Local Government Training Institute

(LGTI) Course Data Curriculum

Data Collection Course

Module 1: Preparing A Data Collection Project

Disclaimer: This work is developed by School of Data with funding from The World Bank Tanzania Data programme. For more information visit https://www.schoolofdata.org.

Module 1: Preparing for a Data Collection Project

STUDENT WORKBOOK

Introduction

Many will not dispute the value that data provides in the development of useful insights for decision making. The GovLab in its Data Life Cycle clearly lays out the different stages data goes through to become useful to its end user.

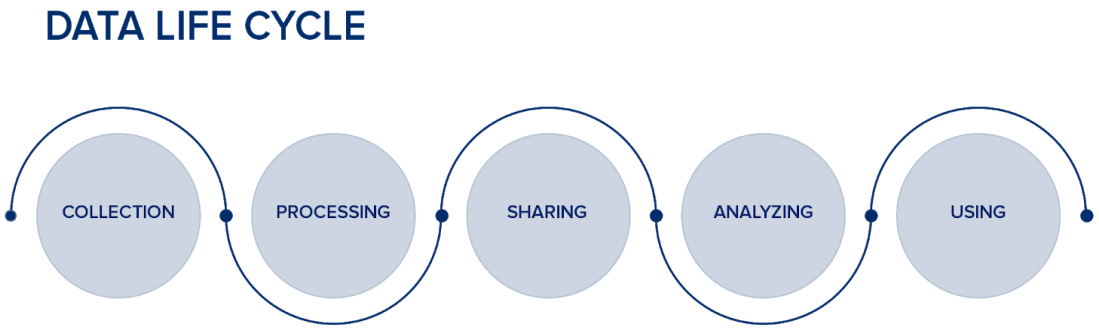


Figure 2: Data Life (The GovLab)

The Data Life Cycle by The GovLab

The Data Fundamentals Course focused on the concepts, techniques and tools one needs to effectively carry out data-related work leveraging on the School of Data Pipeline. However, the course assumed an availability of collected data that needed to be obtained and converted into a format that could then be used in the subsequent stages of the Data Life Cycle or Data Pipeline.

Now that the learner has a grasp of the data fundamental, we focus on data collection. This module takes an in-depth look into the first stage of this process of actually collecting the data to eventually generate insights. Despite the presence and continuing development of many data collection methods, this course will primarily focus on data collection using open-source mobile data collection tools and methods. At the end of of the module, the learner will be able to:

- Explain what constitutes a field data collection project and the different types that can be used.
- Develop a field data collection project scope including the target audience, geography and budget
- Identify the roles and responsibilities needed to carry out a field data collection project
- Design, develop and deploy a survey methodology and questionnaire to collect data

Lesson 1: Introduction to Data Collection Projects

Cases for Data Collection

Example 1: Women Environmental Programme, Abuja Nigeria

Local Government Administrations (LGAs) are the third tier administrative structure created in Nigeria to decentralise governance, bring government closer to the people at the grassroots and render social services pivotal in engendering national development. They are purposefully located and responsible for the governance of about 70 percent of the estimated 170 million people of the Nigerian population. Thus, they are said to be

in a vantage position to aggregate and articulate the needs of the majority of Nigerians and facilitate rural development through the application of the needed financial and human resources in their operations.

The demands and expectations from local government councils over the years have been on the increase while the finance required to deliver the dividend of democracy and good administration at the grassroots continue to dwindle, is inadequate, mismanaged and misappropriated. The LGAs have been characterized by poor accounting systems, unavailability of reliable data required for planning, over-politicization, inadequate finance and poor revenue collection, greed, unnecessary government structure.

Women Environmental Programme (WEP) a civil-society organisation (CSO) based in Abuja, Nigeria wants to promote transparency and accountability in local governments and believes obtaining data on service delivery and infrastructure in local communities will provide great insights into what issues to focus on in their advocacy. The challenge however for WEP was the lack of the required data from government agencies or other

organisations for their topic and location of interest. After some time of searching and consultation, WEP resorted to carrying out a data collection project in 160 local communities to obtain the data need for their advocacy purposes.

The broad objective of the project is to evaluate the root causes of failure in the local government areas and how effective budget controls brings about efficient governance in the local government systems.

Traditional Data Collection

A few years ago, whenever someone referred to carrying out a data collection project, the idea that came to mind was a massive, expensive, paper-based census exercise being carried out by the government. These censuses were crucial to governments as the data collected was used in evaluation, planning and decision-making around a national agenda.

Despite their benefits, these census projects could take several months if not years to provide the data in a form that could be used. On the other hand, most of this data even took longer to be accessible to ordinary citizens. For individuals or organisations like WEP who do not have the bandwidth and time to dedicate to such a massive data collection exercise, easier alternatives that provide similar data collection outputs is critical.

When, Where and How

Obviously, you can only collect data to which you have access. The first part of defining a data collection strategy is understanding when or where will the act of data collection take place.

- **WHEN** applies to something that goes through different steps of a chain: for example, a fish is first caught on a fishing boat, then brought to a harbour, sent to a storage room etc.
- **WHERE** applies to the location where you might have to go to do the data collection: you will not be able to bring the same equipment on a boat or in a office building.

Once you have your answer, you need to define **HOW** you will collect the data. It could be :

- using one or several existing database(s),
- using forms, questionnaires and interviews,
- using direct observations.

Direct observations is making direct measurements of the subject being studied (like the

number of buildings in need of repair after an earthquake), rather than using indirect measurements, such as databases or testimonies from people knowledgeable in the field. Most often, direct observations are made by human observers, sometimes with the aid of specific equipment:

- Pollution sensors to measure air pollution
- Geiger counters to measure radioactivity
- Smartphones to record visual observations and, potentially, sound.

So you have a project in mind, and it requires collecting data using mobile phones. Or maybe you already started collecting data, but your pen and paper method is showing its limits. And on top of that, your experience of mobile data collection is close to zero. Have no fear! You've arrived at the right place.

The spread of cheap smartphones, which started during the 2010s, allowed the potential of mobile data collection to be unlocked. A multitude of projects, from urban mapping to the surveying of remote communities, dropped their previous cumbersome methods (pen and paper, specialised suite of hardware tools...) to switch to collecting data using the now-ubiquitous smartphone. But a tool can only do so much. Correctly collecting data in general requires a mix of knowledge and skills that no smartphone will replace. And if your project is of any significant size,creating an interdisciplinary team with the right mix of skills will be vital to its smooth and successful execution.

Lesson 2: Project Team

Setting Up Your Team

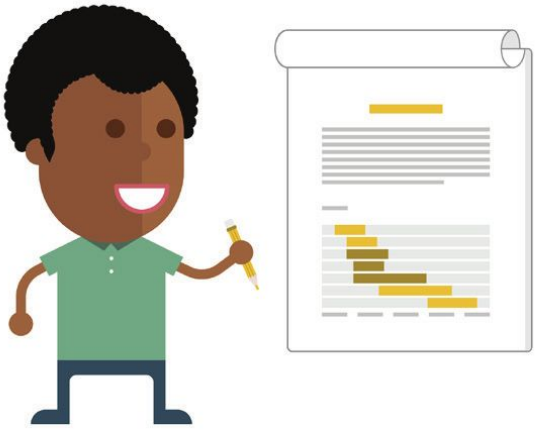
Collecting data using smartphones has many advantages in terms of ease-of-use and cost-reduction, especially over the paper based method:

- They are available in sparsely populated areas, which means that the act of collecting data can be, if needed, crowdsourced by people living in the field
- They often feature at least a basic camera and sound recorder, which allows for multimedia data collection
- The results of the data collection can be sent in real time, provided an internet connection is available.

But there are many ways in which smartphone data-collection can go wrong, making the whole process costlier and more difficult than it should be! This is why a dedicated

preparation phase is vital. To give you the best overview of the tasks involved in collecting data, we've divided them into the four key roles that feature in a data collection team. These functions aren't rigid, of course: your team members may be able to cover multiple roles or there may be a sub-team focused on a particular area. Your choice of resource allocation should encompass the level of expertise of each member, the scope of the project, and the budget available.

The Project Manager



Project Managers are responsible for overall management of the survey. They define the project goal and are responsible for all the major decisions that need to be taken during the project. They are also responsible for communication amongst the survey team, to make sure that everybody is on the same page. Finally they report the project status, maintain a record of lessons learned and lead retrospectives.

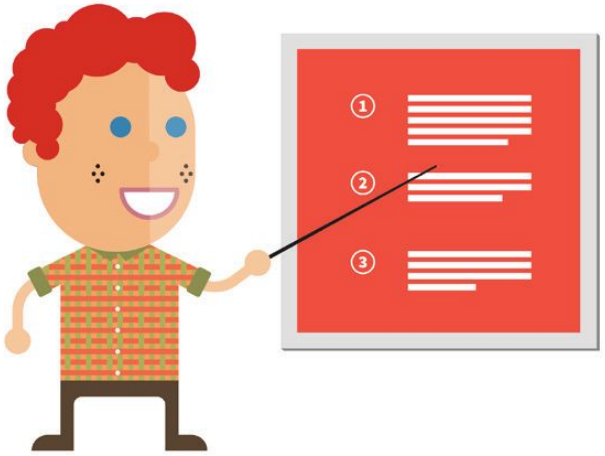
Timeline of Tasks

The Survey Designer

Survey Designers are on a mission to construct questions that elicit just the right data from the survey's respondents. They have to make sure questions are not ambiguous and are as straightforward as possible, encompassing complexity and subtleties in a way the respondents fully understand their meaning. Their role also has a large visual design element: they ensure the survey looks appealing and is engaging to complete.

Timeline of Tasks

The Trainer



Trainers run the training sessions that will help interviewers, or enumerators, learn the skills they need to perform a successful data collection with participants. Even if the survey is well designed the quality of the data collected through the interviews depends ultimately on the data collection skills of the enumerators. Trainers make sure they get acquainted with the mobile devices and the mobile app used to conduct the survey.

Timeline Of Tasks

The Data Manager

Data Managers work with the Survey Designer to ensure the questionnaire used in the interviews will generate relevant data for the analysis phase. They also create variable mappings and define the format in which output is generated. It is their responsibility to track the progress of data collection and to communicate it to the rest of the team.

Timeline Of Tasks

Statistical Expertise

Having the roles and responsibilities above provides a team with a great start towards achieving a successful data collection project. However it is key to ensure that the data collected is representative of the population we are interested in. This is where having some statistical expertise is crucial. If the organisation already has this expertise internally that is a great asset. If this is not present however, this can still be obtained by hiring an external statistician or working with other organisations with statistics

expertise from national statistics offices, particularly if the goal of the data collection project is to make it openly accessible.

The descriptions of the roles above are just a summary. For more detailed explanation of the responsibilities, the learner should check out ***School of Data, The Easy Guide to Mobile Data Collection, 2016*** which lays out these roles in much more detail.

Activity 2: Defining Scope, Stakeholders, Budget and Timeline

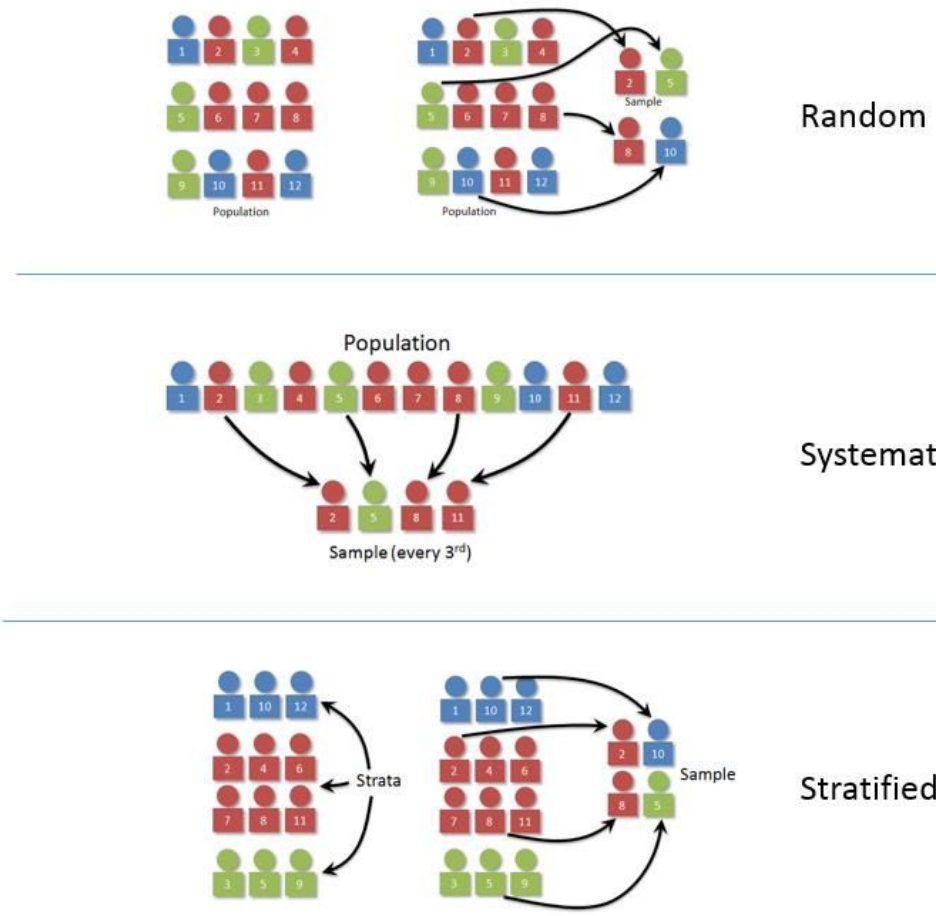
Lesson 3: Methodology and Survey Design

Defining Methodology Characteristics

This stage involved define the structure the data collection project will take. It includes several elements such as deciding on what survey method to use such as questionnaires, interviews or direct observation. Another key aspect of this stage is deciding on what sampling method to use. Some of the common types are random sampling, systematic sampling and stratified sampling. Another key consideration during this stage is how to ensure the privacy, security and confidentiality of everyone

involved in the data collection process. This is particularly important especially in open data collection projects. Failure to demonstrate a clear process and planning for privacy, security and confidentiality could be make or break for your data collection project.

Sampling



In ***random sampling***, every item has an equal chance of being included whereas for ***systematic sampling***, every Nth record is selected. ***Stratified sampling*** on the other

hand involves pulling your sample based on another underlying number – such as population. Rather than pouring all the records for four counties in a pot and pulling randomly – you pull a random sample from each county.

Privacy, Security and Confidentiality

Designing Survey

With a properly structured survey methodology in hand now, you can focus on designing the survey which will be used by the data collectors. This becomes relatively easier since we have the characteristics we want our data collection project to produce, based on the defined methodology. The first stage of designing a survey is producing a ***survey schema***.

Survey Schema

A survey schema is a document that provides a breakdown of each survey question, response format or options and intended objective. It also shows the flow questions such as how the choice of particularly responses could lead to the next questions. Another important function of a schema is to assess the potential sources of error and device techniques to reduce them. For example, the survey team should specify using a number field with a range validation constraint for a question on age.

Schemas are a crucial part of the survey design stage since they serve as an opportunity to think through questions and align them with the survey methodology. When done properly they speed up the development of the final survey as all the steps and flow will have been thought through by then.

This stage is started by the survey designer and could use a simple word processing software such as Microsoft Word, Libre/Open Office Writer or Google Docs. It is recommended to start out simple and build up complexity through several iterations. Using collaborative online tools such a Google Docs can also ease the process of getting feedback on schema from team members. A schema when developed gradually and properly requires only a few changes to be turned into the final survey used by the data collectors.

QUESTIONNAIRE

Women Environmental Programme (WEP), in collaboration with the National Bureau of Statistics (NBS), with support from Open Knowledge Foundation, under the African Open Data Collaboration Fund (AODCF) is conducting a survey in three (3) Area Councils of the Federal Capital Territory, Abuja - AMAC, Gwagwalada, and Kuje Area Councils. The aim of this survey is to gather relevant information that will be used to promote effective budget administration, transparent and accountable local governance of the three Area Councils.

SECTION A – GENERAL INFORMATION

Name of community_____

Area Council_____ Ward_____

Sector (1=Urban 2=Rural) _____ ☐

A0. Name of Respondent (Optional)_____

A1. Role in community ☐

1=Community Head 2=Youth Leader 3=Women Leader 4=Ordinary member 5=Other (Specify)_____

A2. Sex ☐

1=Male 2=Female

A3. Age (in completed years) ☐☐

A4. Marital Status ☐

1=Single 2=Married 3=Divorced 4=Widow 5=Widower

A5. Occupation ☐

1=Civil service 2=Business 3=Student 4=Unemployed 5=Farming 6=Other (specify)_____

A6. Do you have any physical challenge? ☐

1=Yes 2=No **--> SKIP TO A8**

A7. What kind of physical challenge? ☐

1=Deaf 2=Dumb 3=Cripple 4=Blind 5=Amputee 6=Other_____

A8. How could your existence in this community be described? ☐

1=Born here 2=Migrated here

A9. How long have you lived in this community? ☐

1= < 1year 2=1-5years 3=6-10years 4=11-15years 5=16-20years 6=21-25years 7=26-30years 8=30+ years

A10. If 1 in A9, how many months? ☐ ☐

First page of survey schema for WEP data collection project

Turning Schema into Questionnaire

Once the schema has been produced, the next step involved turning this into the final questionnaire for your project. The first step in this stage is for the team to decide on the best tool to use for their data collection. For mobile data collection this means selecting the right data collection software and devices. Once that is done, it is just a matter of transferring the the schema into the suitable format for the data collection software.

Mobile Data Collection

Decades ago, the use of papers and pen was a painstaking and very expensive effort to collect data. Most of us have experienced paper forms getting wet or damaged, or receiving paper forms that were barely answered. But as the age of smartphones and tablets arrived, mobile-based data collection technologies[1] have also gained a huge following. The use mobile data collection tools has also improved the conduct of surveys and assessments. Some of the advantages of using mobile data collection are:

- Most people are using smartphones for SMS messaging and have access to mobile data connection.
- In the absence of laptops and desktop computers, smartphones are cheaper and easier to use.
- Mobile-based data collection compared with the use of paper forms, lessens the risk of losing the data when paper forms are damaged or lost.

[1] https://schoolofdata.org/2014/12/16/mobile-data-collection/

Overview of Mobile Data Collection Tools

- Magpi
- Viamo Mobile Surveys
- Mobile Forms
- Open Data Kit (ODK)
- Kobo Toolbox
- GeoODK
- Cadasta Platform

Open Data Kit

One of the tools that was frequently used by information managers is the Open Data Kit (ODK). While there is a huge variety of online and offline data collection tools, ODK has gained a lot of users because it is free, open source, easy to use and can be used both offline and online. Since ODK is a free and open source set of tools which help organizations author, field, and manage mobile data collection solutions, ODK in itself has evolved in several platforms and formats such as Kobo Toolbox which I prefer to use, GeoODK, KLL Collect, Formhub, Enketo, each one seeking to customize the use of ODK according to their own needs.

ODK provides an out-of-the-box solution for users to:

1. **Build** a data collection form or survey (XLSForm is recommended for larger forms);

2. **Collect** the data on a mobile device and send it to a server; and
3. **Aggregate** the collected data on a server and extract it in useful formats.

Data Collection Using Kobo Toolbox

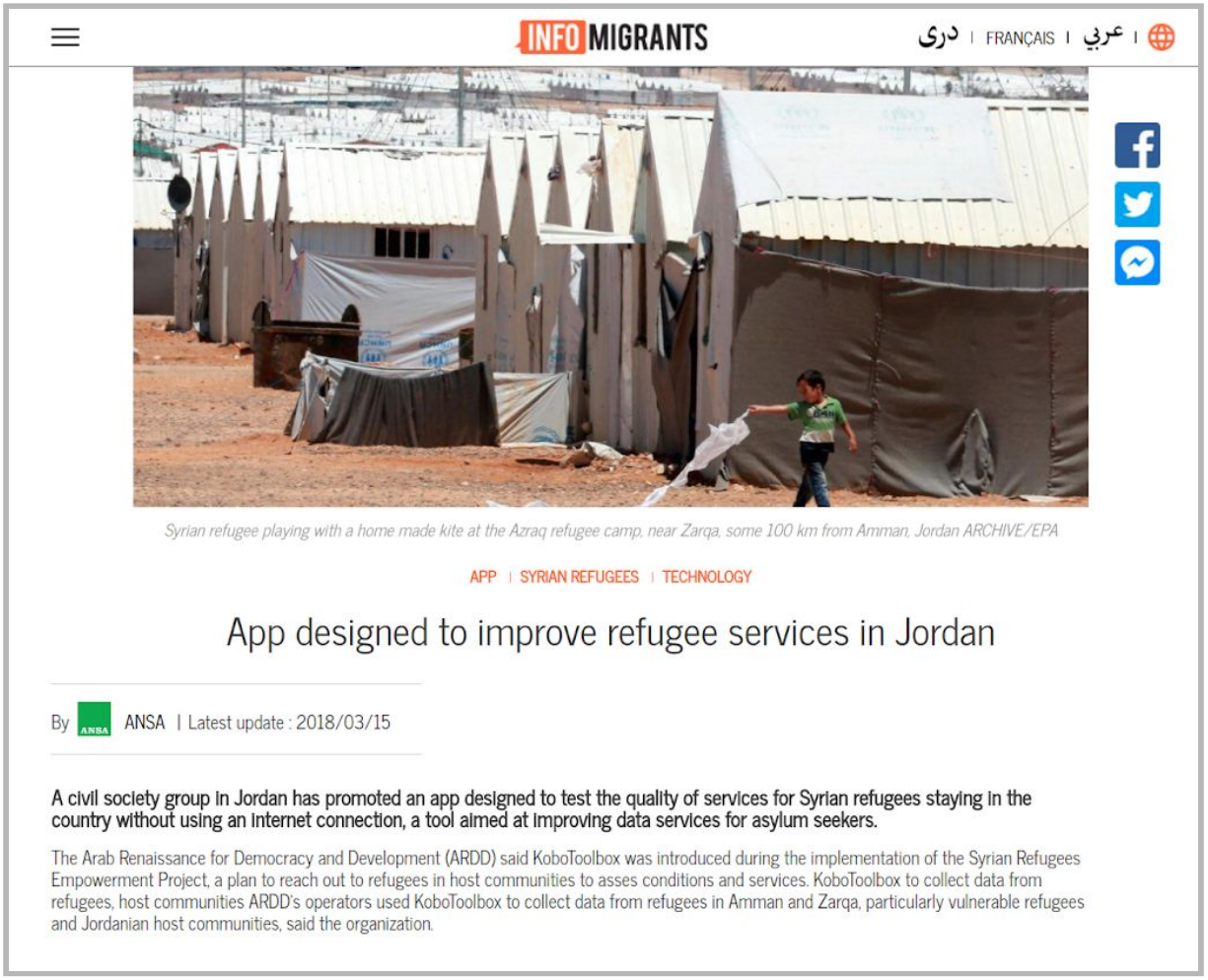
For the rest of this module, we will be using Kobo Toolbox as one of the many platforms in which ODK forms are built, collected and aggregated for better data collection and management. According to Kobo Toolbox, acknowledging that many agencies are already using ODK, a de facto open source standard for mobile data collection, KoBo Toolbox is fully compatible and interchangeable with ODK but delivers more functionality such as an easy-to-use formbuilder, question libraries and integrated data management. It also integrates other open-source ODK-based developments such as formhub and Enketo. Kobo Toolbox can be used online and offline. You can share the data and monitor submissions together with other users and it offers UNLIMITED use for humanitarian actors.

We will be creating a digital questionnaire with an online, open-source mobile data collection tool called KoBo Toolbox.

Before moving forward, make sure you have access the following devices and resources:

- Internet connection
- Personal computer
- Android-based mobile device (smartphone or tablet)

Getting Started



INFO MIGRANTS

دری | FRANÇAIS | عربي

Syrian refugee playing with a home made kite at the Azraq refugee camp, near Zarqa, some 100 km from Amman, Jordan ARCHIVE/EPA

APP | SYRIAN REFUGEES | TECHNOLOGY

App designed to improve refugee services in Jordan

By ANSA | Latest update : 2018/03/15

A civil society group in Jordan has promoted an app designed to test the quality of services for Syrian refugees staying in the country without using an internet connection, a tool aimed at improving data services for asylum seekers.

The Arab Renaissance for Democracy and Development (ARDD) said KoboToolbox was introduced during the implementation of the Syrian Refugees Empowerment Project, a plan to reach out to refugees in host communities to asses conditions and services. KoboToolbox to collect data from refugees, host communities ARDD's operators used KoboToolbox to collect data from refugees in Amman and Zarqa, particularly vulnerable refugees and Jordanian host communities, said the organization.

With the ongoing conflict in Syria forcing many Syrians to flee their country to neighbouring and distant countries, it has become crucial for international organisations to provide improved services to meet their needs. In Jordan, the Arab Renaissance for Democracy and Development (ARDD) has used KoboToolbox to reach out to refugees in host communities to asses conditions and services.

Infomigrants reports that, Aid workers "fill out reports for monitoring and evaluation and research especially at the sites of sessions, as well as the use of the platform to track GPS locations and create a comprehensive map of session

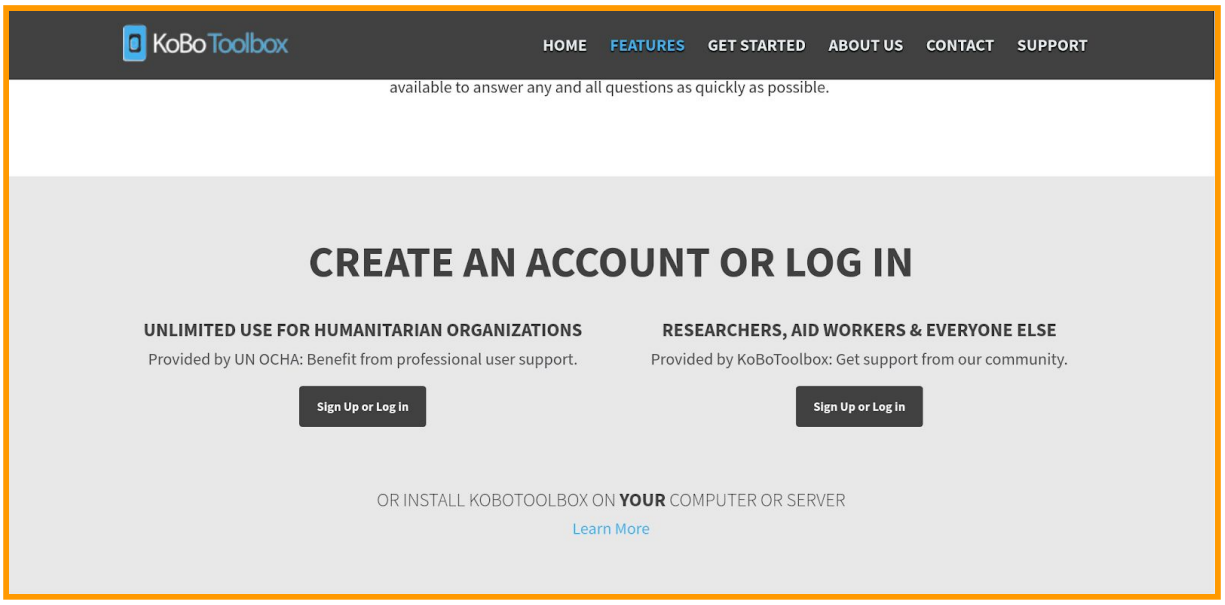
locations and refer individual and legal cases to ARDD's legal team and relevant organizations for help". Read the full article from Infomigrants here: http://www.infomigrants.net/en/post/8091/app-designed-to-improve-refugee-services-in-jordan

Task 1: Create A KoBo ToolBox Account

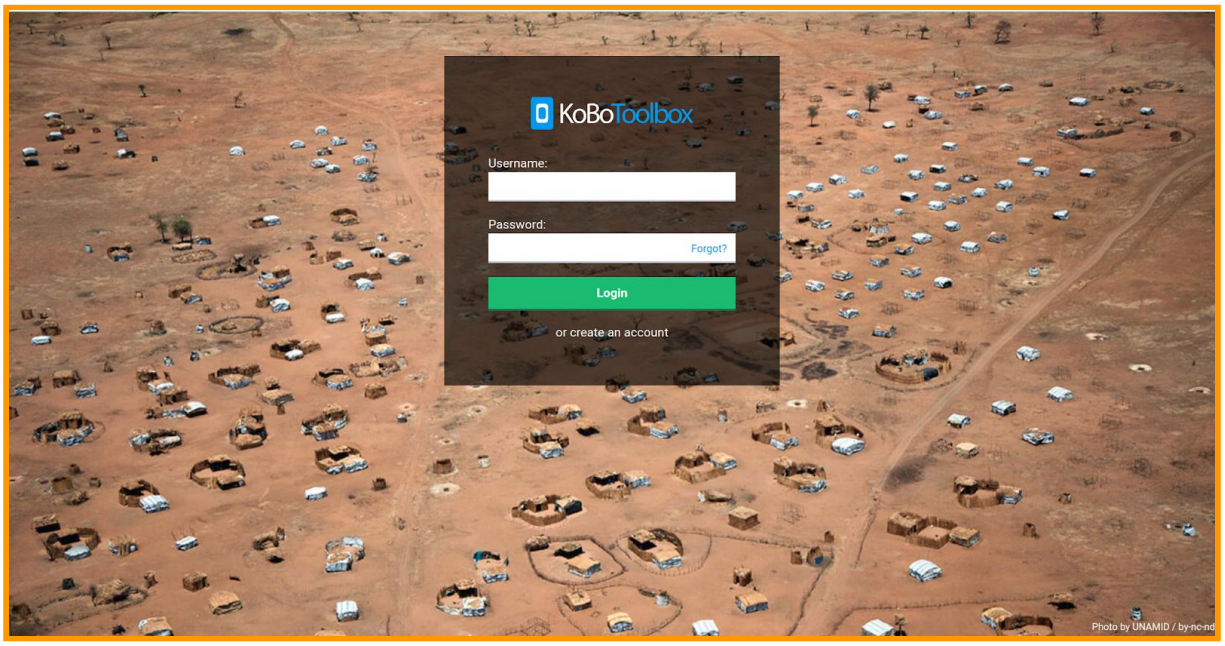
Let's start by creating a KoBo Toobox account. Here are the steps:

1. Open any browser of your choice and visit http://www.kobotoolbox.org/. Take a few minutes to read through the website for any information that interests you.

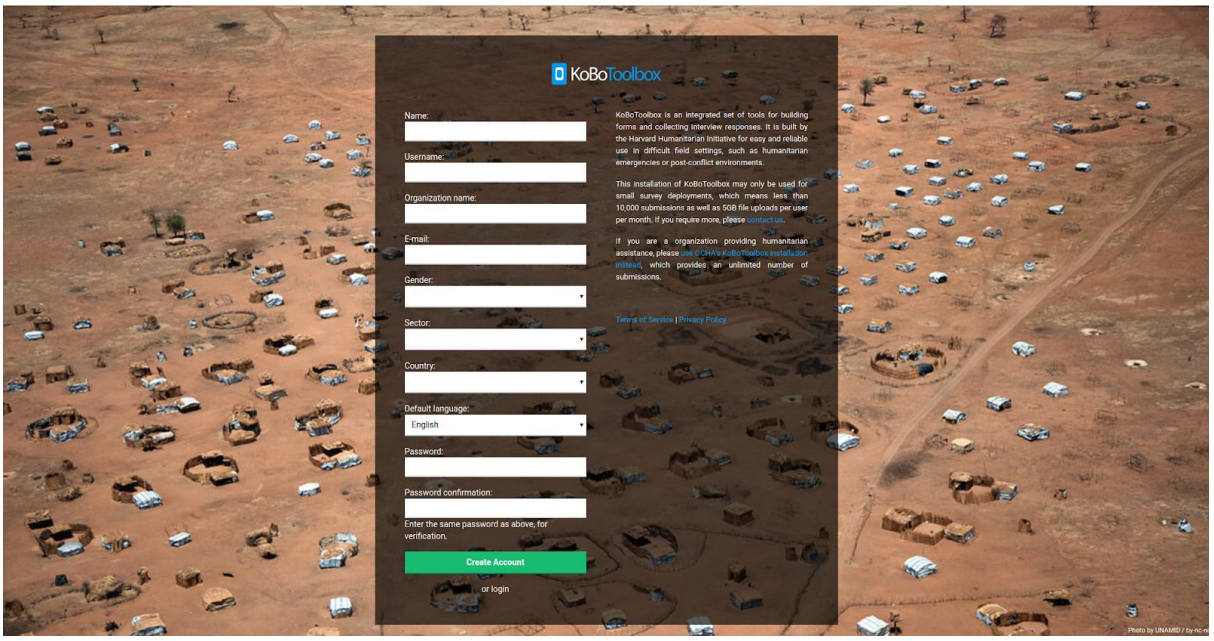
2. Scroll down to the middle of the page to the section that say **"Create An Account or Log In"**.
3. Click the **Sign Up or Log In** button under the **Researchers, Aid Works & Everyone Else** section.



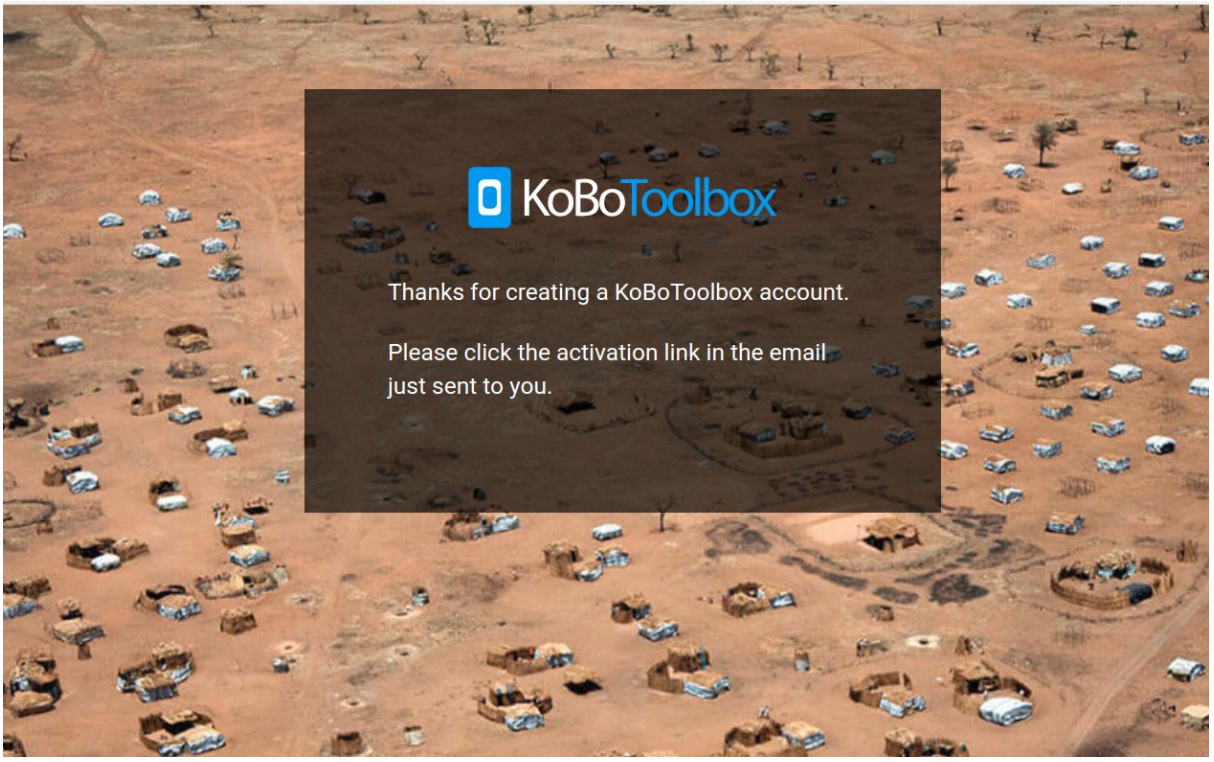
4. If you already have a KoBo Toolbox account, enter your username and password to Login. For new users, click the **create an account** option.



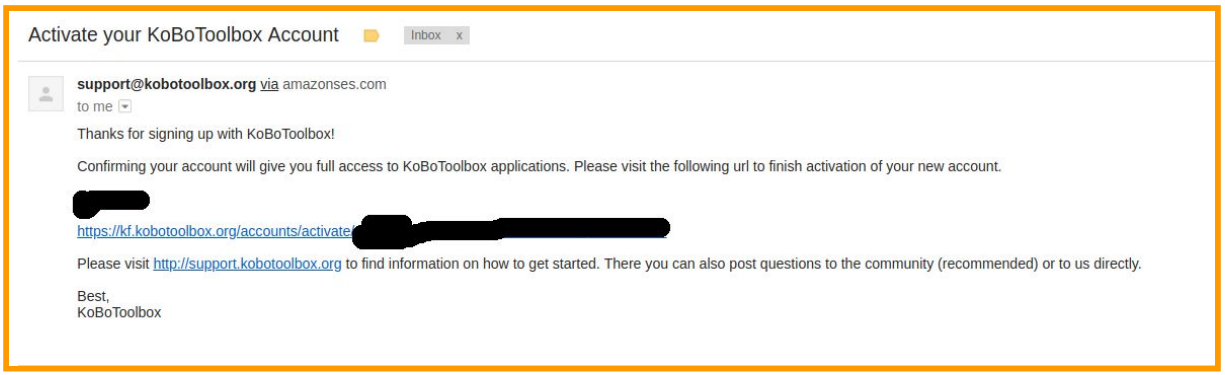
5. On the sign up page that opens up, fill in your details including (Name, Username, Organization name, Email, Gender, Sector, Country). After choosing and confirming your password, click **Create Account**.



6. Check the email you entered in the registration form to confirm your account details and activate your account.



7. Click the activation link in you email.



8. This opens your home page directly. If this does not happen, enter your username and password to access your account.