



library@chuka.ac.ke; www.chuka.ac.ke

APPLICATION OF OPEN SOURCE TOOLS AND CLOUD COMPUTING TECHNOLOGIES IN REAL-TIME DATA COLLECTION AND ANALYSIS

Kirui, T.K.

P. O. Box 15964-20100, Nakuru-Kenya

Email: shakes2way@gmail.com, Mobile: +254712838653

Citation:

Kirui, T. K. (2015). Application of open source tools and cloud computing technologies in real-time data collection and analysis. Isutsa, D. K. (Ed.). *Proceedings of the First International Research Conference held from 29th to 31st October, 2014 in Chuka University, Chuka, Kenya, 267-274 pp.*

ABSTRACT

Cloud Computing technology is a new phenomenon that though fully utilized and implemented in developed countries, its power is yet to be tapped in a number of developing countries. There are different kinds of data that can be entrusted to cloud computing providers that would reduce operating costs and save on time and financial resources. This paper sheds light on a number of cloud computing services such as Google App Engine and Open Source tools such as Open Data Kit and how the two can be used in effective data collection and analysis for organizations both governmental and non-governmental as well as business organizations. Google App Engine offers a combination Infrastructure as a Service (IaaS) and Software as a Service (SaaS) to be able to provide both a mobile application and web application that can be used to receive data for analysis from field research; every instance of the application can be fully customized to meet the specific data collection needs

Key words: *Google App Engine, Open Data Kit, Real-time, Mobile Application, Web Application.*

INTRODUCTION

The crucial steps of every research process are the collection of data and its analysis to formulate accurate and timely information that can be used to make critical decisions. Data collection needs to be conducted prudently to ensure that respondents provide credible data for statistical analysis. Considering all these facts this means that proper care needs to be taken from data collection through to the formulation of a thesis from the research undertaken. A researcher can now easily tap into the extensive power provided by the combination of cloud computing technologies and open source tools.

This paper aims to explain the concept of open source and cloud computing as one of the indomitable technological phenomenon that have been combined can help develop solution prototypes that have been tried and tested and has worked for the good of the end user. Open Source is the term that is used to denote software whose source code is freely available and accessible to anyone for assessment, enhancement and/or modification. Source code is the section of software that is normally available only to the persons who developed the software. It is the part that programmers manipulate in order to change how part or whole of the program works. [10]. Cloud Computing is a concept whereby instead of services being provided through a direct connection to a conventional server, services are provided via the internet through web-based tools and applications [1].

The tools employed in my case study are:

- Open Data Kit Suite [8] - an extensible, open-source suite of tools designed to build information services for developing regions. [2]

- ODK Collect – A smart phone client application that runs on smart devices that run on Android an open source operating system for mobile devices.
- ODK Build – An open source form designer that is used as a tool to design and create Xform that are later downloaded to ODK Collect from ODK Aggregate.
- ODK Aggregate - Aggregate provides a server repository to manage collected data, provide standard interfaces to extract data, integrate with existing systems and is designed to be a generic data storage service that will run on the user’s choice of computing platform. It can receive data from smart devices through the ODK Collect.
- Google App Engine – A cloud service that enables users to upload and run applications as well as giving real time statistical information on bandwidth usage and application access.

MOTIVATION

Having tried and tested these tools in my undergraduate final year project it became vivid that these tools are cost-effective and essential in any kind of data collection process that is intended to produce concise and timely information. A researcher without any programming expertise is able to re-design, customize and deploy these tools as perfect agents to enable him/her to have real-time data collected and input into an aggregate system ready for statistical analysis using statistical software such as SPSS.

The strong belief and closed mindset of researchers in data collection through filling out paper forms and gathering responses through rigidly designed questionnaires delimits the power of research, creativity and hence innovation. The inception of Huduma Kenya by the Kenyan Government to provide its citizens access to public services through integrated technology platforms from One Stop Shop citizen service centres called Huduma Centres located at Postal Offices of major towns, has shown the commitment of the government in embracing Information Technologies. The Government also intends to come up with an online services portal [3].

Automation of data collection processes will surely ensure that the government delivers its services much faster since paperless and cashless technologies are fast becoming a phenomenon in the country. The innovation of cashless payment services in buses and matatus [5] and even payment of parking fees within Nairobi City [7] has shown clearly that computing facilities are being employed to ease the burden of revenue collection as well as inhibiting cases of corruption.

With all these innovations there is surely the need to gather feedback from citizens on their take on the use of Information technology as well as having a forum where users can suggest improvements. A respondent can therefore download a customized version of the ODK Collect and retrieve the Feedback form and post his/her feedback which can be analyzed in real time.

SYSTEM DESIGN

A. Case Scenario

To demonstrate the use of the ODK Suite and the Google App Engine, consider the following case scenario. Uwezo Fund which is a government loan for Youth and Women Groups needs to collect information about Groups applying for the loan and the composition of the group in terms of membership and what they intend to do with the loan if awarded. In the present scenario:

- 1) Applicant representing the group downloads form for printing or picks it at respective constituency offices.
- 2) She/he fills the form in triplicate.
- 3) Mails the form to the constituency offices or travels there to personally submit it.
- 4) Uwezo Fund Constituency Committee goes through all the forms (some being illegible) to check for relevant and qualified groups.
- 5) The committee submits the forms of qualified applicants to a National Committee for approval.
- 6) Upon approval the committee generates a list for qualified applicants to undertake the Capacity Building Programme.
- 7) After successful completion of the training program the loan is disbursed to the bank account of the group

If Uwezo Fund employed open source tools such as ODK Suite and Google App engine the steps would be:

1. Applicant downloads ODK Collect, retrieves the form and starts filling it right from his/her smart device.
OR Authorized Uwezo Fund Agents would move around collecting applicant information of interested

- groups to the designated Smart device that has ODK Collect installed and with the form retrieved.
2. Complete forms are uploaded to the already deployed ODK Aggregate and the forms can be Analyzed by the committee in real time for eligibility.
3. The committee can submit a list of approved eligible applicants on a daily or weekly basis to the National Committee for approval which can be done as soon as the list is e-mailed.
4. Upon approval the committee generates a list for qualified applicants to undertake the Capacity Building Program.
5. After successful completion of the training program the loan is disbursed to the bank account of the group

B. Tool Design

The ODK Suite need to be designed, tested and re-designed before using it in a real scenario, there are various tools and cloud providers that can help in achieving this namely:

- ODK Build - a web-based graphical designer that allows non-programmers to build applications with complex logic and interactions, we designed that allows users to create applications using a drag-and-drop metaphor. [2]
- Google App Engine – a cloud computing service upon which the ODK Aggregate is deployed for submission of filled forms and retrieval of blank forms.

Figure 1 shows a data collection form for Uwezo Fund Loan Application [9] being designed using ODK Build in drag-and-drop fashion, the final form will later be saved as an XML form that will be uploaded to ODK Aggregate for later unlimited retrieval by the smartphone clients ODK Collect. This is similar to the process of formulating a questionnaire for use to interview respondents.

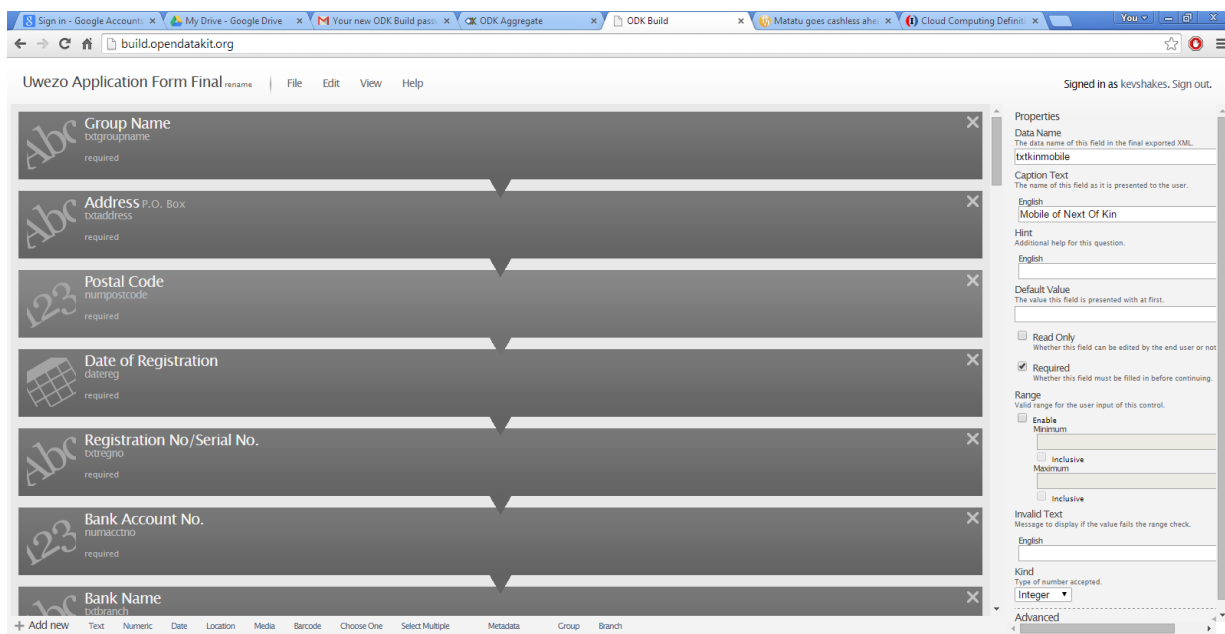


Fig. 1: In ODK Build, prompts appear on the left of the screen while properties appear on the right. Users rearrange prompts using a drag and drop interaction in the web browser.

Figure 2 shows a deployed instance of the ODK Aggregate that accepts form submissions and blank form retrieval for the Uwezo Fund Loan Application Form. [9]

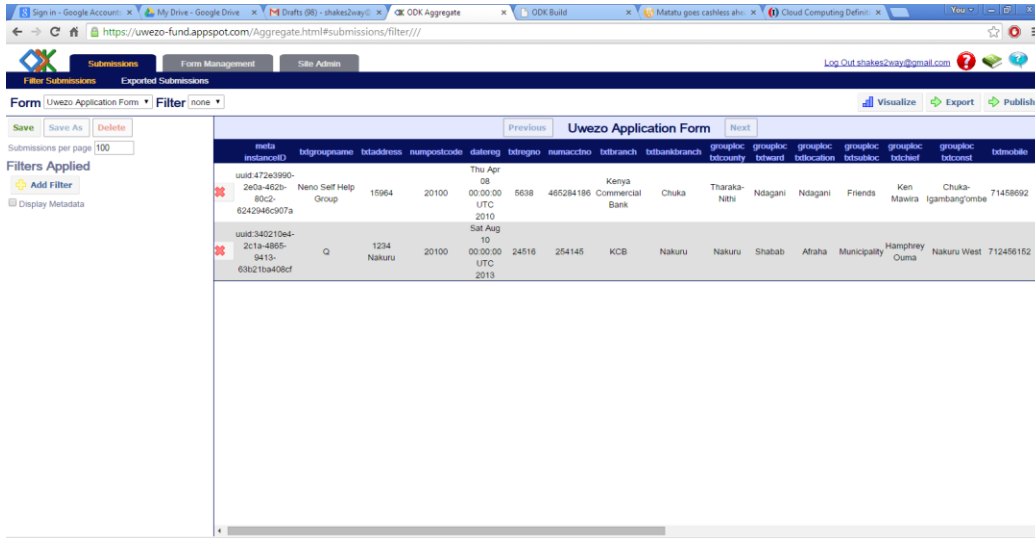


Fig. 2: An instance of ODK Aggregate on Google App Engine’s Cloud Service

SYSTEM IMPLEMENTATION

For implementation of the system, ODK Collect which is a smart phone client is installed on a device running an Android Operating System. Figure 3 a-f shows a series of screen shots of ODK Collect in operation. [6]

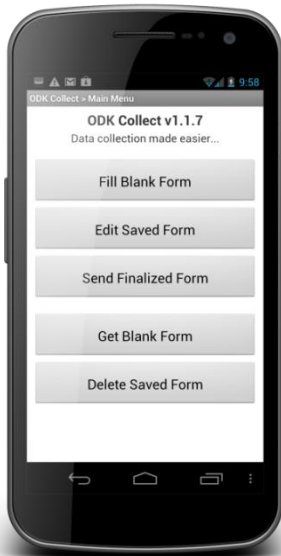


Fig 3.a ODK Collect’s First Screen with command options
 Fill Blank Form – Start filling out an already retrieved form.
 Edit Saved Form – Continue filling a form from where you left.
 Send Finalized Form – Submit a duly filled form.
 Get Blank Form – Retrieve a blank form from ODK Aggregate



Fig 3.b The Settings Menu: URL Configuration. On the settings Menu configure the URL of ODK Collect to point to the URL of where the ODK Aggregate is deployed for retrieval of blank forms and submission of finalized forms to take place.



Fig 3.c After successful connection with ODK Aggregate one is able to retrieve a list of blank forms hosted on the server and select the ones to download. In this case the server has a blank form named Sample. After selecting the for, tap on Get Selected in order to download the selected form.

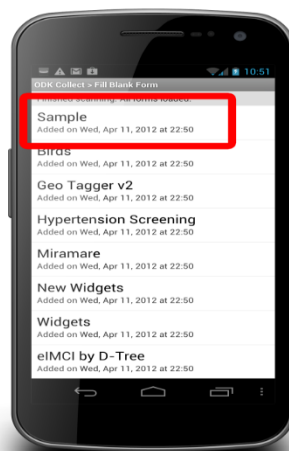
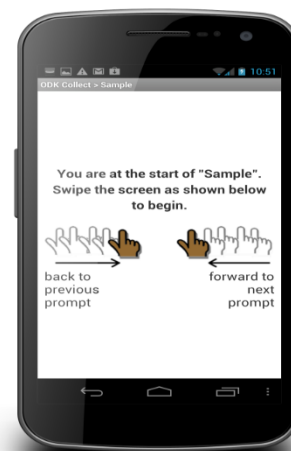


Fig 3.d Once downloaded the form will appear in the list of blank forms when you tap on Fill Blank Form indicated in figure 3.a above. After you tap on Sample(indicated in red) you will have a begin prompt telling you to swipe across either to the left or to the right to start filling the blank form.



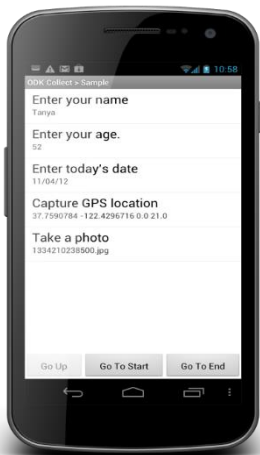


Fig 3.e After all the fields in the form are filled with the required data, as a researcher you can review the details, before marking the form as finalized to make it ready for submission.

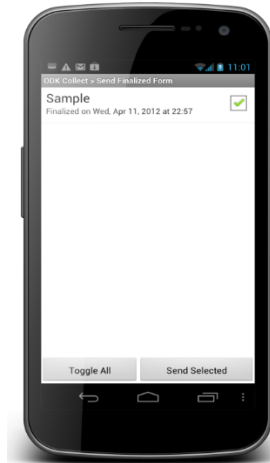
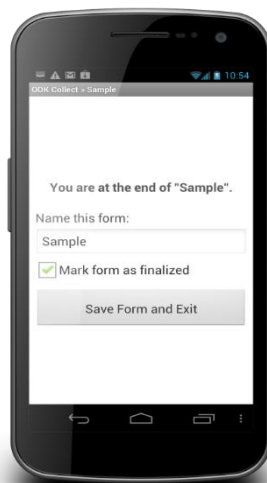


Fig 3.f Tap on "Send Finalized Form" as indicated in figure 3.a above and a listing will appear with all the finalized form that have not been submitted to the ODK Aggregate. Select the Sample Form and tap on "Send Selected". A notification will appear to confirm that the form has been successfully uploaded.



A Summary of ODK Collect

In just six simple steps you are able to understand a simple concept which you can easily use without any additional costs apart from having a smart device with internet connection. The forms can also be filled multiple times before submission such that a researcher can perform a field study where there is no internet connection as long as he/she has already retrieved the blank form. This shows that one does not need any programming expertise in order to make use of the open source tools for data collection.

ODK Aggregate

Without proper deployment of the ODK Aggregate onto the Google App Engine all the functionality and efficiency provided through ODK Collect is impossible to achieve. Figure 4 a-d shows a series of screenshots reflecting on fundamental aspects of the ODK Aggregate. [4]

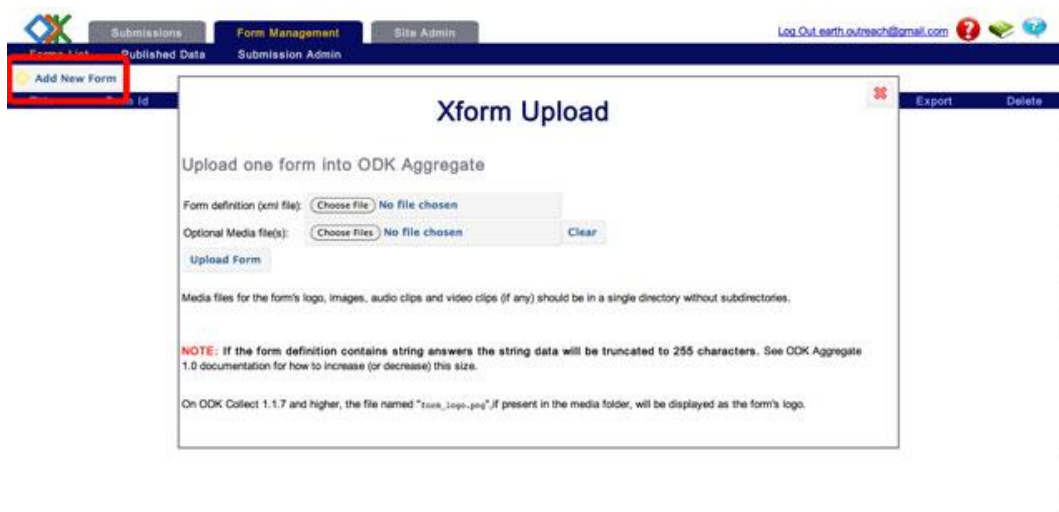


Figure 4 a shows the process of uploading blank form (to be retrieved by ODK Collect) to the Aggregate server

This is the first step after successful deployment which is just a step-by-step wizard using an installer as long as you have a Google ID which is just a Google/Gmail Account. The form is uploaded as a Form definition which is an XML file; one may add any other kind of media that he/she would like an agent to retrieve through ODK Collect that would be relevant to the research being undertaken.

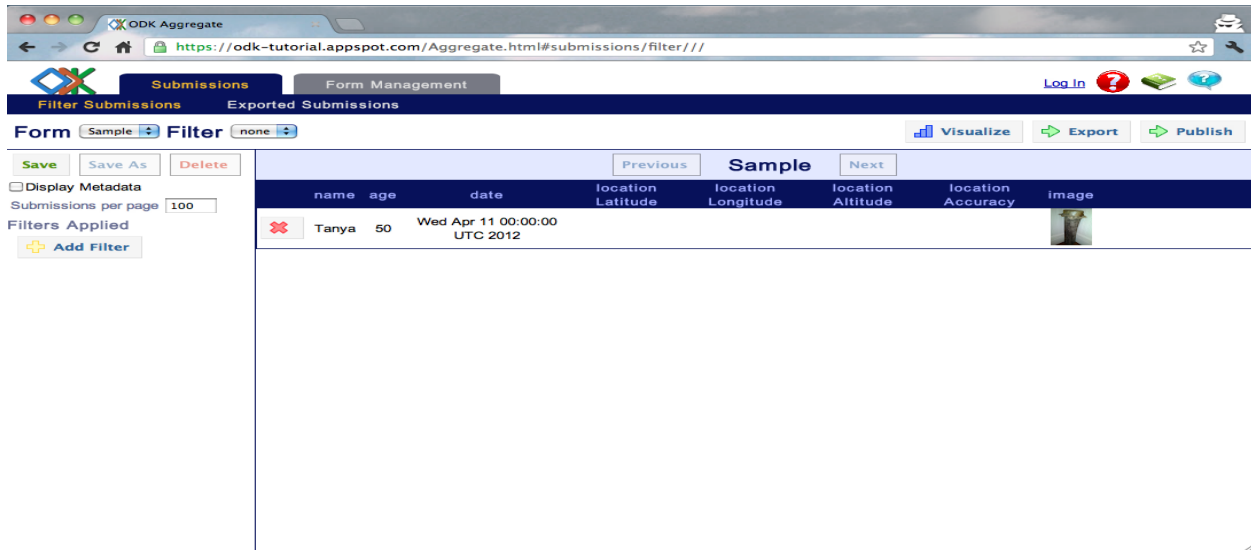


Figure 4 b shows a successful submission of a form to the ODK Aggregate.

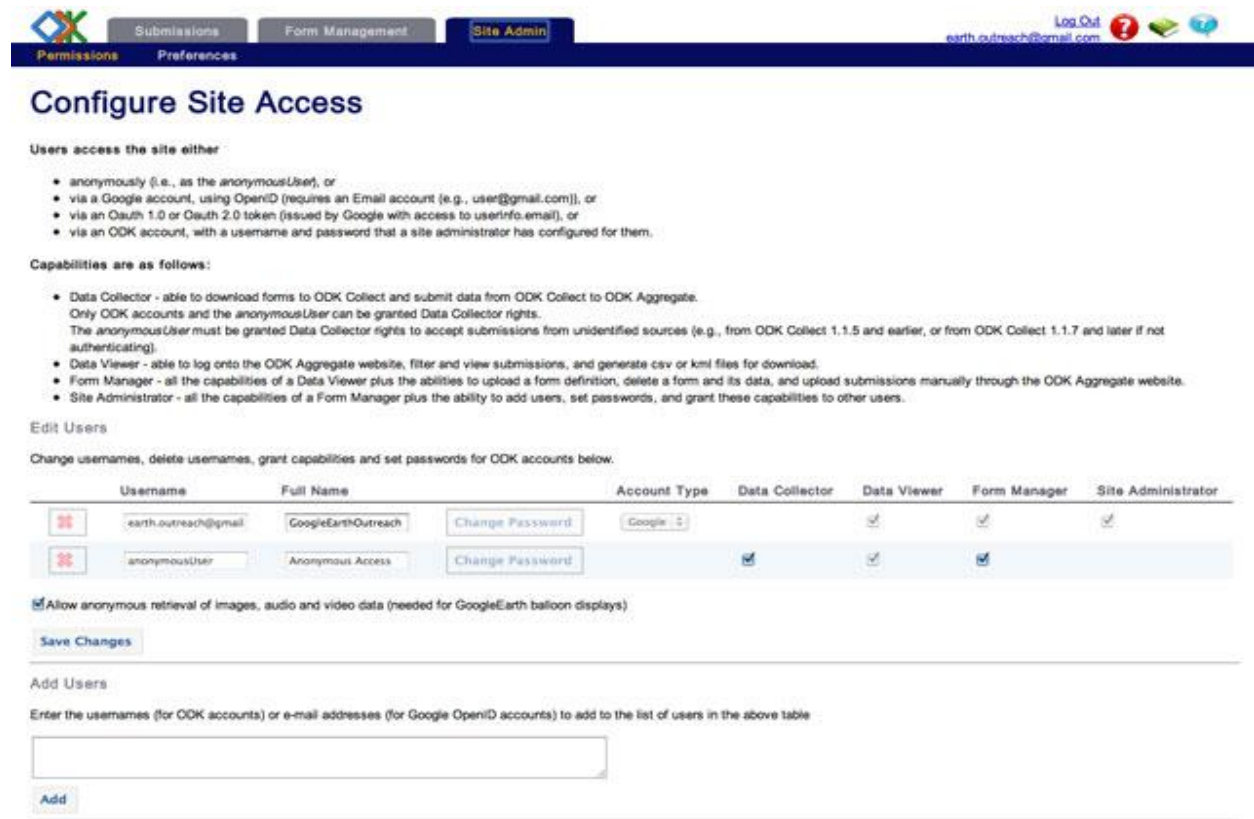


Figure 4 c shows the page to configure access to the ODK Aggregate.

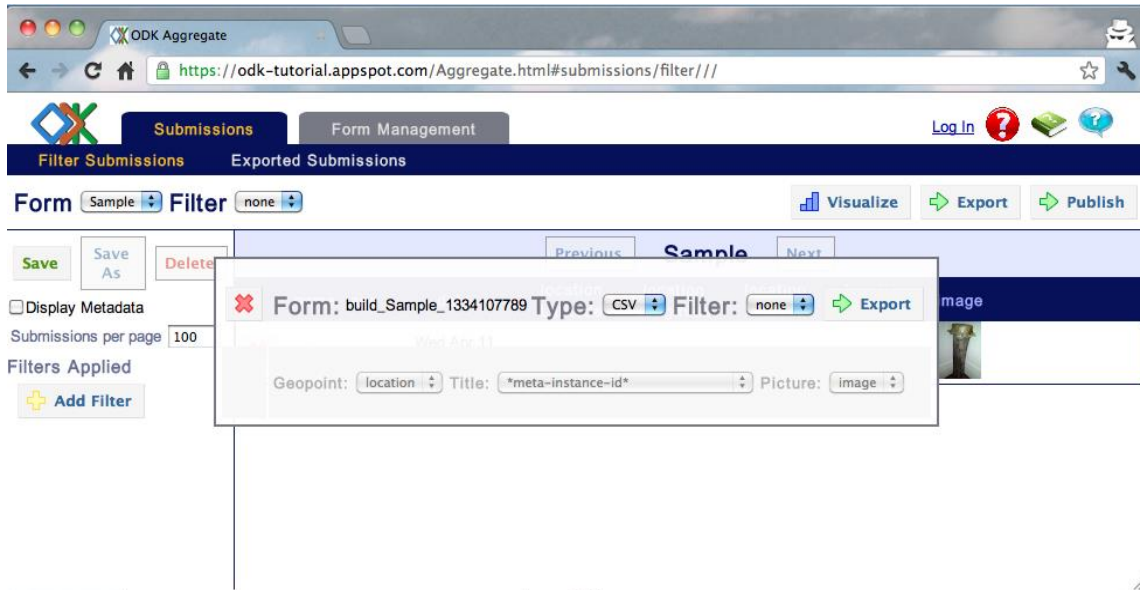


Figure 4 d shows the process of exporting form submission as a CSV file.

CONCLUSION

The concepts of cloud computing and open source are here to stay even beyond the 21st Century. The need to save on financial costs and to make efficient use of limited computing resources has driven cloud computing and open source software to the enormous technological phenomenon in this century. Their adoption by both governmental and non-governmental organizations would see a tremendous impact in donor funding which relies heavily on information gathered and the need for more funding evidenced by the positive feedback gained from the initial funding. The fact that there are tools that have simplified these concepts to the point where no programming expertise is needed has opened their power to all kinds of researchers other than those in the Information Technology field. There are also unlimited online support resources as well as forums and discussions boards on Open Data Kit tools.

RECOMMENDATIONS

Adoption of seamless Information Technology innovations for research and data collection has evidently transformed the way we make decisions that require processing the data into information within the shortest time possible to the point of making real-time decisions that have immediate positive impacts from the prompt feedback collected. This results in a continuous and sustainable development especially when this information that comes from the ground in real time is used in making appropriate budgetary allocation in red-flag areas. With the inception of devolution and with every county having an Information Technology docket, these dockets can be of critical importance in assisting the county leadership in the collection of feedback from citizens and in making significant improvement that would result in a more effective and efficient service delivery. Real-time data can also enable sectors such as security be able to gather evidence from crime scenes by capturing images and collecting statements from witnesses about the happenings before, during and after an accident or a crime. Respondents can also provide intelligence about an impending terror attack which would trickle to the relevant authority so as to make appropriate counteractions.

REFERENCES

- Cloud Computing. September 2014. [Online]. Available <http://investopedia.com/terms/c/cloud-computing.asp/>
- Hartung C et.al, Open Data Kit: Tools to Build Information Services for Developing Regions
Huduma Kenya” September 2014. [Online]. Available: <http://hudumakenya.go.ke/>
- Manage your data using ODK Aggregate September 2014. [Online]. Available:
https://www.google.com/earth/outreach/tutorials/odk_aggregate.html
- Matatu Goes Cashless Ahead of the July Deadline. September 2014. [Online]. Available:
<http://www.businessdailyafrica.com/Matatu-goes-cashless-ahead-of-July-deadline-/-/539546/2320584/-/dnnl5az/-/index.html>

Mobile data collection using ODK Collect. September 2014. [Online]. Available:
https://www.google.com/earth/outreach/tutorials/odk_collect.html
Nairobi City Goes Cashless on Parking Fee. September 2014. [Online]. Available:
<http://www.nairobi.go.ke/home/news/nairobi-city-county-goes-cashless-on-parking-fee-payment/>
Open Data Kit,” September 2014. [Online]. Available: <http://opendatakit.org/>
Twei K, Systems Design Document, Automation of Public Funds Transfer: *A Case of the Uwezo Fund Loan Application Process*. 2014
What is Open Source? September 2014. [Online]. Available: <http://opensource.com/resources/what-open-source/>