measure users’ own feelings toward their proficiency with various information technologies. Many scholars have looked into ICT-related self-efficacy measures, for example computer self-efficacy [11,28], Internet self-efficacy [16,25], and application-specific self-efficacy [50]. These studies have evolved with the development of different ICT skills. As the name of computer self-efficacy illustrates, it focuses on people’s judgment of their capabilities to use a computer. Studies have found that computer self-efficacy is positively associated with outcome expectation on performance, as well as actual performance metrics. For the Internet self-efficacy metrics, they usually cover aspects including terms relating to Internet software, problem solving, and data gathering. Internet self-efficacy has been found to have positive relationships with Internet access, years online, frequency of use, and online skills [25]. In terms of its relationships with psychological metrics, Internet self-efficacy has positive relationships with Internet experience and social outcome expectancy [16]. Last, application-specific self-efficacy is usually adopted to measure perception of efficacy in using specific applications or systems. In particular, with the high usage rate of mobile phones, application-specific self-efficacy is increasingly relevant [50]. To have a metric that covers self-efficacy from all these perspectives, we combined and created our own ICT self-efficacy metric. In summary, we want to measure whether the participants believe they have a range of different ICT abilities: to types, to download data and applications, to send emails, to use websites, and to use applications.

In summary, our third research question investigates the perceived effectiveness of the ICT-based solution and factors influencing it:

RQ3: What factors affect participants’ perceived effectiveness of the participatory data collection and management system in community building?

3 Method
After surveying the related areas, we proposed training on the use of a new humanitarian data management platform, composed of existing systems, which overcomes existing problems by incorporating participatory design and data collection, open data, and data science skills training to empower participants to make informed decisions. Our main goal is to provide evidence of whether it works or not.

3.1 Mixed Methods in Intervention Evaluation

To answer our proposed research questions, we applied a mixed-methods approach. In particular, we used surveys to measure ICT self-efficacy, usability and perceived effectiveness quantitatively. The detailed execution, measures and analyses will be discussed in section 3.5.

We employed qualitative analyses to gain insights from every step of the process of the participatory data collection and management. In addition, we used focus group discussions after the project was concluded to gain preliminary understandings of potential usage of the system by community members who did not participate in the project. This is particularly important: other than the participants from the experiment, ultimately, it relies on all the members to fully make use of the system. The discussions were guided by the following four questions: (1) What are the problems that your community is facing?; (2) What asset data are interesting to you?; (3) Can the asset data be used in solving those problems, and if so, how?; and (4) Do you plan to use the data to implement your ideas? Can you develop a course of action to solve the problem? The last question is designed to incentivize participants to engage in strategizing potential solutions.

Similar participatory data collection projects in refugee camps have been conducted before in Jordan [48]. This study expanded the paradigm to urban refugees, thereby generating broader design and data management implications for humanitarian contexts.

3.2 Research Context

In this study, we worked with Burundian and Congolese refugees living in the urban areas of Rwanda. By the end of 2017, the number of refugees in the world reached 19.9 million [43]. Rwanda, compared to other countries, has a very complex refugee situation. First, it has the history of producing its own refugees and handling the returning refugees. In addition, since the year of 1995, Rwanda has been hosting waves of Congolese refugees given the prolonged violence in DRC. Rwanda also began hosting refugees from Burundi since its political turmoil in 2014. Currently, Rwanda has more than 73,000 Congolese and 84,000 Burundian refugees, with its own population being 11 million.

As refugees arrive in a host country they sometimes have a choice of whether they want to live in camps or in urban areas. Those who live in camps can get access to centralized services, such as schools, medical care, food and water. Furthermore, within the camps, refugees have more interaction and camaraderie with fellow refugees. However, those living in urban areas generally have more confidence in their ability to survive as a result of job skills or access to financial assets. Nevertheless, the average length of being refugees is 17 years [17].

Kigali, the capital of Rwanda, hosts the highest number of urban refugees. There are more than 23,000 Burundian and 2,000 Congolese refugees living there. The majority of Burundian refugees have only been living in Rwanda for two years, while most urban Congolese refugees have been there for almost twenty years. In the urban area of Huye, which is the second biggest area with urban refugees living in Rwanda, there are about 3,500 Burundian refugees. All the Congolese refugees living in the region are living in camps.

In terms of refugees’ capacities in using ICTs, we learn and derive from both reports on mobile phone and Internet usage and from the UNHCR Rwanda. By 2016, the mobile phone penetration rate in Rwanda is close to 80%. The number of Internet subscriptions has reached almost 4 million, representing 35% of the population, and the annual Internet growth rate has kept above 8% [35]. Among these Internet subscribers, almost all of them use mobile Internet [35]. From our own anecdotal experience, the majority of the refugees used mobile phones, including many smart phones, to communicate with others and access the Internet. In addition, cellular data in Rwanda is relatively affordable: roughly 1 USD for 1GB.

3.3 Training

The participatory data collection and management was implemented through the form of training. The training lasted for five weeks and was conducted in UNHCR facilities in both Kigali and Huye. Table 1 shows our five-week schedule of both research sites. For each week, we had one to two sessions in each location. We commuted by four-hour buses in between. It is worth to note that, it generally took our participants in Kigali two hours of travel to reach our training site and three hours of travel for our participants in Huye. They usually had to take multiple public buses, which cost was compensated.

<table>
<thead>
<tr>
<th>Table 1: General Schedule of the Participatory Data Collection and Management Project Conducted in Kigali and Huye, Rwanda in the Summer of 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week</strong></td>
</tr>
<tr>
<td>Kigali</td>
</tr>
<tr>
<td>Huye</td>
</tr>
</tbody>
</table>

Our training had two major components: ICTs supporting the participatory data management tool, and the process of implementing it.

3.3.1 Structure of the Data Management System. The data management tool was mainly built using the UNHCR Kobo server (https://kobo.unhcr.org/), and the ODK Collect (http://opendatakit.org/) V1.10.0 mobile application created by
Participatory Data Collection and Management in Low-Resource Contexts

ICTD’19, January 4-7, 2019, Ahmedabad, India

researchers from the University of Washington on Android devices. The general structure of the system is shown in Figure 1. We chose to use the UN’s servers to store and access data to ensure the data is secure, and to use ODK Collect to collect digital data with no continuous Internet connection required. Participants only needed to use the Internet to upload the collected data to the server, and to view the uploaded data through a URL. All the collected data is openly accessible to all the refugee members in the community through an URL.

3.3.2 Procedures of the Participatory Data Management Training. First, we provided a general overview of the training procedures. The training included the following stages: participatory data inventory design, data collection, data analysis, and system management.

In Kigali, UNHCR helped to provide a training center. Additionally, they provided technical support including Android devices, a projector, and a Wi-Fi network. They also helped with translation services in both French and Kinyarwanda. We used slides to train our participants on the purposes of the project and the detailed guidance in using the tool. Translators helped us translating the training to French or Kinyarwanda. Given Huye is a much smaller city and the UNHCR office there is small, support for hardware for training was not available. However, UNHCR’s support in facilitating the training in terms of providing transportation compensation for the participants and translators was essential to carry out the project.

![Figure 1: Participatory Data Management System](image)

Participatory data inventory design helped us identify relevant capacities and resources within the community. We then aggregated all the items into a digital form. After the form was finalized, it was uploaded to the Kobo server to be downloaded on the ODK Collect mobile application. We trained our participants to download and use the ODK Collect application on mobile devices either of their own or provided by UNHCR. With ODK Collect on Android devices, our participants visited other refugee households to collect data based on the created inventory structure. After they collected the data from some households, they can choose to upload with their own cellular data or wait until they return to the training center to use WI-FI. Later, once data was collected from every participant, we performed some data analysis to explore the potential use cases of the data. Finally, we trained the participants to manage the server by creating their own accounts using their email addresses. Additionally, we trained them how they could create their own data inventory to conduct follow-up studies using the whole system.

3.4 Participants

To recruit our participants to join the participatory data management training, we collaborated closely with UNHCR. We used stratified sampling method to take refugees’ gender, location and nationality into account. Then, we applied random sampling in each stratum to identify participants. While we considered taking age into consideration, UNHCR Rwanda only records age into three groups: under 18 (38%), 18-55 (60%) and over 55 (2%). Since we only recruited participants who are over 18, the need to separate those under or over 55 was unnecessary given their relative proportions. To contact our potential participants, we called them via phone, though this likely introduced some bias to our sampling methodology as not every refugee had a mobile phone number listed. We finalized the participants by providing information of our project and then asking for their consent.

In Kigali, first, we gathered the general information about refugees who are registered with UNHCR in Kigali. Table 2 displays the basic statistics of the sample population. To only select participants who are over 18 years old, we sampled our participants through the pool shown in Table 2. In the end, there were 30 participants when we launched the data management activity. Among the 30 participants in Kigali, 3 participants were resettled outside of Rwanda during the training and another three dropped out for various reasons. In the end, only 24 participants finished the whole program.

In Huye, there are 1651 female refugees and 1738 male urban refugees, with a total number of 3389 refugees registered with UNHCR. They are all from Burundi. For the 10 participants in Huye, all of them were present throughout the whole training.

To summarize, all of the participants were iterate with an average of 10 years of education. Their average age is 35 years old. In total, we worked with 16 female and 18 male participants.

<table>
<thead>
<tr>
<th>Gender (population)</th>
<th>Nationality (population)</th>
<th>Sample in Each Stratum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female (7,415)</td>
<td>DRC (449)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Burundi (6,966)</td>
<td>14</td>
</tr>
<tr>
<td>Male (7,484)</td>
<td>DRC (450)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Burundi (7,034)</td>
<td>13</td>
</tr>
<tr>
<td>Total (14,899)</td>
<td></td>
<td>30</td>
</tr>
</tbody>
</table>

Table 2: Breakdown of Stratified Random Sampling for Participants in Kigali, Rwanda (Age > 18)

<table>
<thead>
<tr>
<th>Gender (population)</th>
<th>Nationality (population)</th>
<th>Sample in Each Stratum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female (7,415)</td>
<td>DRC (449)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Burundi (6,966)</td>
<td>14</td>
</tr>
<tr>
<td>Male (7,484)</td>
<td>DRC (450)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Burundi (7,034)</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 3: Breakdown of Stratified Random Sampling for Participants in Huye, Rwanda (Age > 18)
Table 4: Summary of Participants of Focus Groups

<table>
<thead>
<tr>
<th>Sub-Communities</th>
<th>Women Association</th>
<th>Community Leaders</th>
<th>Peace Volunteer</th>
<th>Huye Gathering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Participants (Female)</td>
<td>7 (7)</td>
<td>7 (2)</td>
<td>9 (2)</td>
<td>5 (3)</td>
</tr>
</tbody>
</table>

3.4 Measures and Analyses

To answer our research questions, we issued a survey on three metrics at the end of the training, which include a 10-item usability test - SUS, an 8-item ICT Self-Efficacy scale, and a 7-item Perceived Effectiveness metric, in addition to a pre-training test on the same ICT Self-Efficacy scale. We also collected our participants’ age, gender, number of years of education, number of years since moving to Rwanda, and the city they live in during the training.

We issued pen-and-paper surveys when we finished the whole project. All these items were presented in both English and Kinyarwanda. UNHCR staff offered important help in translating the items into Kinyarwanda before we landed in Rwanda. In the appendix of this paper, we present the English version of the metrics that we constructed.

To determine whether there is a significant change of our participants’ ICT self-efficacy, we use paired t-test. To analyze the relationships among ICT self-efficacy, usability and perceived effectiveness, we use linear regressions.

4 RESULTS

In this section, we will first present results from the participatory data management procedures we described earlier. Then, we will provide some descriptive analyses of the three most important metrics before we dive into inferential statistics on the relationships among them.

4.1 Participatory Data Management Procedure

The procedure covers three main steps: co-design, data collection, and data analysis and management. Unlike the previously explored data management project in Za’atari refugee camp in Jordan, which only moved partially down the participatory continuum, this project aims to engage participants in every possible way. We will highlight some of the major differences in details.

4.1.1 Co-Designing Data Management Processes. For each research site, we initially outlined to all the participants the goal of this participatory data management project, which is to empower them to be the leaders in collecting and using relevant data to build their communities. Then, we described and showed the technical affordances of the data management platform. This took rounds of clarification for participants to understand that any data of potential interest for community development can be collected for their own use. These data might include skills, hobbies, education and other assets that could be leveraged by the community to solve its own problems. Furthermore, the tool can be used just by themselves, without the involvement of government agencies or humanitarian organizations, to collect more data after the research is concluded. Throughout the training, we found that it was useful to remind them of these goals since it is a big mindset shift for them to realize they are the owners of the whole procedure. During this repeated process, we were able to reinforce our participants’ understanding of assets as well, which is not an easy concept to communicate cross culture and languages. Essentially, assets meant any social, physical, natural and human capacities that they find useful [40].

Then, we conducted a participatory design session with all the participants involved. We first covered some basic data collection ethics, which include explaining the purpose and asking for permission from the households before starting collecting information, terminating the procedure whenever the household member asked to, as well as emphasizing the open source nature of the data even after the study is concluded. Then, we discussed the details of the data collection procedure. To adhere to the open data principles such as openness and accessibility, we had to take privacy concerns such as exposing the households’ personally identifiable information into consideration.

An important element of participatory design was reviewing the technical affordances of the system, which included a mapping function. Participants were asked to discuss whether or not they wanted to use the function and the potential privacy implications of recording respondents’ locations. Our participants decided to include a question on collecting addresses but only fill it after gaining consent from the household they collect the asset data from. In this way, they can trace back to the assets when needed easily.

4.1.2 Co-Designing Data Inventory. Later, in both Kigali and Huye, we brainstormed on the capacities and resources among the refugee families in the community separately. Refugee participants were interested in various assets across physical, human and social aspects, which include skills, hobbies, education background, community activities and places to socialize. These assets could be useful in bringing community members together with shared interests, enhancing members’ skills by recruiting voluntary instructors, or finding members with complementary skills for starting a business.
Some of the hobbies suggested by our participants include music, drumming (Burundian drummers are popular in the region), reading, painting, cooking, and singing. Some interesting skills include agriculture, security, beer-making, handicraft, English, medical skills, and jewelry making. Interested gathering locations include family and friends’ houses, churches, playgrounds, stadiums, hospitals, bus stops, markets, and bars.

The majority of the questions and choices for each question are similar between Kigali and Huye. Among all the categories, places to socialize vary the most. For women, Duschirenwe Women Association is a popular place for women to socialize in Kigali; however, there are no equivalent facilities in Huye.

![Image](image1.png)

Figure 2: Participants Brainstorming on Asset Data to Collect

4.1.3 Collecting Data. We let our participants grow their data collection network using their own methods. They usually started by reaching out to other families living in their neighborhood, through whom they were introduced to others they did not know before. In either case, they usually called the families to set a time before going to visit them. Since the cities are relatively big, it took our participants an average of more than one hour to reach each family. Figure 3 shows an example context of data collection. One of our participants was using ODK Collect on his own mobile phone to ask questions about the household’s capacities and assets. Outside of asking questions, they usually took extra time to get to know one another as well.

By the end of the project, the refugee participants were able to collect data from 214 households in Kigali and 57 households in Huye. In both Kigali and Huye, about 20% of our participants collected about 50% of these households’ data. For these seven participants, their average age is 38 years old, their average years of education is 10 years and three of them are females. Among all the data collected, 49 households (23%) agreed to have their location geo-tagged in Kigali, while only 5 households (9%) did so in Huye. Our participants only record location information when the households they visit permit them to do so.

In Rwanda, while the project was embraced by UNHCR, their level of involvement in the data collection design process was less than our previous project in Za’atari camp, Jordan. In Jordan, as they were concerned with representativeness of the data for their own purposes, we used stratified random sampling to ensure a representative sample of refugee households. Here, given the constraints of an urban setting, namely that it is unclear, even to authorities, where all refugees live, we could not nor did we want to provide a mandatory sampling strategy to our participants. Instead, they were left to their own devices to determine how to reach other refugees and collect their data based on the produced inventory. While this likely had negative effects on the representativeness of the data, it allowed flexibility. We also did not set specific goals for each participant to achieve, such as the number of refugee households they need to visit.

The lack of specification over household data collection targets did create an additional difference. Unlike in the camp environment, we did not have the opportunity to engage in data quality control before the participants submitted their collected data to the server. Yet, without a quota we would assume there would be limited motivation for our participants to make up data. The similar project that was conducted in the Za’atari camp used some quality control techniques because the researchers provided a list of households for each participant to collect data from [42].

![Image](image2.png)

Figure 3: A Participant (left) was Collecting Data from a Burundian Refugee Who Lives with Her Two Children

4.1.4 Understanding Collected Data. Every week we met face-to-face with our participants, we learned about their unique experiences with the data collection process. They usually liked to share their new experiences of meeting other interesting people through this project. In addition, we collaboratively explored the collected data with the participants.

We used the Kobo UNHCR server to store and access the data. Kobo also provides built-in visualizations like histograms of specific questions and maps of the visited households’ locations. An example interface is shown in Figure 4. Given the sensitivity on geographical locations, we do not provide its example visualization here. Additionally, users had the option to download the collected data and use external tools to analyze them. However, since we did not have computers or laptops available in the training center, we only viewed the histograms and maps generated by the online server on individual phones or tablets.

In terms of the topics our participants were interested in analyzing, because most of our participants did not have a full-time job, they were interested in opportunities to learn additional skills and gain resources to accomplish their own goals. For
example, because majority of the refugee participants only spoke French and Kinyarwanda, they wanted to know how they could learn English, either from their fellow refugees who were already fluent or by forming study groups. Some refugees wished to buy or rent land to produce food for other refugees, but they were short of many necessary resources to accomplish this goal, such as the land itself, initial investments in agricultural equipment, and the necessary agricultural skills. Some refugees wanted to have more social and entertainment activities, such as making beer and having better places to hang out. By using the system, they had an improved chance to access relevant information within their community to help them accomplish these goals.

Next, we will use different statistical tests on the survey data we collected. In table 4, we provide the mean, standard deviation, and Chronbach’s alpha of our three primary metrics. In order to understand how perceived effectiveness is influenced by usability, we use ICT self-efficacy as a control variable together with other demographic features.

Table 4: Descriptive Statistics of the Key Metrics

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Number of items (Scale)</th>
<th>Mean</th>
<th>SD</th>
<th>Chronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICT Self-efficacy</td>
<td>8 (0-3)</td>
<td>1.78</td>
<td>.55</td>
<td>.88</td>
</tr>
<tr>
<td>Usability</td>
<td>10 (1-5)</td>
<td>3.26</td>
<td>.53</td>
<td>.80</td>
</tr>
<tr>
<td>Perceived effectiveness</td>
<td>7 (1-5)</td>
<td>3.32</td>
<td>.46</td>
<td>.66</td>
</tr>
</tbody>
</table>

4.2.1 Perceived Effectiveness. Among the 34 participants who answered the survey on perceived effectiveness, there were 26 participants (76%) who agreed that the data collection improved their awareness of their community; 23 participants (68%) agreed that data collection helped them engage more with their community, and 14 participants (41%) agreed that data analysis helped them become more aware of the capacities in their community. In addition, 17 participants (50%) agreed that the collected data will be useful in solving problems in their everyday lives, while 14 participants (41%) agreed that helping others using the system could increase their sense of responsibility in building a better community. There were 20 participants (59%) who expressed that they would like to use the system frequently, and 24 participants (71%) who believed that they are confident to use the techniques to conduct similar project for their community in the future. These numbers were likely inflated due to social desirability. However, our participants were more critical on rating usability, as discussed below.

From Model (1) in Table 5, we find that usability contributes significantly to the participants’ perceived effectiveness when controlled for demographics and ICT self-efficacy level.

4.2.2 Usability. For usability test results, after scaling, the score is 65 out of 100, which indicates only a moderate level of usability. The reasons for this moderate usability might be explained by the fact that the tool was the first product focused on participatory interventions introduced to the community and the usage period was only five weeks long. When we categorize usability into high (mean = 72, min = 62, max = 84) and low (mean = 55, min = 48, max = 60), we find that those who have lived in Rwanda longer and have higher education are more likely to rate usability higher. Additionally, those who rated usability higher also experienced higher perceptual effectiveness.

Looking into details, among the 34 participants who answered the usability tests, 5 of them (15%) found that the system was unnecessarily complex. There were 12 participants (35%) indicating that they might need support of a more technical person to be able to use the system, and 3 participants (9%) strongly felt
that they needed to learn a lot before they can comfortably use the system. Therefore, we can consider the following modifications in the future to increase usability: additional training could help increase their familiarity with the tool, and the usability test could be issued after longer period of usage [34].

Table 5: Statistics on the Usability and Effectiveness of the Participatory Data Management System

<table>
<thead>
<tr>
<th>Control variables</th>
<th>(1) Perceived Effectiveness</th>
<th>(2) Usability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>.04</td>
<td>.11***</td>
</tr>
<tr>
<td>Gender - M</td>
<td>-.02</td>
<td>-.08</td>
</tr>
<tr>
<td>Age</td>
<td>-.00</td>
<td>.00</td>
</tr>
<tr>
<td>Location - Kigali</td>
<td>.26</td>
<td>-.50*</td>
</tr>
<tr>
<td>Years in Rwanda</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>ICT Self-Efficacy</td>
<td>-.01</td>
<td>-.01</td>
</tr>
</tbody>
</table>

Predicting Variables

- Usability: .49*
- N: 34
- DF: 26
- Adjusted R²: .31 *

Note: *p < .05. **p < .01. ***p < .001

We also ran linear regressions on the factors contributing to usability scores. Model (2) in Table 5 presents the result for usability of the participatory data management system. We find out that, (1) the more education our participants have received, the more likely they rated usability higher; (2) our participants in Huye thought the system had a higher usability; (3) ICT self-efficacy did not have significant effect in usability; (4) gender, age, location, and years living in Rwanda did not contribute to explaining usability; and (5) no interaction effects were detected among all the variables. In summary, other than education, the only other significant factor is location. Between Kigali and Huye, the biggest difference is that the number of participants is much smaller in Huye. Some studies on communities have found that smaller groups can allow for rich interactions [33].

5 Discussions

Our findings have implications for humanitarian data collection and management with refugees and, more broadly, participatory design in low-resource contexts. We also generated some practical recommendations for research in multi-cultural environment.

5.1 Participatory Data Collection and Management for Camp and Urban Refugees

Our ICT-enabled participatory data management solution enabled (1) the open data principle, which requires data be shareable with stakeholders, especially participants; (2) participatory design of the data inventory, which gave participants a voice in determining the kinds of information they wanted to collect and the actions they want to take with the data; and (3) participatory approaches to data analysis and management, which trained participants to be leaders in using data to make informed decisions to build their communities on their own behalf.

As a whole, the training and implementation process promotes various types of community interactions: interactions among the refugee participants during the whole process, interactions between participants and other community members through data collection, and interactions among community members especially as they helped the participants contact other refugees they know who could participate in the survey. Through these interactions, refugees who participated in any capacity gained some knowledge of their community.

This participatory data collection and management project was introduced to urban refugees for the first time. In contrast, our previous work in a camp demonstrated some advantages of that context, namely: (1) members are co-located within a defined geographical area, which makes face-to-face interactions easier; (2) various stakeholders including UNHCR, NGOs, and the local government are present to promote such development activities; and (3) common facilities like community centers can be used to facilitate activities [48]. It is less likely for urban refugees to have these advantages given the differences in resettlement procedure between these populations [42]. However, lacking these facilities, urban refugees generally have more means and responsibilities to support themselves. Therefore, the greater independence of urban refugees might serve as an important advantage for community building activities.

An important difference between the urban and camp contexts was sampling methods, where snowball sampling was used in the former versus stratified random sampling in the latter. There are pros and cons for both methods. Mainly, it is a balance of data quality, the effort it takes, and what is feasible for the context. In line with our participatory approach, the trade-offs between these issues should be left to data managers to decide. However, regardless of whether or not representativeness is deemed important, it is beneficial for the benefits and limitations of representativeness to be made clear, so that data managers can accurately reflect the content of the database. It is also important to note that the boundaries of what constitutes ‘the community’ must be defined.

5.2 Participatory Design in Low-Resource Contexts

A key feature of our project is the engagement of refugee participants not only in the process of designing the data inventory, but in the design of key aspects of the overall data management system, taking into account cultural norms. This project extends refugees’ involvement from passive acceptors of help or paid enumerators of fixed studies [41] to active contributors with initial skills to use data to solve problems identified by the community.
Refugees participating in the project made numerous decisions. The decisions varied, including from whom to collect data, the types of data that were the most intriguing, and the types of analyses they want to conduct. They also reflected on how they can act upon the data they find interesting, and further, how they might collaborate with others to solve the problems they care about. Participatory design serves as a critical starting point to democratize the data management process and maximize the power of information for stakeholders in the community [24].

We also found that participatory design was not an intuitive concept for aid agencies and refugees. When working with aid agencies and refugees in Za’atari camp, we had to keep reminding them that the data was openly accessible not only for the agencies but also for refugees. Similarly, when working with urban refugees in Rwanda, it was necessary to repeatedly explain to the participants that the data and the system were built for them and they were the ultimate consumers of the collected data. Given the limited rights of refugees as compared to their citizen counterparts generally, this need for continued reassurance may stem from their marginalized status and hence may be unique to this community. Future research on data management systems might further explore the relative perceptions between these groups.

Even within refugee communities, we find divergence on perceptions of rights to and value of these data. In both our urban and camp-based studies, we used focus groups to infer broader community members’ utilization of the data. In both the urban and camp contexts we found that different sub-communities have different sentiments and understanding of the project. For example, women’s groups actively working to improve lives for themselves, their children and their communities, are more likely to see the value of data in solving specific problems. However, groups like community leaders, typically and in our case consisting of older men, are more resistant to change. This suggests the broader effectiveness of our system is likely to depend on having the right group of leaders who understand not only the potential value of the information but also changes in the humanitarian context that are driving a greater emphasis on self-reliance rather than traditional mechanisms of aid. In turn, this will require appropriate strategies and implementation plans.

5.3 Multicultural Research Context

Invaluable local knowledge from the service providers is one of the researchers’ greatest assets, from testing the feasibility of the system and assisting in coordination, to implementing the project and brainstorming for improvement. As a methodological note, we would like to draw attention to the need for flexibility in conducting research in multilingual contexts [39]. In our case, before we started our field study, we created all the assessment metrics and worked with UNHCR Rwanda staff to translate them to Kinyarwanda. However, when we showed the survey to our participants, they requested us to translate it into French since they use Kinyarwanda primarily as a spoken language and they read and write in French. Later on, when we developed the asset data collection forms, we used French instead. Therefore, even though we have close collaboration with local agencies while conducting field studies, being prepared for changes is an important mindset for conducting multilingual field studies.

5.4 Limitations and Future Studies

As a rather ambitious undertaking, our project served as merely a proof-of-concept for what we hope will someday become a fully functional system used by urban refugees.

As a participatory project, throughout the procedure, we tried to maximize refugees’ participation. However, it can still be significantly improved from multiple perspectives. First, for design, participants decided on both the types of questions and choices for each question with guidance from researchers. As an pilot study, this was deemed necessary. However, future studies can let participants control the selection of interested data. Second, for data collection and analysis, participants set goals and methods for themselves without any explicit influence by the researcher. However, third, system management was conducted in a form of training to prepare participants in using the system for future needs. Participatory efforts could be done between researchers and participants in the future.

To understand the challenges in systemic use beyond those mentioned above, future research should consider how to engage local aid agencies and governments to integrate the capacities of refugees into the broader humanitarian data management process. Key to such an investigation would be developing knowledge on how participatory approaches and basic data science training can be more fully integrated into ICT-based livelihoods training common in a variety of humanitarian contexts.

6 Conclusions

As data has been increasingly pervasive and accessible, skills to manage them become critical in making data-driven decisions for community building. However, in low-resource areas, both access to and control of data are limited. In this study, we proposed a participatory data management system, which enabled open data principles, participatory design of the data inventory, and data science skill training in order to improve participants’ data collection, analysis and management capabilities.

The proposed data management system was implemented with urban refugees in Rwanda during a period of five weeks. Using mixed methods, the study confirmed the feasibility of the system and found a significant improvement of ICT self-efficacy among the participants. The results also indicated that refugees’ rated usability is strongly correlated with their perceived effectiveness in building communities.

This study, for the first time, contributes to participatory data collection and management with urban refugees. It generated different implications for participatory design and data management skills for low-resources and multicultural contexts, as well as opened door for more future improvements.
APPENDIX

I. ICT Self-Efficacy Scale
1. I am confident that I know how to use information and communication technologies (such as the Internet and mobile phones) to gather information.
2. I am confident that I know how to use information and communication technologies (such as the Internet and mobile phones) to communicate.
3. I am confident that I can learn advanced information and communication skills within a specific program.
4. I am confident that I know how to download a file.
5. I am confident that I know how to send an email.
6. I am confident that I know how to use my favorite websites.
7. I am confident that I know how to use mobile applications.
8. Please rate your general skill in using information and communication technologies.

Beginner, average, advanced, or expert

II. Perceived Effectiveness
1. The participatory data collection, in general, improves my awareness of my community.
2. Data collection helps me engage more with my community.
3. Data analysis in the lab helps me become more aware of the assets in my community.
4. Collected data are useful in solving my problems/challenges in life.
5. Helping others using data inventory increases my sense of responsibility.
6. I feel I am collaborating with NGOs in conducting this participatory project.
7. Helping others using data inventory increases my sense of responsibility.
8. Collected data are useful in solving my problems/challenges in life.
9. I am confident that I know how to use my favorite websites.
10. I am confident that I know how to use mobile applications.
11. The participatory data collection, in general, improves my awareness of my community.
12. Data collection helps me engage more with my community.
13. Data analysis in the lab helps me become more aware of the assets in my community.
14. Collected data are useful in solving my problems/challenges in life.
15. Helping others using data inventory increases my sense of responsibility.
16. I feel I am collaborating with NGOs in conducting this participatory project.

REFERENCES

Alfred P Rovai. 2002. Building Sense of Community at a Distance. International Review of Research in Open and Distance Learning 3, 1, 1–16.


