Open Data Kit: Mobile Data Collection Framework For Developing Countries

Patrick Loola Bokonda, Khadija Ouazzani-Touhami, Nissrine Souissi

Abstract: To evaluate and improve the delivery of health care services and programs; to collect and share information on agro-biodiversity; to fill the gaps in the data of artisanal or industrial fisheries, the use of computer tools of data collection is a major asset. It is now obvious that most of these tools are difficult to use in developing countries. And this is because of the complexity of use, either because they are commercial solutions or simply because these tools require a stable internet connection. This study is the result of thorough research on Open Data Kit (ODK), a suite of tools fully compatible with the context of developing countries and fully open source software. This article traces the research work done for the design, implementation, and improvement of ODK. A description of the different ODK extensions that can be found in the literature is given in addition to the reasons that explain the success of ODK in developing countries. To highlight the utility of ODK, an analysis of the areas of use of ODK has been made. This analysis concludes that ODK is mainly used to improve the intervention effectiveness of health system agents in developing countries.

Keywords: Data Collection, Mobile Data, ODK, Open Data Kit.

I. INTRODUCTION

The new global gold is called 'data'. Today, the effectiveness of a project, an organization or a country is measured by its ability to collect, store and analyze, as quickly as possible, a large amount of data. This race for data collection has led to the emergence of several methods of data collection, integrating new information and communication technologies, some more complex than others [1], [2], [3], [4], [5].

But, what is perhaps less known is that when it comes to the use of digital technologies, developed and developing countries do not have the same challenges.

The purchasing power of a population and organizations, the availability of an Internet connection, the level of education of a population as well as the technical level of the agents of organizations and companies are some of the major difficulties that we find in developing countries.

In such a context, computer science engineers and researchers must propose tools that can be used despite all these difficulties. In other words, tools that adapt to the context of developing countries.

Over the years, after the advent of smartphones, several systems for collecting data by mobile devices have emerged, such as CyberTracker [6], Pendragon [7], FrontlineForms [8], EpiSurvey [9], Comcare [10], JavaRosa [11].

But unfortunately, most of these systems are not adapted to the context of developing countries. Three main reasons explain the inadequacy of these systems:

- Some of these systems (CyberTracker and Pendragon) are commercial solutions.
- Some of these systems are difficult to deploy and use. For example, FrontlineForms, EpiSurvey, Comcare and JavaRosa require a certificate or electronic signature before any interaction with storage, hardware accessories or networking can occur [12].
- Some of these systems require of a stable internet connection.

Unlike other systems, Open Data Kit (ODK) [12], [13] is a platform that has been specifically designed to fit the context of developing countries.

This article is the result of extensive research on this tool suite (ODK). This study has 3 research objectives. To analyze work involved in the design and implementation of the ODK software suite, to synthesize studies that aim to improve the ODK or to create an extension of it, and to summarize research that uses ODK to meet a specific need.

To complete this study, two main research questions (RQs) were asked:

RQ1: What are the extensions of ODK in literature?
RQ2: What are the fields of use and some examples of application of ODK?

This article is subdivided as follows: section 2 presents the method followed during the research; Section 3 gives the result obtained after the application of the research method; Section 4 presents the data extracted from the research work studied while providing the answer to the two research questions; and section 5 includes the conclusion of this research as well as a proposal for possible improvement work of ODK in order to increase its adaptation to the context of developing countries.

II. RESEARCH METHODOLOGY

The research method followed is summarized in two stages, described in Fig. 1:

- Stage 1: Four steps whose purpose is to prepare the necessary elements for the research, selection and exclusion of articles.
- Stage 2: Three steps using the elements of Stage 1. The research was made in the following search engines and scientific databases: Google Scholar, Springer Link and Science Direct. The research keywords used are: Open Data Kit (ODK); Open Data Kit (ODK) extension; Open Data Kit (ODK) architecture; Open Data Kit (ODK) workflow; Open Data Kit (ODK) visualization. After obtaining a considerable number of papers, elimination and selection was made on the basis of the following criteria and means:
  (1) Use of the relevance ranking tools provided by the scientific databases;
  (2) Reading the title and/or abstract to identify articles that answer one of the research questions;
  (3) Prioritization of articles published after April 2008: date of launch of the ODK project by Google [14];
  (4) Prioritization of articles published in peer-reviewed scientific journals;
  (5) Selection of articles that concern an ODK extension or enhancement;
  (6) Selection of articles that concern the use of ODK;
  (7) Careful examination of the abstract of the articles;
  (8) Full reading of the content of articles.
The Fig. 2 summarizes the steps mentioned above.

III. RESULTS
The research made in the abovementioned scientific databases, using successively the search keywords selected in each of them gave 4,261 occurrences for all the databases combined. Table- I gives the number of articles obtained by database and by combination of keywords.

Table- I: Number of articles obtained by database and combination of keywords

<table>
<thead>
<tr>
<th>Search engine or Scientific database</th>
<th>Keyword</th>
<th>Number of obtained articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Scholar</td>
<td>Open + Data + Kit + ODK</td>
<td>2420</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>Open + Data + Kit(ODK) + extension</td>
<td>571</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>Open + Data + Kit(ODK) + architecture</td>
<td>530</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>Open + Data + Kit(ODK) + workflow</td>
<td>238</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>Open + Data + Kit(ODK) + visualization</td>
<td>114</td>
</tr>
<tr>
<td>Total of obtained articles from Google Scholar</td>
<td></td>
<td>3873</td>
</tr>
<tr>
<td>Springer Link</td>
<td>Open + Data + Kit + ODK</td>
<td>157</td>
</tr>
<tr>
<td>Springer Link</td>
<td>Open + Data + Kit(ODK) + extension</td>
<td>37</td>
</tr>
<tr>
<td>Springer Link</td>
<td>Open + Data + Kit(ODK) + architecture</td>
<td>21</td>
</tr>
<tr>
<td>Springer Link</td>
<td>Open + Data + Kit(ODK) + workflow</td>
<td>6</td>
</tr>
<tr>
<td>Springer Link</td>
<td>Open + Data + Kit(ODK) + visualization</td>
<td>19</td>
</tr>
<tr>
<td>Total of obtained articles from Springer Link</td>
<td></td>
<td>240</td>
</tr>
</tbody>
</table>

Starting from the ninety-two articles previously selected, we applied criteria (4), (5) and (6) to them, to extract those which will be the subject of an in-depth study, which will then serve as a thread for future works.
In addition to these criteria, a careful analysis of the summaries according to criterion (7) and the full reading of the content of each of these ninety-two articles according to criterion (8) identified thirty (30) articles which have been adopted as objects of study for the present work. We will present our synthesis in the following section.
IV. RESEARCH SYNTHESIS AND DISCUSSION

In the available literature, the oldest ODK article dates back to October 2009 [12] and the most recent to January 2019 [15]. The thirty articles selected explore several aspects of the ODK tool suite.

In this section, for a better analysis, the articles have been grouped into four main themes. Table- III classifies the articles by theme. After the in-depth study of these thirty articles as well as the reading and the re-reading of the official documentation [44], several observations were made.

The ODK tool suite has been designed from the ground up to respond to the need of data collection in developing countries [12], [16]. Designed initially with three tools [12]: ODK Collect, ODK Aggregate and ODK Manage; the framework quickly evolved to offer four [16]: ODK Collect, ODK Aggregate, ODK Voice and ODK Build.

At this point, although correctly answering a large majority of applied cases, the wide adoption of ODK has showed the limitations of the first version of the suite. This first version was rather intended for users with no strong technical skills, it was therefore easy to deploy and use. The main limitation of this version is that its customization is quite limited. It was therefore difficult to use for studies with complex workflows.

Table- III: Studied articles by theme

<table>
<thead>
<tr>
<th>References</th>
<th>Theme</th>
<th>Number of articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>[12], [16], [17], [18], [19]</td>
<td>Presentation of a new ODK architecture.</td>
<td>5</td>
</tr>
<tr>
<td>[20], [21], [22], [23], [24], [25], [26]</td>
<td>Improvement or Creation of a new component of the ODK tool suite.</td>
<td>7</td>
</tr>
<tr>
<td>[27], [28], [29]</td>
<td>Proposal for an ODK extension.</td>
<td>3</td>
</tr>
<tr>
<td>[30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43]</td>
<td>Use case of ODK.</td>
<td>15</td>
</tr>
</tbody>
</table>

This is what prompted the ODK team to design a new ODK tool suite, originally named ODK 2.0 [17], [18]. This new suite is not intended to replace the first, but to meet the need to implement mobile data collection applications for studies with complex workflows. It is important to note that the versions of the tools of both suites continue to evolve together until today. To avoid confusion causing users to automatically choose the ODK 2.0 suite believing that it is necessarily better than ODK, ODK 2.0 was renamed ODK-X in July 2019 [44].

To date, the ODK suite has six main tools [44]: ODK Collect, ODK Aggregate, ODK Central, ODK Build, ODK XLSForm and ODK Briefcase, while ODK-X has seven: ODK Tables, ODK Survey, ODK Services, ODK Application Designer, ODK Suitcase, ODK SyncEndpoint, ODK Aggregate Extension Tables.

Another observation from this analysis is related to the architecture of ODK itself. ODK is a suite of independent but communicating tools. It is the same for ODK-X. On the other hand, the tools of ODK are incompatible with those of ODK-X. This architecture allows engineers and researchers:

- To propose new tools compatible with the ODK suite or with the ODK-X suite [20], [21], [22], [23], [24];
- To make improvements to a single component of the ODK suite or the ODK-X suite, without having to touch all the others [25], [26];
- To use one or some of the components of the ODK or ODK-X suite without having to use the whole suite. The tools of the two suites can even be used with other technologies that do not belong to ODK [31], [32], [34], [38], [40], [41];
- To extend a component of the ODK or ODK-X suite by adding new functionalities and then to use this extension with the other components of the suite [27], [28], [29].

15 papers, whose main purpose is to present ODK applied cases, have shown that health is the most frequent domain where ODK is used. Health alone accounts for 11/15 papers, while there are only two in the field of agriculture, one in fisheries and one in social domain. Fig. 3 presents as a...
Open Data Kit: Mobile Data Collection Framework For Developing Countries

Fig. 3: Rate of papers by application area of ODK

Although ODK is adaptable to studies in different fields, most of the time, tools of ODK or ODK-X suite are used for medical investigations.

The use of the tools of ODK for data collection has helped to improve many health programs and systems by enabling health workers to rapidly collect and analyze large amounts of data in many countries.

This study identified nine countries in which ODK was used. These are: Kenya, Mali, India, Nigeria, Ethiopia, Madagascar, Tanzania, Mozambique and the Dominican Republic.

One of the most important findings of this study is that all applied cases found in the research works of ODK have been realized in a region of a developing country.

The success of ODK tools in these countries can be explained by four reasons:

- Its open source character: Users and organizations can install and use ODK tools without spending anything [13];
- The ability to collect data without an internet connection and submit them as soon as a connection is available: This aspect of ODK is very useful in developing countries where the internet connection is not always stable;
- The ability to submit data using SMS. This feature is one of ODK’s greatest advances for better adoption in developing countries. It was added for the first time in 2011 [20]. This feature is only valid from the ODK Collect v1.16.0 release. Unfortunately in October, 2018, the Google Play Store restrictions on high-risk data led to the removal of this feature from the version 1.20 of ODK Collect. However, versions prior to ODK Collect v1.20 still have this functionality [44];
- ODK can be customized and deployed by a non-programmer or a person with limited technical skills.

We conclude this section by presenting the three ODK extensions found in the literature:

- GeoODK [29]: The ODK extension that offers the possibility of collecting and storing geo-referenced information. It is an extension of ODK Collect;
- ODK-Ext [27]: An extension of ODK Collect. The work was to add two new features to ODK Collect. The first is the addition of the concept of pre-filled forms using previously collected data. The second is the addition of the feature of sending data by SMS;
- Mezuri [28]: An extension of ODK 2.0 (ODK-X). Mezuri is a comprehensive platform for collecting, managing and distributing data designed for use by managers of global development programs and researchers.

V. CONCLUSION AND FUTURE WORKS

Open Data Kit is a mobile data collection framework specifically designed for use in developing countries. This article is the result of a thorough study of this tool suite. The Google Play Store has banned several applications, including ODK Collect, from using SMS access authorization [45]. This decision was made by Google to force applications to respect the privacy of users. This led to the removal of ODK Collect SMS data sending functionality from Version 1.20.

Given the importance of this feature for developing countries, a possible improvement is to find a secure technical way for ODK to reuse the SMS sending functionality while respecting the conditions set by the Google Play Store respecting the privacy of users.

REFERENCES

Khadija Ouazzani-Touhami is a fulltime professor at the Mines Rabat School, Morocco. She is economist statistician engineer - IT option of the National Institute of Statistics and Applied Economics – Rabat – Morocco. Her research interests include process engineering and optimization, business process management, data lifecycle, smart data, simulation, and information system.

Nissrine SOUISSI is a fulltime professor at the MINES-RABAT School, Morocco. She obtained a Ph.D. in computer science from the UPEC University in 2006, France and an Engineer degree from Mohammadia School of Engineers in 2001, Morocco. Her research interests include process engineering, business process management, databases, data lifecycle, smart data, hospital information system, and