Dynamic Data Collection for Participatory Science in Open Data Kit

Yaw Anokwa
Computer Science and Engineering
University of Washington
Seattle, WA 98195
yanokwa@cs.washington.edu

Gaetano Borriello
Computer Science and Engineering
University of Washington
Seattle, WA 98195
gaetano@cs.washington.edu

Carl Hartung
Computer Science and Engineering
University of Washington
Seattle, WA 98195
chartung@cs.washington.edu

Waylon Brunette
Computer Science and Engineering
University of Washington
Seattle, WA 98195
wrb@cs.washington.edu

Jeff Beorse
Computer Science and Engineering
University of Washington
Seattle, WA 98195
jeff@beorse.net

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
Open Data Kit (ODK) is a modular, extensible, open-source set of tools originally designed to enable data collection and information delivery for developing regions. As ODK has matured, users have requested functionality that goes beyond the original vision of the platform. In this paper, we describe the user-requested features being added to enable more dynamic data collection for participatory science.

Keywords
data collection, citizen science, participatory sensing, mobile phones, ictd

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
In recent years, information management in developed regions has relied on technological solutions for increased efficiency, speed, and decision-making. Despite this, much of the world has been unable to benefit from these advancements due to constraints ranging from poor infrastructure to technically-savvy users. In these developing regions, paper (the default practice) limits the scope and complexity of the

information management services that can be provided. To help resource-constrained organizations leverage newer technologies, Open Data Kit [1] was created as free and open-source solution for building data collection and information delivery services.

System Description

Open Data Kit is a modular set of components that can be used individually or in various configurations to build a data collection or information delivery system. To create an end-to-end system, our users rely on three tools: ODK Build, ODK Collect and ODK Aggregate.

ODK Build is an HTML5 web application where users drag and drop "prompts" to create a "form" the user will interact with. The form is a W3C XForm that describes both the application logic and data schema. Prompts in the form support various data types (text, numbers, dates, etc.), branching logic, input constraints, and multiple languages.

For example, for a "birding" form, a prompt might ask the user to capture an image of a bird. The next prompt would ask for a location (e.g., obtained through GPS), and if an image and location were captured, a third prompt would ask for a description of the bird. In addition to capturing images, audio or video, the forms themselves can use multimedia in the prompts. For example, the birding form might include a series of images and sounds to aid users in identifying birds.

Once created, the form can be given to a device to display. ODK Collect is a mobile application for Android that displays the forms. Collect runs on phones, tablets and netbooks and is designed for asynchronous and offline use. Once forms are completed, users can send the resulting data over USB, Wi-Fi or the cell network to a server like ODK Aggregate.

Aggregate hosts the submitted data and provides extraction interfaces such as spreadsheets, maps, and queries. It can also stream data to external servers. Aggregate runs on local machines as well as in the cloud. To better support non-technical users, Aggregate installs in a few clicks and automatically creates the required databases from the XForm that Collect uses.

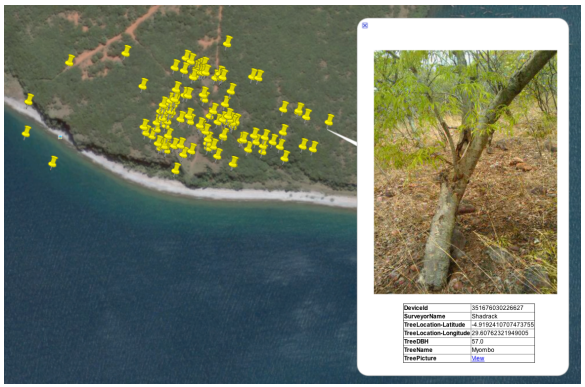

figure 1. Forestry workers with the Jane Goodall Institute in Tanzania submit data from ODK Collect to ODK Aggregate and then visualize the data in Google Earth. Managers can click on points to reveal forest survey data.

ODK tools are designed to maintain compatibility with other XForms systems. For example, users can use a variety of form designers (PurcForms, XLS2XForms, OpenRosa Designer), other mobile clients (JavaRosa, Episurveyor Mobile) and other backends (Kobo, DataHQ, OpenMRS). For systems without native XForms support, APIs for integration are provided.

figure 2. ODK Collect prompts can be multimedia rich (play a sound or video of an animal) and support logical branching (i.e., ask questions based on what was observed).

Current Uses

Uses of ODK tools vary widely. The Academic Model Providing Access to Healthcare (AMPATH) has used ODK Collect to gather information on over 63,000 patients in 18,000 households in Western Kenya. Collected data is submitted directly into their OpenMRS medical record system. They report higher user satisfaction, better data quality, faster turnaround of data and cost-savings over their previous PDA-based system.

The Berkeley Human Rights Center (HRC) used ODK Collect to gather 4000 offline surveys in Liberia for a report on the country's transition to peace. They note that ODK is easy for non-technical users to put into practice. Furthermore, the HRC has contributed to the ODK ecosystem by developing the Kobo PostProcessor for aggregating data on local machines for deployments with no wireless connectivity.

ODK tools are also being used in developed countries. Examples include PTP's use of ODK in surveying transit workers in North Dakota, LETS GO's use of ODK in fostering high school student learning in Sweden, and UW NatureMapping's use of ODK in monitoring biodiversity across the United States by enabling amateur nature biologists to properly identify animal species through direct sightings and other observations such as calls and tracks [6].

Requested Modifications

Participatory science enlists average citizens who may require support in making observations and identifications.

There are also data collection scenarios where data capture must be done quickly as the opportunity presents itself. For example, when trying to record a birdcall or take a photo of an animal, users do not want to be encumbered by the navigation requirements of a large form.

Also, to improve the quality of the data gathered, context-dependent prompts could help elicit more accurate information from the user. These are necessarily highly specific to the species, its habitat, and the type of observation being made (whether it is tracks, scat, or disturbance of the environment).

To better support these scenarios, several changes are being made to ODK. New prompt types are needed to

allow new functionality such as: adding just-in-time or historical data from external sensors (via USB, Bluetooth, etc.), reading and caching information from external (local and remote) data sources (i.e., location-specific selection lists, prompts based on previously submitted user submissions, etc.), and allowing data (i.e. multimedia, location traces) to be gathered in an ad-hoc manner at the time of observation and transferred into a form later. New UI components are also being added to better support the filtering and selection of ad-hoc and external data.

A remodeled XForms core is also needed to support forms with thousands of prompts and data elements (in a large decision tree). As forms grow larger, so too do the data structures that store the embedded logic. When designing ODK, such large forms were not anticipated and so a rethinking of how forms are modeled is needed. Instead of parsing the entire form into memory, lazy-loading of prompts from an on-device database and support for chaining together forms are being added. Form designers, like Build, must also be improved to make creation of large forms easier.

The dynamism of new prompt types and remodeled core will better support crowd-sourced data collection in ODK. For example, if a birding organization discovers a bird that was new to that locale from one of its submissions, forms on the mobile devices of other users in the same locale would automatically update with images and sounds of that particular bird.

Related Work

The ODK set of tools comes from a long line of generic data collection systems. The most notable of these in the academic space are CAM [2] and MyExperience [3]. In the commercial space, Pendragon Forms [4] and Episurveyor [5] are well-known and widely used.

Broadly speaking, ODK differs from previous work by focusing on modular, single purpose tools for smartphones and the cloud. We prioritize easy to use configurations that can build end-to-end or piecemeal systems. Additionally, our tools treat multimedia as first-class objects, support a variety of input modalities, comply with existing open standards, and integrate with other systems.

Conclusion

As ODK has matured, users have pushed the platform beyond its original vision. In this paper, we have described ODK and the modifications being made to the platform to enable more dynamic data collection.

References

[1] Hartung, C., Anokwa Y., Brunette, W., Lerer, A., Tseng, C., Borriello G. Open Data Kit: Tools to Build Information Services for Developing Regions. *In Proc. ICTD 2010*.

[2] Parikh, T.S. and Lazowska, E.D. Designing an Architecture for Delivering Mobile Information Services to the Rural Developing World. *In Proc. WWW 2006*.

[3] Froehlich, J., Chen M.Y., Consolvo S., Harrison B., Landay, J.A. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. *In Proc. of MobiSys 2007*.

[4] Pendragon. http://pendragonsoftware.com

[5] Episurveyor. http://www.datadyne.org

[6] NatureMapping. http://depts.washington.edu/natmap