Development Engineering 1 (2016) 36–44

Contents lists available at ScienceDirect

Development Engineering

journal homepage: www.elsevier.com/locate/deveng

A proposed integrated data collection, analysis and sharing platform for impact evaluation

Andreas Kipf[a], Waylon Brunette[b], Jordan Kellerstrass[a], Matthew Podolsky[a], Javier Rosa[a], Mitchell Sundt[b], Daniel Wilson[a], Gaetano Borriello[b], Eric Brewer[a], Evan Thomas[c,*]

[a] University of California, Berkeley, 450 Sutardja Dai Hall, MC: 1764, Berkeley, CA 94720-1764, United States
[b] University of Washington, 101 Paul G. Allen Center, Department of CS&E, 185 Stevens Way, Seattle, WA 98195-2350, United States
[c] Portland State University, Department of Mechanical and Materials Engineering, 1930 SW 4th Ave., Portland, OR 97201, United States

ARTICLE INFO

Article history:
Received 24 March 2015
Received in revised form
25 November 2015
Accepted 8 December 2015
Available online 10 December 2015

Keywords:
Cloud computing
ICT4D
Impact analysis
Provenance
Global development

ABSTRACT

Global poverty reduction efforts value monitoring and evaluation, but often struggle to translate lessons learned from one intervention into practical application in another intervention. Commonly, data is not easily or often shared between interventions and summary data collected as part of an impact evaluation is often not available until after the intervention is complete. Equally limiting, the workflows that lead to research results are rarely published in a reproducible, reusable, and easy-to-understand fashion for others. Information and communication technologies widely used in commercial and government programs are growing in relevance for international global development professionals and offer a potential towards better data and workflow sharing. However, the technical and custom nature of many data management systems limits their accessibility to non-ICT professionals. The authors propose an end-to-end data collection, management, and dissemination platform designed for use by global development program managers and researchers. The system leverages smartphones, cellular based sensors, and cloud storage and computing to lower the entry barrier to impact evaluation.

1. Introduction

Efforts to assess the impact of global poverty reduction projects, such as solar lighting installations, latrines, water pumps and filters, and cookstoves, often rely on data collected through person-to-person surveys, subjective observations, and/or expensive and time-consuming experimental studies. Data is frequently recorded by hand and processed on a per-project basis. These conventional approaches have limitations that can impact the value of the derived data. In the case of surveys and observations, research has shown surveys often overestimate adoption rates due to courtesy bias (where the participant is attempting to please the surveyor) (Manun'Ebo et al., 1997) or recall bias (tendency to forget details in more distant past) (Stanton et al., 1987). Furthermore, the presence or repeated visits of observers or enumerators can cause reactivity—influencing the behavior they are measuring (Zwane et al., 2011). And even with well-designed experimental studies such as randomized controlled trials, the data collected and subsequent impact analysis are often not available until well after the intervention is considered complete. This can delay providing input to subsequent interventions. Overarching these challenges is the bespoke nature of most data collection, analysis and sharing systems that are either (1) basic and limited or (2) expensive.

A user-configurable, online data management platform supported by robust electronic data collection tools may improve data quality and sharing and reuse as well as program accountability and performance in the global development sector. Further, such a platform could reduce the redundant implementation and management of many common components (e.g., database servers, transmission protocols, and privacy arrangements) and enable researchers to spend their time and money on the development project and impact analysis instead. In this paper, we propose a design for a platform like this called Mezuri (Esperanto for "measure").

Kepler (Altintas et al., 2004), Conveyor (Linke et al., 2011), Taverna (Hull et al., 2006), Mobyle (N'eron et al., 2009), DHIS2 (Manya et al., 2012) and Open Foris (Miceli et al., 2011) are

Abbreviations: ODK, Open data kit; M&E, Monitoring and evaluation; SUMs, Stove use monitors

* Corresponding author.
E-mail addresses: kipf@cs.berkeley.edu (A. Kipf), wrb@cse.uw.edu (W. Brunette), kellerstrass@cs.berkeley.edu (J. Kellerstrass), podolsky@cs.berkeley.edu (M. Podolsky), javirosa@cs.berkeley.edu (J. Rosa), msundt@cse.uw.edu (M. Sundt), dlwilson@berkeley.edu (D. Wilson), gaetano@cse.uw.edu (G. Borriello), brewer@cs.berkeley.edu (E. Brewer), evan.thomas@pdx.edu (E. Thomas).

http://dx.doi.org/10.1016/j.deveng.2015.12.002

examples of existing scientific workflow and data management platforms. These platforms have several limitations, such as their supported programming languages. For example, while Kepler allows users to integrate R and MATLAB code into workflows, code written in other languages can only be integrated in the form of web services, which might already be too difficult for most users. Additionally, some of these platforms are domain specific and thus only provide a limited set of algorithms. Mobyle, for instance, focuses on algorithms for the bioinformatics domain. DHIS2 is exclusively for health workflows and while it provides built-in analysis features and a web API, any customized processing code is run outside of the system. Further, these platforms only, if at all, track the provenance of data (i.e., how data was collected or processed) from the point where it enters to the point where it leaves the platform. As these platforms do not contain data gathering functionality they neither capture the provenance of the originally collected data nor the provenance of the final outputs which leave the system (e.g., visualizations).

In contrast to most of these platforms, Mezuri is conceived as a more broadly applicable, user-configurable data collection, analysis and sharing platform for global development professionals. Mezuri aims to provide end-to-end support for collecting provenance data and allows users to choose from a variety of programming languages and even combine different languages when implementing their workflows. Our proposed platform builds on existing efforts to collect data with smartphones and cellular-based sensors and digital surveys in global development settings, and combines these technologies with online data tools. Specifically, Mezuri will extend Open Data Kit (ODK) by building upon ODK 2.0's (Brunette et al., 2013) infrastructure to create an integrated data collection platform with provenance, processing, analysis, and sharing. Mezuri will use ODK 2.0 infrastructure and protocols to enable end-to-end integration with the ODK 2.0 tool suite. Additionally, Mezuri will leverage existing remote sensor data collection systems like Get All The Data (Pannuto et al., 2013) and SWEETSense (Thomas et al., 2013). These systems were selected because of the team's existing expertize with these platforms. In this paper, we identify the needs of researchers based on user surveys (see Section 2) and example applications (see Section 3), derive corresponding engineering requirements (see Section 4), and propose a design that addresses these requirements (see Section 5). Finally, we will outline future technology challenges (see Section 6).

2. Identification of needs

Mezuri aims to be a broadly accessible data collection and processing platform that helps global development experts build workflows that meet their field operational and research needs. To identify those needs, we conducted interviews with potential users. We interviewed 17 researchers engaged in global development impact analysis projects. Our interview consisted of 19 questions organized into the following categories: collection, processing, analysis, sharing, provenance, security and privacy. Each of those categories addressed different aspects of the potential users' workflows.

In most cases, interviewees reported that survey data collection is conducted by enumerators in the field using phones or tablets with tools like ODK Survey. Conducted surveys have up to several hundred questions and thousands of participants. In some cases, interviewees are collecting instrumentation data from energy, power, and temperature sensors. Some of the sensors store their data on local storage and wait for it to be manually collected by humans, whereas others have a cellular connection and are able to automatically send their data to the cloud. Some projects produce raw sensor data of many terabytes per month. Once collected, survey data is mostly cleaned using tools such as Open Refine. We found that the programming language R is the most common among our interviewed researchers, followed by Python and C/C++. Our interviewees stated that a common processing step is correlating survey and sensor data in time, especially when both data sets contain GPS coordinates.

Once the data has been processed and analyzed, researchers often share their data. We found that the methods of sharing vary strongly among our interviewees, and include emails, cloud storage services, scientific data platforms, and web sites or databases of research groups. Only about half of our interviewees stated that they share their code on platforms such as GitHub, using emails, or describe it in their publications. Others do not share it as they do not want it to be public or consider it as highly project-specific. Of the 17 interviewed researchers, 13 reported that they need to keep track of the provenance of their data. Further, we found that 13 of our interviewees are dealing with sensitive data, including, but not limited to, personal identifiable information. Security and privacy arrangements include access control checks using databases or protected files, encryption of data that is in transit, and de-identification of data prior to its publication.

3. Example applications

The following examples of monitoring and evaluation applications using ICTs helped inform Mezuri's requirements and design. These applications guided the development of the Mezuri prototype as archetypal use cases that combine sensors, smartphone surveys, and data storage, analysis, and sharing.

3.1. Cookstove use monitoring

One example of a typical monitoring and evaluation case study is our cookstove work in Darfur, Sudan. In this study we compared objective cookstove adoption measured by sensors versus user-reported adoption measured by surveys (Wilson et al., 2014).

The Berkeley–Darfur Stove is a high-efficiency wood-burning cookstove developed by Lawrence Berkeley National Laboratory and the University of California, Berkeley. As of December, 2014, more than 35,000 of these stoves had been distributed in and around internally displaced persons (IDP) camps in North and South Darfur. Beginning in July 2013, 180 participants were chosen by local camp leaders (Omdas) based on the criteria that participants had not previously been recipients of a Berkeley–Darfur Stove. Of the 180 participants, 170 received instrumented stoves that included sensors. The purpose of these sensors was to record a time-series of cookstoves' temperatures that could validate adoption of the cookstoves. Previous work has shown that temperature measurements can be used to detect and measure cookstove usage over time (Ruiz-Mercado et al., 2012). The sensors, termed Stove Use Monitors or SUMs, were built upon the Maxim DS1922E model iButton data logger.

Over the 3.5 months of the experiment, 180 participants were surveyed twice using ODK Collect for baseline and follow-up surveys. A third interaction took place in the form of a second follow-up when SUMs were removed from stoves. The five administrative units all received their cookstoves and baseline surveys over one week in July 2013. However, follow-up surveys were not conducted after the same interval for all groups. Instead, administrative units were followed up with one at a time at two-week intervals to spread out enumerator resources (the follow-up survey was much more onerous than the baseline survey) and to distribute sensors with different sampling rates across the population.