# Ohmage: A General and Extensible End-to-End Participatory Sensing Platform

H. TANGMUNARUNKIT, C. K. HSIEH, B. LONGSTAFF, and S. NOLEN, UCLA
J. JENKINS, C. KETCHAM, and J. SELSKY, Cornell Tech
F. ALQUADDOOMI, D. GEORGE, J. KANG, Z. KHALAPYAN, J. OOMS, and N. RAMANATHAN, UCLA
D. ESTRIN, Cornell Tech

Participatory sensing (PS) is a distributed data collection and analysis approach where individuals, acting alone or in groups, use their personal mobile devices to systematically explore interesting aspects of their lives and communities [Burke et al. 2006]. These mobile devices can be used to capture diverse spatiotemporal data through both intermittent self-report and continuous recording from on-board sensors and applications.

Ohmage (http://ohmage.org) is a modular and extensible open-source, mobile to Web PS platform that records, stores, analyzes, and visualizes data from both prompted self-report and continuous data streams. These data streams are authorable and can dynamically be deployed in diverse settings. Feedback from hundreds of behavioral and technology researchers, focus group participants, and end users has been integrated into ohmage through an iterative participatory design process. Ohmage has been used as an enabling platform in more than 20 independent projects in many disciplines. We summarize the PS requirements, challenges and key design objectives learned through our design process, and ohmage system architecture to achieve those objectives. The flexibility, modularity, and extensibility of ohmage in supporting diverse deployment settings are presented through three distinct case studies in education, health, and clinical research.

Categories and Subject Descriptors: H.4 [**Information Systems**]: Information Systems Applications; H.3.4 [**Information Systems**]: Information Storage and Retrieval—*Systems and software*; H.3.4 [**Information Systems**]: Information Storage and Retrieval—*Online information services*; J.4 [**Computer Applications**]: Social and Behavioral Sciences; J.3 [**Computer Applications**]: Life and Medical Sciences

General Terms: Design, Experimentation, Human Factors, Management, Measurement

Additional Key Words and Phrases: Participatory sensing, mobile data collection, citizen science, experience sampling, mobile sensing, data sharing, educational tool, clinical research

---

This work was made possible with funding from the UCLA Center for Embedded Networked Sensing (NSF CCF-0120778), the National Science Foundation Math Science Partnership (NSF DUE-0962919), the National Institutes of Health (NHLBI 5RC1HL099556), the Intel Science and Technology Center for Pervasive Computing (ISTC-PC), Cornell NYC Tech, and gifts from Google and Nokia.
Authors' addresses: H. Tangmunarunkit (corresponding author), C. K. Hsieh, B. Longstaff, S. Nolen, F. Alquaddoomi, D. George, J. Kang, Z. Khalapyan, J. Ooms, and N. Ramanathan, UCLA, 3551 Boelter Hall, Los Angeles, CA 90095; emails: {hongsudt, changun.tw, brent.longstaff, technolengy}@gmail.com, falquaddoomi@ucla.edu, donygeorge@gmail.com, jinhakang@ucla.edu, {zkhalapyan, jeroenooms, nithyaar}@gmail.com; J. Jenkins, C. Ketcham, J. Selsky, and D. Estrin, Cornell Tech, 111 8th Ave #302, New York, NY 10011; emails: {john.paul.jenkins, cketcham, joshua.selsky}@gmail.com, destrin@cs.cornell.edu.

DOI: http://dx.doi.org/10.1145/2717318

**ACM Reference Format:**
H. Tangmunarunkit, C. K. Hsieh, B. Longstaff, S. Nolen, J. Jenkins, C. Ketcham, J. Selsky, F. Alquaddoomi, D. George, J. Kang, Z. Khalapyan, J. Ooms, N. Ramanathan, and D. Estrin. 2015. Ohmage: A general and extensible end-to-end participatory sensing platform. ACM Trans. Intell. Syst. Technol. 6, 3, Article 38 (April 2015), 21 pages.
DOI: http://dx.doi.org/10.1145/2717318

## 1. INTRODUCTION

Advances in mobile technology and the ubiquity of mobile phones have enabled phones to become a convenient, affordable, and scalable real-time data collection platform. Participatory Sensing (PS) is a distributed data collection and analysis approach that takes advantage of smartphones [Burke et al. 2006]. In a PS project, participants use their phones to systematically collect data relevant to themselves or their community.

PS can facilitate and enhance many different applications. In health, it can assist individuals and their care providers in monitoring and managing symptoms, side effects, and treatments for chronic illness outside the clinical setting [Chen et al. 2012; Ramanathan et al. 2013; Krishna et al. 2009]. In emergency response scenarios, it can be used by citizens or emergency response teams to rapidly gather information in the field [Starbird and Palen 2011; Okolloh 2009]. In education, it can be used as an innovative instructional tool for teaching data collection and analysis concepts and methods [Mobilize 2010; Heggen et al. 2012]. All of these scenarios share in situ data collection, centralized data storage and management, as well as the need for deployment administration, data analysis, and visualization. However, construction of such a PS system is not the main objective of any of these projects, and it would draw time and resources away from their primary activities. Indeed, prior to developing ohmage, we developed multiple siloed PS systems, each specifically implemented for an individual study [Acker et al. 2010; Mun et al. 2009]. This approach discouraged code reuse, incurred high development and maintenance overhead, and did not scale well. These drawbacks motivated the development of a general, reusable, and extensible platform to help rapidly design and iterate on PS deployments.

Developing an effective general PS platform for collection, management, analysis, and visualization of diverse data streams needs to be driven by stakeholders in the system. Ohmage emerged from an iterative participatory design process that incorporates qualitative focus groups and interviews, as well as empirical iteration with hundreds of end users. During the course of its development and evolution, ohmage has been used as an enabling platform in numerous independent behavioral, health, wellness, and education projects addressing different populations (breast cancer survivors, new moms, HIV+ men, ADHD young adults, high school students, etc). Through our participatory design process, the following three groups of participants emerged over time:

(1) *project authors* who design data collection projects to answer a specific set of questions,
(2) *data collectors* who use their mobile devices to capture diverse spatiotemporal data through both intermittent self-report and continuous recording from on-board sensors and applications, and
(3) *software developers* who want to rapidly prototype data collection and analysis modules on the phone and on the server.

Unique challenges emerged within each group. These requirements drive the design of ohmage as an extensible end-to-end PS system with broader capabilities than just a data collection tool. Section 2 summarizes the requirements, challenges, and key design principles learned through our participatory design process.

There are several other PS platforms with varying capabilities and design objectives. Ushahidi [Okolloh 2009] supports mapping and analyzing user-reported data in various forms (e.g., photos, text). Sensr [Kim et al. 2013], Liquid [Liquid 2014], and EpiCollect [Aanensen et al. 2009] similarly support scripted surveys for smartphone apps. CitSci.org [Newman et al. 2011] supports scripted surveys for ecology with reusable data fields. Open Data Kit (ODK) [Brunette et al. 2013] allows semiprofessionals to make scripted surveys for Android devices and includes support for internal and external hardware sensors. Ohmage was specifically designed to allow easy configuration of systematic data collections in which individuals capture both self-report survey data and passive data in the course of their everyday lives using smartphones. Both types of data captures are scriptable and authorable. Ohmage self-report apps support a rich set of question types (e.g., choices, multimedia), as well as configurable temporal and special reminders on both Android and iOS devices. Although ohmage does not explicitly have a sensor abstraction framework like ODK, it can consume passively collected data streams through its application programming interface (API), and these streams are also scriptable like the surveys. Various mobile sensing apps have been integrated with ohmage to automatically collect contextual information such as the user's mobility status. In addition to data collection, ohmage also provides other end-to-end capabilities through data management tools to flexibly create and manage projects of different settings, as well as visualization tools to explore and make sense of collected data. Ohmage achieves its modularity and extensibility through a rich set of backward-compatible APIs that provide unified data access across all applications. We describe the system architecture in Section 3.

Among many ohmage deployments, we explore in more detail in Section 4 three unique case studies that demonstrate the breadth and flexibility of ohmage in supporting different research disciplines. These studies are as follows:

(1) *Mobilize*, in which the ohmage end-to-end PS process is used for teaching computational thinking in 81 math, science, and computer science high school classrooms to date;
(2) *Moms*, in which 44 young moms used ohmage for 6 months to study diet, stress, and exercise behaviors to better manage risk factors related to heart disease; and
(3) *PREEMPT*, in which ohmage was extended to set up and monitor personalized randomized control trials (N-of-1) of pain treatments in more than 100 participants with chronic pain.

For performance evaluation, Section 5 summarizes system efficiency based on our empirical observations and systematic load testing. Application usability from end users is also provided. Section 6 presents related work, and Section 7 concludes the article.

## 2. OHMAGE PARTICIPATORY DESIGN PROCESS

The design of ohmage was developed around two very different types of use cases:

(1) individualized studies, where data was collected and analyzed ultimately to better understand an individual's behaviors, and
(2) citizen science projects, where data was collected in a more coordinated fashion by a group of data collectors to better understand their shared experiences or context.

Ohmage development methodology is driven by an iterative participatory design process that includes qualitative focus groups and interviews, as well as empirical iteration, to solicit detailed needs and feedback from participants. Five focus groups were conducted with 20 participants living with HIV across two groups and 24 young mothers across three groups to discuss preferences, barriers, and attitudes regarding the use of smartphones for data collection and analysis [Ramanathan et al. 2013]. Thousands of authors, data collectors, and software developers in different research areas, including behavior, health, wellness, and education, have used ohmage to collect and analyze data in their studies over the course of years. Many think-aloud sessions were conducted to gain feedback on user experience and interface design. Formal structured and informal unstructured interviews and evaluations were conducted with hundreds of participants. Feedback from participant interactions and lessons learned from past deployments have been continuously incorporated into ohmage.

In this section, we summarize the received feedback, challenges, and key design objectives that emerged from three user groups—project authors or researchers, data collectors, and software developers.

### 2.1. Project Authors or Researchers

Project authors design data collection projects to answer a specific set of questions. Authors want authorability of diverse mobile data collection and analytical functions built in that enable interpretation of multiple streams of data at the same time. Authors also want to be able to set up, monitor, and manage data collection projects of tens or hundreds of data collectors. For example, some authors want to examine the incoming data and rectify any deployment problems that have arisen. Depending on their project needs, authors might want to grant different data access to different users (e.g., sharing data among participants in PS projects but not health studies). In some cases, the ability to customize the look and feel of the apps (e.g., different brandings, icons, or keywords) for their deployments were requested.

As for the data collection component, many authors desire personalized messages or interventions generated based on an individual's submitted data. For example, if there is no data submitted over the past 24 hours, a warning message should be generated to remind users to submit their data. Or if a participant reports a high level of stress, a personalized message that helps the user relax should be generated. For studies in which data collectors are non-English speakers, the foreign language support would be requested. In addition, there are needs for a feedback module on the mobile device. For example, in one study, the authors wanted young mothers to receive immediate feedback on their smartphone that explicitly compared their current diet, stress, and exercise behaviors to the baselines to help the moms improve eating habits, reduce stress, and increase their physical activity.

The key design objectives for this user group are as follows:

—*Scriptable data streams*: The system should support the integration, storage, management, and visualization of heterogeneous data collection requirements. These data streams should be scriptable and authorable. The system should also support self-report in foreign languages. Currently, ohmage allows different languages in the survey questions and entries. Full internationalization support remains our future work.

—*Study manageability*: The system should provide an interface for authors to manage their studies, monitor the incoming data, visualize study analytics, and manage fine-grained access controls based on user roles.

—*Configurable and customizable*: The system should be configurable to support diverse deployment scenarios. For example, different deployments need different settings (e.g., max data size, feature on/off, branding). Ohmage provides mechanisms to change app/system settings, as well as group management to support deployments isolation on the same backend.

—*Personalized interventions*: The system should support intervention mechanisms based on the collected data. Currently, ohmage supports visual feedback comparing users' current behaviors with their baselines, as well as messages that can be conditioned based on the current survey entry. We plan to support more sophisticated forms of interventions in the future.

### 2.2. Data Collectors

Data collectors use their mobile devices to capture diverse spatiotemporal data through both intermittent self-report and passively captured data streams. With mobile devices, data collectors expect to be able to collect data anywhere and anytime with or without network connectivity. Therefore, the applications need to operate smoothly in a variety of environments. Further, we found that asking participants to carry an extra device was burdensome and often led to data loss (e.g., participants may forget to carry or charge the extra devices). Therefore, data collection applications have to be available on a wide range of mobile platforms to enable more users to use their personal devices to collect data. Battery life was another key concern raised by many data collectors, especially when continuous data collection was enabled in the background (e.g., the GPS or accelerometer sensors for activity tracking).

For the app features, data collectors place a high premium on being able to customize their data collection and make the data collection more convenient. For example, across all five focus groups conducted with data collectors, customization of reminders was considered necessary. This suggestion agrees with many users' complaints on forgetting to collect data in deployments where reminders were not explicitly introduced. The five focus groups also suggested that there should be shortcuts, such as home screen buttons, to quickly record events. In addition, they identified the ability to set goals and track progress toward those goals to be extremely attractive for sustaining engagement. Young mother focus groups were enthusiastic about using phone camera to photograph their meals to help increase their self-accountability when trying to improve their diet. Many teenagers also found taking pictures to be fun and engaging.

An important challenge across all PS studies is user engagement. Ohmage currently does not implement sophisticated user engagement or gamification mechanisms, such as those in UbiFit [Consolvo et al. 2008], BeWell [Lane et al. 2011], Noom (us.noom.com), or Fitocracy (fitocracy.com). To proceed with our health studies, we adopted the monetary incentive, which is commonly used by most health researchers. For our education deployments, other techniques (e.g., classroom management) and incentives (e.g., extra credit, pizza parties) were used to encourage students' participation. However, as we move into a larger-scale operation, traditional incentives should be used alongside engagement mechanisms that will engage people more broadly and for longer time periods. Effective user engagement remains our future work. For user recruitment, ohmage assumes that it is done independently and hence provides no mechanism for online or crowdsourcing recruitment [Welbourne et al. 2014]. However, ohmage provides a self-registration mechanism for users to participate in public PS projects.

The key design objectives for this user group are as follows:

—*Highly available apps*: The mobile apps should allow users to collect data anywhere and anytime. The apps should be available on a wide range of mobile devices. The ohmage self-report apps currently support iOS and Android devices, which together account for 93.9% of the smartphone market share in Q3 2013 [comScore 2013]. The apps should be battery conscious and, in the case of passive data collection, allow users to adjust the collection frequency.

—*Mobile app usability*: In addition to ease of use, the mobile apps should provide mechanisms to support user participation. Examples of these mechanisms are reminders to prompt users to take surveys that fit their schedule, shortcut buttons for quick data entry, and simple feedback to encourage user participation.

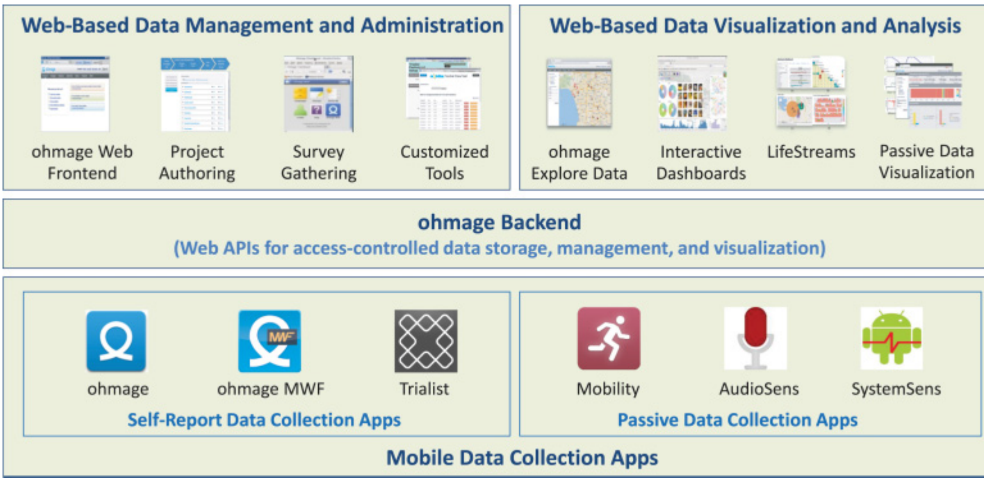



Fig. 1. Ohmage system architecture and its end-to-end PS software suite.

### 2.3. Software Developers

Developers require secure, flexible, and reliable APIs to support rapid prototyping and building of robust applications on both the phones and servers. In addition, as in many open source projects, many ohmage components are developed independently and in parallel by different groups of developers. Thus, a comprehensive set of stable and well-documented APIs that allow different components to be integrated in a loosely coupled fashion is critical to project success.

The key design objective for this group is as follows:

—*Modular and extensible*: The system should allow different modules to be changed independently without impacting the rest of the system. Ohmage provides a well-defined set of APIs that are backward compatible within major releases to securely and centrally create and manage ohmage objects. These APIs allow different applications (e.g., new mobile apps, custom visualization) to be independently developed and integrated.

## 3. OHMAGE SYSTEM ARCHITECTURE

The ohmage architecture (Figure 1) consists of four primary components: (1) the ohmage backend, which serves as a datastore and provides a unified interface for data access; (2) mobile data collection apps that run on participants' phones for data collection and feedback; (3) Web-based data management and administration tools for study management and administration; and (4) Web-based data analysis and visualization tools for exploring, analyzing, and visualizing the collected data.

### 3.1. Ohmage Backend

The ohmage backend is the central component of the ohmage platform. It provides secure communication, authentication, account management, access control, and data storage and management through an extensive set of APIs. These APIs support the create, read, update, and delete (CRUD) operations to ohmage objects such as data streams. To support consistent security across all applications, access to these CRUD operations by phone apps, administrative interfaces, and all other tools is orchestrated solely through the ohmage Web APIs [Ohmage 2010a]. As the backend evolves, these APIs remain backward compatible (within the same major release). The ohmage backend provides the foundation for achieving many key design objectives described in

Section 2, especially those for the project authors and software developers. We summarize its key features in this section.

*3.1.1. Scriptable Heterogeneous Data Streams.* One of the main design objectives of ohmage is to allow users to create and manage heterogeneous data streams that they want to collect. Two types of data streams are supported: (1) self-report data that are entered by the users and (2) passive data that are passively and continuously collected from mobile sensors or apps. These two different types of data are created and managed in a unified approach through scriptable data schemas. A data schema describes the content, structure, and metadata of the data to be collected. It serves as a contract between the data collectors and data consumers. Ohmage supports scriptable data schemas by allowing the schema definitions to be specified in XML. These schemas can be manually created or automatically generated through an authoring tool. Once the schemas are defined, the ohmage backend can immediately start to receive the data. Every data upload will be validated against its corresponding schemas. By default, ohmage includes a small set of standardized metadata, including timestamps and GPS. These metadata allow researchers to perform simple temporal and spatial analysis across deployments. However, for deeper cross-deployment analysis, the agree-upon data needed for the analysis can be added to the data definitions. We describe the detail of the two data stream types next.

*Self-report data.* Self-report data refers to data captured by a prompted experience sampling method where participants are asked to record their observations and experiences in the form of surveys [Froehlich et al. 2007]. A *project*, a self-report data schema, contains a set of surveys, each of which is a sequence of messages or prompts to be displayed. A message (e.g., “Please get up and walk for 2 minutes”) is a brief communication that does not require user input. A prompt solicits a response from the user, either an answer to a question (e.g., “Did you exercise today?”) or some other form of input (e.g., “Take a picture of your snack”). Ohmage supports a rich set of prompt types, including: (1) single/multiple choices with the option for users to define additional choices; (2) number; (3) free text; (4) timestamp; (5) multimedia, such as pictures, videos, or audios; and (6) a *remote activity* that launches a third-party application, such as a game app, to assess attention. Each prompt and message can optionally set a condition based on previous prompt responses that determines whether or not it should be displayed. For instance, a prompt “Did you take your medication before, after, or without food?” can be set to display if the user responded “Yes” to the previous prompt “Did you take medication X this morning?” This capability enables a more interactive and responsive survey answering experience.

*Passive data.* Passive data refers to data streams that are passively and continuously collected from mobile apps. Examples of these data streams are location traces and audio samples. Similar to self-report data, the ohmage backend allows mobile app developers to arbitrarily define a data schema, called an *observer*, to describe the passive data payloads for their apps. Currently, there are three mobile apps that submit data to ohmage: Mobility, AudioSens, and SystemSens (described in Section 3.2). We are also working on integrating third-party sensing apps, such as Funf (funf.org), to use the observer APIs.

*3.1.2. Account Management and Access Controls.* A data collection project usually involves many people with different roles and hence requires different data access restrictions. Ohmage provides account management and role-based access control to help manage different studies with different needs in diverse settings. Group management allows a collection of accounts, each with a different group role, to be managed together. This capability enables isolation of projects from one another, thus allowing multiple

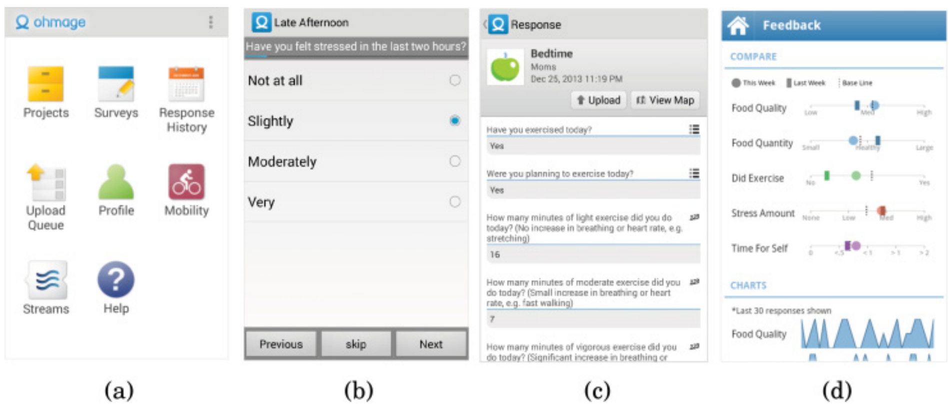



Fig. 2. Ohmage Android app. (a) The app home screen showing different menu options. (b) A prompt for user input. (c) A completed survey. (d) User feedback showing their current status versus their baseline in the Moms study (Section 4.2).

studies to be securely managed on the same backend. In addition to a group role, a user is also assigned system-wide roles (e.g., system admin) and project roles (i.e., supervisor, owner, analyst, participant). A combination of these roles is flexibly used to determine user access privileges in the system.

### 3.2. Mobile Data Collection Apps

The ohmage mobile data collection component consists of mobile apps that interact with the data collectors and collect various data streams. Based on the types of data streams, these apps can be classified into two groups: self-report apps and passive data collection apps. A collection of these apps and their collective features are designed to address the highly available apps, mobile app usability, and personalized interventions objectives. Additionally, when applicable, libraries or APIs are also provided to facilitate the integration of third-party apps with ohmage. We describe these apps next.

*3.2.1. Self-Report Data Collection Apps.* Self-report apps allow participants to record their observations or experiences by answering a set of survey questions. A self-report app downloads the project schema from the ohmage backend and renders the corresponding graphical interface based on the prompt order, type, and display condition. Survey responses are automatically timestamped, geocoded, and uploaded to the ohmage backend.

*Ohmage Android app.* The ohmage Android app supports self-report data collection, temporally and spatially triggered reminders, and data summary; it also implements a resource-aware upload manager that makes it simple to integrate other mobile sensing apps into ohmage. For self-report data collection, the app allows users to browse and participate in available projects in real time. It supports all message and prompt specifications described in Section 2, including video/audio capture and third-party app integration through the remote activity prompt type. For example, instead of describing one's emotion in text or numerical rating, a project can utilize the Photographic Affect Meter (PAM) app [Pollak et al. 2011] that allows users to choose a picture that reflects their current mood. Survey responses can be submitted even when a network connection is not available. These responses will be buffered in an upload queue and uploaded when the network is available. Figure 2 shows different screenshots of the ohmage app.

A few additional features are provided to facilitate survey taking. Shortcut buttons can be configured for users to quickly submit short prefilled surveys from the home screen. For example, a stress button can be used to quickly report a stressful event. The app also provides an interface to set reminders based on time and locations to take surveys. Users can view their previous survey responses based on time (i.e., calendar view) and location (i.e., map view), and visualize a data summary that compares their current performance with a configurable baseline (shown in Figure 2(d)). This comparison is one instance of data feedback intended to sustain user participation.

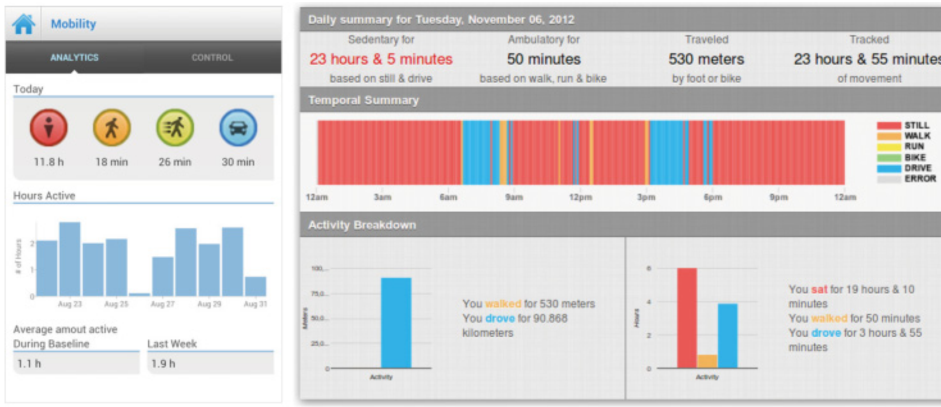
Finally, to minimize the resource consumption and streamline the integration with other mobile sensing apps, a unified ohmage data uploader was developed. The uploader supports authentication, secure data upload, and error handling. It can be configured to upload only via Wi-Fi and will not upload any data if the battery level is below 20%. These functions are made available through the Android interprocess communication interface, allowing other mobile sensing apps to upload the collected data to ohmage in simply one function call.

*Ohmage MWF.* One of the ohmage design objectives is to allow participants to use their own mobile devices to collect data by making ohmage available on multiple mobile platforms. To achieve this goal, we implemented ohmage MWF using the UCLA-developed Mobile Web Framework (MWF) (mwf.ucla.edu). Ohmage MWF is a lightweight cross-platform mobile Web app that provides a subset of functionality of the native Android version, namely self-report data collection and time-based reminders. It makes use of HTML5, Javascript, and MWF technologies for dynamic and responsive page rendering and data cache for offline operations. The PhoneGap library is used as middleware to access the native device features such as GPS, camera, and reminders across platforms. For manageability across apps, ohmage MWF adopts similar navigation flow and GUI design as the native Android app. Ohmage MWF is currently available on Android and iOS.

*3.2.2. Passive Data Collection Apps.* Passive data collection apps are mobile apps that collect data from sensors or other applications on mobile devices. Along with the raw data, apps may also provide high-level features (e.g., mobility mode or voice vs. nonvoice) extracted by applying specific algorithms to the raw data. The main design objective of these apps is to reliably collect data while consuming minimal resources, such as battery and network, to minimize the impact on the participants' phone availability. In this section, we briefly introduce three apps and their approaches for resource efficiency. Each of these apps collects a different type of data and has a dedicated Web-based tool (described in Section 3.4) to analyze and visualize the data.

*Mobility.* Mobility infers the user's mobility mode—still, walking, running, or driving—using accelerometer, Wi-Fi, and GPS data. It uses the variance and FFT features of acceleration to classify whether the user is walking or running and uses changes in ambient Wi-Fi signals or GPS to distinguish between driving and being still. The use of a power-consuming GPS sensor is limited to the situation where a WiFi signal is not available. To further reduce the energy consumption, Mobility duty cycles the accelerometer and GPS (if used). The user can choose to sample the mobility data every 1 or 5 minutes. Figure 3 shows the data summary on the phone and the Web-based visualization.

*AudioSens.* AudioSens samples and processes audio data obtained from the phone microphone. It uses a speech recognition algorithm to detect whether an audio sample contains speech and to investigate the user's social interactions. To reduce the energy consumption, the app dynamically adjusts the sampling rate. When the app has not detected speech for a certain amount of time (e.g., when the user is sleeping), it exponentially decreases the sampling rate.




(a) (b)

Fig. 3. Mobility. (a) The app analytics screen showing mobility and ambulatory summaries for 1 and 10 days. (b) The Web-based Mobility Dashboard showing a daily summary of the user's mobility status, mobility times series, and activity breakdown based on time and distance.

*SystemSens.* SystemSens captures and analyzes the phone usage data. It records more than 30 types of phone usage information covering (1) event-based records (e.g., phone calls and SMS) generated whenever a system state changes and (2) polling-based records (e.g., CPU, memory) generated periodically. These data are useful in analyzing phone usage patterns [Falaki et al. 2011] or inferring user behaviors such as sleeping patterns.

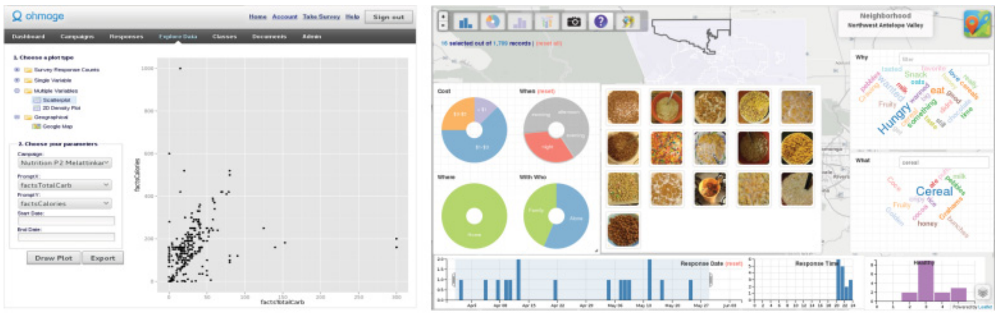
### 3.3. Web-Based Data Management and Administration

Managing a data collection project can be challenging and time consuming. Ohmage provides a set of tools to address the study manageability objective of PS deployments. In addition, a mobile alternative data collection tool is also available as an option for participants without mobile devices. These tools are described next.

*Ohmage Web Frontend.* Ohmage Web Frontend is the main project, data, and user management portal. A system administrator can use the Frontend to create and manage user accounts and groups, as well as audit the ohmage log to evaluate system performance and identify anomalies. A project author can use the Frontend to create and manage projects, monitor the incoming data, visualize study analytics, and control project accesses. Participants can use the Frontend to view or delete their data, change privacy states of their responses, and export their data. A self-registration interface is available for new users to create their own accounts. The Frontend also has a unified login system whereby other Web applications on the same server can simply redirect their logins to the Frontend login and allow it to handle all client authentication.

*Project authoring.* A project definition can be scripted in an XML file. However, for nontechnical authors such as researchers or teachers who are not familiar with programming, ohmage also provides a GUI-based Project Authoring tool. The tool guides users through a step-by-step project creation process and automatically creates the project XML content that can be uploaded to the backend. Users can click to view the temporary XML content at any time during the process. On-screen tool tips describe each configuration parameter, as well as potential values for that parameter.

*Survey gathering: A mobile alternative.* Even though PS focuses mainly on the use of mobile devices as data collection tools, in some deployments such as Mobilize



(a) (b)

Fig. 4. The ohmage Explore Data and public Snack Interactive Dashboard. (a) A scatter plot of the Snack Nutrition data collected by students in an algebra class showing Total Carbohydrates versus Total Calories. (b) A dashboard showing multiple single-variable distributions (e.g., Snack Cost, Healthy Level, images) of the Snack data filtered by "Cereal" snacks that users had at night.

(Section 4.1), not all participants have access to those devices. To allow such users to participate in the data collection, a generic Web-based application, extended based on the ohmage MWF codebase, was created. This tool allows users to complete their surveys using desktop browsers with two drawbacks: (1) there is no photo taking capability, so instead a user will be asked to upload an existing image file, and (2) the GPS obtained from the browser might be inaccurate or unavailable. Despite its limitations, this tool has proven to be very useful in the educational deployments because their analysis does not rely solely on these data types.

### 3.4. Web-Based Data Visualization and Analysis

The ultimate goal of many PS studies is to make sense from the data and make the results available and actionable to the end users. Instead of a one-size-fit-all solution, ohmage provides a set of tools that facilitate data exploration, visualization, and analysis for different usage scenarios. For self-report data, there are the generic ohmage Explore Data visualization and the Interactive Dashboard. For passive data streams, there is a specific visualization for each ohmage passive data collection app. Finally, Lifestreams takes both self-report and passive data into its data analysis stack and makes sense of both types of data simultaneously. All tools dynamically retrieve their data from the ohmage backend. This section describes each of these tools.

*Ohmage Explore Data.* Ohmage Explore Data allows project authors and end-users to explore and visualize high-level analytics of self-report data. There are three different views provided: (1) a project monitoring view that focuses on the data collection progress, as well as user participation (e.g. leader board); (2) a spatial map view displaying locations of survey responses associated with the selected users with zooming capability; and (3) a prompt detail view that provides single variable time-series and distribution plots of different prompts associated with one or all users, and two-variable scatter and density plots of all users. An example of a two-variable scatter plot is shown in Figure 4(a).

*Interactive dashboards.* The Interactive Dashboard was designed to support the exploration of large multivariate datasets through "Overview first, zoom and filter, then details-on-demand" design pattern [Shneiderman 1996]. A set of tailored graphs (i.e., pie charts, bar plots, maps, wordcloud, image pane) can be chosen to present different survey questions. Users can directly interact with individual graphs to filter and zoom into the data. For example, one can select a range within a bar plot or lookup a specific word in a wordcloud. Once a filter is applied to a graph, other graphs will immediately be updated accordingly. The dashboard can be customized for a specific project through a configuration file. Figure 4(b) shows a dashboard customized for the Mobilize Snack project (Section 4.1) that can be used to explore the healthy-level rating of different snacks during different times of the day. A public version of it is available at https://lausd.mobilizingcs.org/snackdemo.

*Passive data visualization.* Ohmage provides visualization dashboards for the passive data streams collected by Mobility, AudioSens, and SystemSens apps. The Mobility dashboard displays a daily summary of participants' mobility status (shown in Figure 3(b)), mobility maps, and historical analysis that looks at mobility breakdowns over time. The AudioSens dashboard shows individuals' speech patterns over time and space, as well as comparisons of speech patterns between multiple users. Finally, the SystemSens dashboard shows time series of different resource consumption, such as battery level, and network, as well as the user's interaction with the phone, such as screen on and off events. These dashboards are used to provide feedback to participants.

*Lifestreams.* Lifestreams [Hsieh et al. 2013] is a more advanced statistical analysis stack designed to systematically and simultaneously reason from ohmage's self-report and passive data streams, as well as answer questions regarding a person's behavior. It transforms the raw ohmage data into more meaningful information (e.g., extracting time at home or at work from a person's location traces), selects relevant features either manually or automatically through feature selection methods, and performs statistical inference to identify behavioral trends and patterns. It has been used to identify behavior changes and recognizes behavioral patterns in a long-term behavioral study (Section 4.2).

## 4. OHMAGE CASE STUDIES: SUPPORTED RESEARCH AND FINDINGS

Ohmage has played a significant role in numerous independent behavioral, health, wellness, and education research projects. Examples include high-risk behaviors in HIV+ men, young adults with ADHD, PTSD in veterans, smoking cessation in teenagers, and monitoring diabetes in a personalized randomized control trial (N-of-1).

In this section, we present three distinct ohmage use cases: (1) Mobilize, mobilizing for innovative computer science teaching and learning; (2) Moms, studying diet-, stress-, and exercise-related risk factors for cardiovascular disease (CVD) in young mothers; and (3) PREEMPT, N-of-1 trials using mHealth in chronic pain. These use cases demonstrate different research disciplines, captured data, targeting populations, and scale. Mobilize presents a broader use of end-to-end PS practice in a larger educational deployment setting. The Moms study presents a typical application of ohmage in behavioral and wellness studies that collect and analyze participants' Ecological Momentary Assessment (EMA) and passive contextual data. Finally, PREEMPT demonstrates a role of ohmage in personalized, mobile-based, N-of-1 randomized experiments.

### 4.1. Mobilize: Mobilizing for Innovative Computer Science Teaching and Learning

The Mobilize project (mobilizingcs.org) is a targeted National Science Foundation Math Science Partnership project funded for 2010 to 2015. Mobilize aims to engage STEM (science, technology, engineering, and mathematics) students in data collection and analysis activities that promote computational thinking and civic engagement via innovative math and science instructions.

In Mobilize, students develop their computational thinking skills through the following PS process: (1) creating their data collection projects by choosing the topic and designing the project details; (2) collecting data by making key decisions about what, where, and when to report observations; and (3) exploring/analyzing the collected data. Ohmage provides an integrated and iterative learning platform for students to practice the PS process. Curriculum units in four different subjects have been developed to incorporate this learning process: (1) an 8-week Exploring Computer Science (ECS) Data Analysis unit; (2) 6-week math lessons in Algebra 1; (3) a 3-week science unit in biology; and (4) a full-year Introduction to Data Science (IDS) to introduce students to data science and statistics. Since 2012, ohmage and the curricula have been deployed to 106 LAUSD classes and 3,278 users with 20,492 survey responses.

*Ohmage application.*For project creation, canonical projects (e.g., Snack, Nutrition) are precreated for many introductory lessons. For advanced lessons, students are expected to create their own projects by writing XML scripts or using the Project Authoring tool. For data collection, Mobilize relies primarily on the ohmage self-report data collection tools. During the 2012 deployment, when only the native Android app was available, 250 Android phones were loaned to students. Since 2013, ohmage had been extended to provide self-report apps on Android, iOS, and the Web. Thus, all students had been asked to use their own mobile devices and school computers for their data collection exercises. For data exploration, Mobilize relies on the ohmage Explore Data and the interactive dashboards. The dashboards are easy to navigate and explore, and can be used to quickly engage students in data inquiry. Alternatively, the ohmage Explore Data allows students to systematically explore different scatter plots (Figure 4(a)). In addition, an R-based analysis tool is used for in-depth data analysis lessons.

An important step behind a successful deployment is the accurate and timely setup of user accounts and groups (or classes). Typically, accounts are created by an administrator using the Web Frontend. However, in a large deployment, the centralized approach is inefficient. Thus, ohmage has been extended to provide a Class Management tool for teachers with proper privileges to dynamically create and manage their own classes. This tool can be generalized to support other deployments.

*Results and findings.* The Mobilize project evaluation of the past deployments [Ong et al. 2012, 2013] have been encouraging. For the 2011–2012 implementation, ECS teachers found the unit to be innovative and enjoyable, as well as technically and pedagogically challenging. Students also had a positive learning experience with the unit. For the 2012–2013 implementation, ECS teachers felt more confident in engaging students in deeper discussions related to data. Math and science students found the units engaging and developed a better appreciation of the role of data in their lives. An example of students' feedback is "I enjoyed the fact that we were able to personally collect various data about our own lives and recording them with phones, rather than pencil and paper." Finally, most students found ohmage applications easy to use (see Section 5.3).

The Mobilize project has successfully demonstrated a broader use of ohmage as an innovative teaching and learning software in STEM education. In addition to high school classes, ohmage was piloted in undergraduate political science classes at UCLA in 2013 and is in the process of offering it as a UCLA-wide instructional tool.

### 4.2. Moms: Studying Diet-, Stress-, and Exercise-Related Risk Factors for Cardiovascular Disease in Young Mothers

CVD is the leading cause of death among women. One of the major risk factors for CVD is weight. More than 60% of women in the United States are overweight [NIH 2013]. This situation is worsened when women become mothers. Such trends emphasize the need for prevention or reduction in CVD risk factors for young mothers. The Moms study aims to assess the validity and reliability of using mobile technologies to monitor CVD risk factors of diet, stress, and exercise in young mothers.

The study collected participants' diet, stress, mood, exercise, and activity traces throughout the day. Between January 2012 and March 2013, 56 young moms participated in and completed the pilot. Among them, the experimental group consisted of

44 moms who used our apps to collect data for 6 months—the first 3 months of data were used to establish an individual baseline, and during the subsequent 3 months, moms received feedback that compared their current reports with their established baselines.

*Ohmage application.* During the 6-month pilot, each mom was given an Android phone installed with the ohmage self-report and Mobility apps. Surveys were launched four times a day with reminders in the morning, mid-day, late afternoon, and nighttime to prompt participants to report their diet, stress, mood, and exercise routines. Shortcut buttons to capture meals and stress events in real time were available. After the baseline period, participants would get a positive reinforcement feedback on behaviors that showed improvement compared to the established baselines. Participants could also go to the feedback screen to compare their current performance with the past week's performance and their baselines (Figure 2(d)). In addition, moms were asked to run the Mobility app to collect their mobility states. The ohmage Frontend and Explore Data were used by the project authors to create and manage the project. Lifestreams was later used for a qualitative study involving 8 of 44 moms during in-person interviews.

*Results and findings.* The Moms study is a relatively ambitious data collection project in terms of survey frequency and duration. Among the 44 moms, 15,599 survey responses across the four surveys were collected. On average, moms answered 2 surveys per day during periods where they answered at least once. For mobility, 3,834 days of data were collected with a minimum of 5, a maximum of 202, and an average of 87 days. This information on participation can be useful for other studies aiming for similar scale.

Almost all moms reported that the study made them more aware of their eating behaviors and conscious of unhealthy eating habits. The surveys and buttons, as well as the positive feedback received at the end of surveys, helped increase the moms' awareness of their habits. In addition, the study coordinator found Lifestreams helpful in providing more insight into an individual's behavior data and in guiding the discussion with the participants. The behavior change detection algorithm [Hsieh et al. 2013] was deemed as one of the most useful analyses.

There were a few common issues that arose in asking moms to use the study phones. Some reported difficulty in remembering to carry or charge their phone. This was especially a problem in analyzing the moms' mobility traces. These lessons emphasize the needs for cross-platform mobile apps to support the use of participants' own phones.

The Moms study represents a general use case of ohmage application in health and wellness research. Similar studies that used ohmage include the Family Wellness study that used the Android app and AudioSense to evaluate the feasibility of using mobile technologies for familial behavior studies, as well as a study on health-related behaviors among gay men and their peers that used the ohmage MWF and SystemSens to monitor gay men's behaviors and their social interactions (e.g., phone and SMS traces) among supporting groups [Comulada 2014].

### 4.3. PREEMPT (Personalized Research for Monitoring Pain Treatment): N-of-1 Trials Using mHealth in Chronic Pain

Chronic musculoskeletal pain is an enormous problem affecting more than 100 million Americans [Boudreau et al. 2009]. Pain treatments are often prescribed in a trial and error fashion that takes time to identify a successful treatment. N-of-1 trials are randomized controlled crossover trials conducted in a single patient. PREEMPT seeks to take advantage of personal phones to allow patients with pain to participate in N-of-1 studies as part of their everyday life. Under PREEMPT, an application ("Trialist") will be created to allow patients and their healthcare providers to run personalized experiments comparing two pain treatments.

*Ohmage application.* Trialist is a software package consisting of three components: a backend, a Clinician-Facing component, and a Patient-Facing component. The ohmage platform is used as the foundation for Trialist. The ohmage backend provides secure data storage and access. For the Clinician-Facing component, there are two Web applications: Trialist Setup and Trialist Dashboard. Trialist Setup allows clinicians and patients to collaboratively design an N-of-1 pain treatment trial in a clinical setting. Clinicians can set the experiment configurations, such as treatments to be compared and medication schedules. At the N-of-1 trial's conclusion, Trialist Dashboard displays the results in actionable items (e.g., the probability of long-term success with treatment A vs. treatment B) so that the clinician and patient can make a joint and rational decision on the best treatment.

For the Patient-Facing component, the ohmage MWF app has been extended and customized for the Trialist mobile app to provide the following functionality: (1) allows patients to collect data via configured surveys; (2) provides reminders informing patients of when to begin or switch a treatment, as well as when to enter clinical survey data; and (3) provides user feedback (e.g., study progress, time remaining before completion).

*Deployment.* PREEMPT has been live since June 2014. About 250 patients will be recruited to participate in this study; half of the patients will use Trialist, and the other half will adopt the usual care. The study consists of N-of-1 trials conducted over 4 to 12 weeks and 3-, 6-, and 12-month follow-ups. Trialist demonstrates how mobile health and ohmage can be used to support mobile-based N-of-1 studies and will pave the way for broader use of such trials in clinical practice across other chronic health conditions. It will also promote more personalized patient-centered healthcare.

## 5. OHMAGE PERFORMANCE EVALUATION

To demonstrate ohmage performance and scalability, we present empirical data of a live ohmage server that has been in operation since 2012 and results of a systematic load testing of frequently used APIs. Finally, usability feedback is also reported.

### 5.1. Empirical System Performance

Mobilize is so far the largest ohmage deployment. In this setting, ohmage runs on an Intel Xeon E5-2960 2.9GHz server, with an assigned eight core and 16GB of memory. Nginx 1.4.7 is used as a reverse proxy serving static content and handling SSL communication for the Tomcat 7 server. MySQL 5.5 is used for database. From the ohmage audit log from January 1, 2014, to June 30, 2014, there were more than one million API calls submitted. Among these, the top three read and write APIs are survey_response/read (58.3%), image/read (13.2%), and project/read (3.9%), and survey/upload (12.3%), user/setup (1.5%), and survey_response/update (0.8%), respectively.

Figure 5(a) shows the empirical cumulative distribution function of the observed response time of the top three read and write API calls. The response time directly affects user experiences; a high response time may cause a request to be dropped. We found that 90% of these calls were returned within 0.2 seconds. User/setup is slower than the others due to its complex password hashing. Upon further investigation, we found that the system had a low number of concurrent requests (i.e., 84.5% of calls were serviced while there were two or fewer concurrent requests and the maximum concurrency was 40). The number of dropped requests due to timeout was less than 0.001%.

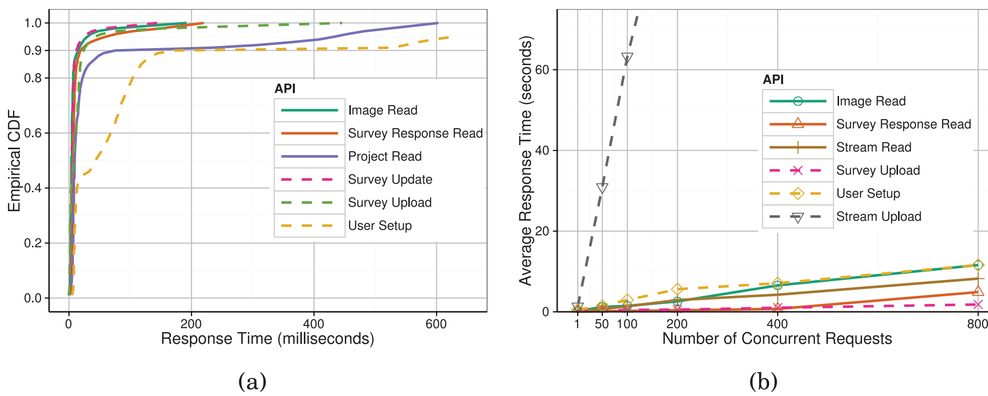



Fig. 5. Performance evaluation. (a) The empirical CDF of the response time in a 6-month 1,000-participant deployment; 90% of calls were returned within 0.2 seconds. (b) The average response time with increasing numbers of concurrent requests in load testing. Most of APIs can scale up to 800 concurrent requests.

### 5.2. Systematic Load Testing

We demonstrate how ohmage scales under different load conditions by performing systematic load testings on frequently used APIs. The experiment was done on a virtual server with the same setup as the Mobilize server. The database consists of 336,000 survey responses from 280 projects by 8,400 users in 280 different groups (a projected Mobilize load at the end of 2015). Each project has 1,200 survey responses. For passive data stream, mobility data from 150 users over a 4-month period with a 1-minute sampling interval were created. Our initial database contains 145GB of data; only about 1.4GB is used for survey data, whereas the rest is for passive data.

The Apache JMeter stress testing tool was used for our load testing. JMeter allows different numbers of concurrent requests to be sent to the load server simultaneously. The JMeter client was co-located on the same network as the load server. The database cache was clear before each test to minimize a caching effect between tests. Each test was performed 10 times. The concurrency tests were conducted to the top three read and write API calls with typical request parameters observed empirically, as well as stream/read and stream/upload for passive data. Survey/upload submitted one survey. Survey_response/read requests an individual's data. Stream/write and stream/read uploaded and requested 1 hour of mobility data, respectively.

Figure 5(b) shows the average response time observed by JMeter for different numbers of concurrent requests and APIs. For clarity, we removed the third read and write ranked APIs from the plot. The performance of these calls linearly depend on the number of concurrent requests. At 400 concurrency, survey-related operations were able to returned within 1 second. For most APIs, ohmage was able to support 800 simultaneous requests, which is much bigger than the observed load to date, with 0% to 5% of dropped calls. The exception is stream/upload, which could only scale to 400 concurrent requests due to the bottleneck on the payload validation (which we plan to improve in future releases). Nevertheless, the ohmage Android app has a randomized upload mechanism to alleviate the upload synchronization on the server.

Many read APIs allows different subsets of data to be returned depending on the requesting parameters. As the returned data get bigger, the bottleneck shifts from the database lookup to the serialization. For example, the most verbose form of survey_response/read returns 2.8MB for 30 users and 300KB for a single user. Although the file size of 2.8MB is 9.3 times bigger than 300KB, with 100 concurrent requests, its average response time is 27 times longer than the 300KB. Similar observation is true for the stream/read. For large datasets, it is more optimized to make multiple requests for smaller chunks than to request all of them at once.

### 5.3. Application Usability

In the Mobilize deployments, in addition to direct feedback from teachers and curriculum designers, a usability questionnaire was given to student data collectors to measure the ease of use of three different ohmage applications: the mobile data collection app, the Web-based Survey Gathering tool, and the Frontend. In this questionnaire, rating of 1 (very hard) to 5 (very easy) is used to represent students' overall experiences. The questionnaire was completed by 650 high school students, with 11th grade as a median grade, across the 3 years of deployments. The median across all three components is 4, whereas the mean is 4.11, 3.56, and 3.75, respectively. In general, most students found ohmage easy to use. Students felt more comfortable with the mobile apps than the mobile alternative data collection tool. This could be because the Web data collection tool used the same process flow and interface designed for a mobile context, which could be unfamiliar to students who did not have experience with mobile devices.

For project authors and software developers, we were able to obtain their feedback directly through interviews and discussions. In particular, the project authors found two features very helpful: the campaign authoring tool that allows them to design a project without involving a technical person, and the campaign monitor page, which reports the deployment progress. Software developers appreciated the flexibility of the ohmage APIs when extending system functionality, as well as the effort saved by the Android stream upload library when integrating their app with ohmage.

## 6. RELATED WORK

Many PS efforts have successfully demonstrated the feasibility of applying the PS paradigm to scenarios where traditional approaches fall short because of time, cost, or complexity. These efforts can be divided into two categories based on user participation. The first category focuses on citizen science projects that engage the public or community members in the data collection activity. For example, EarPhone [Rana et al. 2010] is a PS noise pollution mapping system for urban areas that uses phones' microphones to collect audio samples. The bus time prediction app of Zhou et al. [2012] uses sequences of detected cell towers, reported by the phones of bus riders, to detect bus routes and locations. GreenGPS [Ganti et al. 2010] uses vehicles' on-board diagnostics and GPS to learn the fuel efficiency of different routes through a city. Project Noah [Noah 2013] allows people to define missions consisting of organism types and geographic context, then app users participate by photographing the organisms.

The second category focuses on individualized data collected by users to address their personal issues related to health and wellness. For example, PEIR [Mun et al. 2009] combined personal location traces and geographic information to infer the users' impact and exposure to air pollution. Mappiness [MacKerron and Mourato 2013] studied how local environment affects happiness by using an iPhone app to collect mood, GPS, and ambient noise from more than 45,000 people in the United Kingdom. Many research and commercial systems help users track their personal fitness by logging nutrition and exercise information [Consolvo et al. 2014]. These solutions use automatic or manual data entry. For example, UbiFit [Consolvo et al. 2008] tracks physical activity automatically and also allows users to manually record additional activities not detected by the sensor. Nutrition tracking has mostly relied on manual logs, but services like Fitbit and MyFitnessPal provide databases of nutritional information for common foods. Kitamura et al. [2009] used computer vision to classify food by nutritional group (e.g., vegetable, dairy) and serving size, although the accuracy was low.

Whereas some PS systems are specifically designed for particular deployment scenarios, others, like ohmage, are more general and can be applied in broader settings, which allows the tools to be reused and shared across projects. Many general PS platforms with different capabilities and goals have emerged over time. Ushahidi [Okolloh 2009] was originally designed for Kenyans to report violence incidents, primarily via SMS, on feature phones. Since then, it has become a more general platform for gathering, analyzing, and visualizing user-reported data (text, photos, or other multimedia) from various data streams like SMS, email, Twitter, and Web forms. Ohmage also has a rich set of data management, analysis, and visualization tools for researchers and users to collect, manage, and explore their data. However, ohmage only supports data collected through mobile and Web applications.

Sensr [Kim et al. 2013] is a tool to create, share, and manage citizen science projects. Sensr provides a phone UI widget-based survey authoring tool with limited supported data types (i.e., photo, text, and radio buttons). The Sensr mobile app only supports iOS devices. Liquid [2014] was designed to allow collaboration on measuring rainwater quality but is now a general PS platform providing survey authoring, Android-based data collection, and Web-based charting tools. Liquid supports simple data types (i.e., text, Boolean, date, time, number, choice) and a mechanism to invite project collaborators. EpiCollect [EpiCollect 2014; Aanensen et al. 2009] was designed to facilitate the data collection of field workers in studies with geographic contexts, such as epidemiological and ecological studies, and is a general PS platform providing a Web-based survey form builder, Android and iOS data collection apps, and Web-based data visualization. Similar to ohmage, EpiCollect supports a rich set of data types, including multimedia and supports conditional branching, although some features like branching and audio/video prompts are only available on Android. CitSci.org [Newman et al. 2011] is an ecological data management infrastructure providing Web-based tools for creating scripted citizen science projects, managing participants, choosing the data to study (species and environments), analyzing the data, and gathering feedback from users. It enables data sharing between projects with standard data formats with vetted lists of organisms and their attributes, as well as site characteristics.

Compared to the preceding self-report systems, ohmage supports a broader set of data allowing both scriptable survey and passive data to be collected by participants through their everyday lives. It has authoring tools for nontechnical users, and scripting extends to passive data streams as well as surveys. Instead of domain-specific data types provided by CitSci.org, ohmage data types are more generic. Unlike the previously mentioned apps, ohmage self-report schemas support not only questions but also messages that require no user input. In addition, ohmage apps provide participation-assisting features such as personal reminders and data feedback. Like EpiCollect and CitSci.org, Ohmage data collection tools are available for iOS and Android devices. As with Liquid and CitSci.org, users can also enter data through a traditional Web application.

Finally, ODK [Brunette et al. 2013] was designed to help semiprofessional users (such as fieldworkers) in developing countries collect survey data using Android devices. ODK has been used in public health and environmental monitoring, as well as for reporting human rights abuse, in many countries. It began with tools for scripting and collecting mobile surveys (text, multimedia, etc.) and organizing the data for analysis. ODK later added support for internal and external sensors with a sensor driver framework [Brunette et al. 2012]. Rather than providing a sensor driver framework, ohmage provides a general framework (i.e., observer) and the Android stream upload library to facilitate the collection and integration of passive data streams. These streams can be used to capture sensor data and higher-level information.

## 7. CONCLUSION

A general, reusable PS platform can benefit many applications in multiple areas. It helps save time and resources and allows projects to concentrate on their primary objectives instead of technology. In this article, we have presented the PS requirements, challenges, and primary system design criteria, as well as the architecture of ohmage, an open source platform that provides end-to-end PS functionality. We have demonstrated the breadth and flexibility of ohmage in supporting different research disciplines through three distinct use cases in education and health. Finally, we have summarized ohmage system performance and its usability.

Ohmage can be set up in multiple ways to support a PS deployment. A project can host its own ohmage instance by installing ohmage software components on a JVM-supported operating system. Simple installation packages are available for Ubuntu and Fedora [Ohmage 2010b]. Due to its flexible account management and access control mechanisms, ohmage can also be offered as a software as a service (SaaS) to projects that do not have resources or want to deal with system administration. This service model is adopted by the Mobilize project to provide PS services for individual classes in LAUSD. In addition, there is a public sandbox of ohmage (play.ohmage.org) that allows individuals to experiment with their personal data collection. In this sandbox, users can create their own accounts, create new projects or participate in existing ones, and collect and visualize their own data.

Ohmage will continue to evolve to incorporate new technologies and functionality through its iterative design process. While a native iOS self-report app is being developed, the ohmage MWF will be extended to provide more features, such as video and audio prompts and data visualization. In addition, we are investigating a scriptable personalized reminder mechanism triggered by the user's previous submissions and other contextual information. Using the Open mHealth architecture (openmhealth.org), we will integrate with other third-party sensing apps, such as Funf [Aharony et al. 2011]. This integration will help broaden the sensing capability available for diverse studies. For data analysis and visualization, we are developing a Web application that allows users to dynamically explore multidimensional relationships in survey data through the customizable composition of statistical plots. All of these extensions will strengthen and broaden ohmage capability, which in turn will benefit a larger user community.

## ACKNOWLEDGMENTS

The authors would like to thank J. Burke and M. Hansen for their contributions to the concept of PS and its use in Mobilize; M. Monibi, H. Falaki, K. Parameswaran, and T. Pham for their earlier contributions to the codebase; and R. Rocchio, E. Sakabu, and C. Trance for their support of the development of ohmage MWF.

## REFERENCES

David M. Aanensen, Derek M. Huntley, Edward J. Feil, Fada'a al-Own, and Brian G. Spratt. 2009. EpiCollect: Linking smartphones to Web applications for epidemiology, ecology and community data collection. *PloS One* 4, 9, e6968.

Amelia Acker, Martin Lukac, and Deborah Estrin. 2010. *Participatory Sensing for Community Data Campaigns: A Case Study*. Technical Report. Center for Embedded Network Sensing.

Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. 2011. Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing* 7, 6, 643–659.

Denise Boudreau, Michael Von Korff, Carolyn M. Rutter, Kathleen Saunders, G. Thomas Ray, Mark D. Sullivan, Cynthia I. Campbell, Joseph O. Merrill, Michael J. Silverberg, Caleb Banta-Green, and Constance Weisner. 2009. Trends in long-term opioid therapy for chronic non-cancer pain. *Pharmacoepidemiology and Drug Safety* 18, 12, 1166–1175.

Waylon Brunette, Rita Sodt, Rohit Chaudhri, Mayank Goel, Michael Falcone, Jaylen Van Orden, and Gaetano Borriello. 2012. Open data kit sensors: A sensor integration framework for android at the application-level. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and services (MobiSys'12)*. ACM, New York, NY, 351–364.

Waylon Brunette, Mitchell Sundt, Nicola Dell, Rohit Chaudhri, Nathan Breit, and Gaetano Borriello. 2013. Open data kit 2.0: Expanding and refining information services for developing regions. In *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications (HotMobile'13)*. ACM, New York, NY. Article No. 10.

Jeffrey A. Burke, Deborah Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B. Srivastava. 2006. Participatory sensing. In *Proceedings of the World Sensor Workshop (SenSys'06)*.

Connie Chen, David Haddad, Joshua Selsky, Julia E. Hoffman, Richard L. Kravitz, Deborah E. Estrin, and Ida Sim. 2012. Making sense of mobile health data: An open architecture to improve individual-and population-level health. *Journal of Medical Internet Research* 14, 4, 10.

comScore. 2013. U.S. Smartphone Subscriber Market Share. Available at http://www.comscore.com.

W. Scott Comulada. 2014. Mobile phone-based assessment of health-related behaviors among gay men and their peers. *Connections: International Network for Social Network Analysis* 34, 1–2, 43–51.

Sunny Consolvo, Predrag Klasnja, David W. McDonald, and James A. Landay. 2014. Designing for healthy lifestyles: Design considerations for mobile technologies to encourage consumer health and wellness. *Foundations and Trends in Human-Computer Interaction* 6, 167–315.

Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. 2008. Activity sensing in the wild: A field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*. ACM, New York, NY, 1797–1806.

EpiCollect. 2014. EpiCollect. Available at http://www.epicollect.net/.

Hossein Falaki, Ratul Mahajan, and Deborah Estrin. 2011. SystemSens: A tool for monitoring usage in smartphone research deployments. In *Proceedings of the 6th International Workshop on Mobility in the Evolving Internat Architecture (MobiArch'11)*. ACM, New York, NY, 25–30.

Jon Froehlich, Mike Y. Chen, Sunny Consolvo, Beverly Harrison, and James A. Landay. 2007. MyExperience: A system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th International Conference on Mobile Systems, Applications, and Services (MobiSys'07)*. ACM, New York, NY, 57–70.

Raghu K. Ganti, Nam Pham, Hossein Ahmadi, Saurabh Nangia, and Tarek F. Abdelzaher. 2010. GreenGPS: A participatory sensing fuel-efficient maps application. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (MobiSys'10*). ACM, New York, NY, 151–164.

Scott Heggen, Osarieme Omokaro, and Jamie Payton. 2012. MAD science: Increasing engagement in STEM education through participatory sensing. In *Proceedings of the 6th International Conference on Mobile Ubiquitous Computing, Systems, Services, and Technologies (Ubicomm'12*).

Cheng-Kang Hsieh, Hongsuda Tangmunarunkit, Faisal Alquaddoomi, John Jenkins, Jinha Kang, Cameron Ketcham, Brent Longstaff, Joshua Selsky, Betta Dawson, Dallas Swendeman, Deborah Estrin, and Nithya Ramanathan. 2013. Lifestreams: A modular sense-making toolset for identifying important patterns from everyday life. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys'13)*. ACM, New York, NY, Article No. 5.

Sunyoung Kim, Jennifer Mankoff, and Eric Paulos. 2013. Sensr: Evaluating a flexible framework for authoring mobile data-collection tools for citizen science. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'13)*. ACM, New York, NY, 1453–1462.

Keigo Kitamura, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2009. FoodLog: Capture, analysis and retrieval of personal food images via Web. In *Proceedings of the ACM Multimedia Workshop on Multimedia for Cooking and Eating Activities (CEA'09)*. ACM, New York, NY, 23–30.

Santosh Krishna, Suzanne Austin Boren, and E. Andrew Balas. 2009. Healthcare via cell phones: A systematic review. *Telemedicine Journal and e-Health* 15, 3, 231–240.

Nicholas D. Lane, Mashfiqui Mohammod, Mu Lin, Xiaochao Yang, Hong Lu, Shahid Ali, Afsaneh Doryab, Ethan Berke, Tanzeem Choudhury, and Andrew Campbell. 2011. BeWell: A smartphone application to monitor, model and promote wellbeing. In *Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth'11)*. 23–26.

Liquid. 2014. Liquid. Available at https://getliquid.io/.

George MacKerron and Susana Mourato. 2013. Happiness is greater in natural environments. *Global Environmental Change* 23, 5, 992–1000.

Mobilize. 2010. Mobilizing for Innovative CS Teaching and Learning. Available at http://www.mobilizingcs.org/.

Min Mun, Sasank Reddy, Katie Shilton, Nathan Yau, Jeff Burke, Deborah Estrin, Mark Hansen, Eric Howard, Ruth West, and Peter Boda. 2009. PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services (MobiSys'09)*. ACM. New York, NY, 55–68.

Greg Newman, Jim Graham, Alycia Crall, and Melinda Laituri. 2011. The art and science of multi-scale citizen science support. *Ecological Informatics* 6, 3, 217–227.

NIH. 2013. Overweight and Obesity Statistics. Available at win.niddk.nih.gov/statistics/.

Project Noah. 2013. Project Noah. Retrieved March 21, 2015, from http://www.projectnoah.org/.

Ohmage. 2010a. Ohmage API Documentation. Available at https://github.com/ohmage/server/wiki/.

Ohmage. 2010b. Ohmage Installation Manual. Available at http://ohmage.org/usage.html#developer_panel.

Ory Okolloh. 2009. Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action* 59, 66–70.

Christine Ong, Noelle Griffin, K. Binning, Girlie Delacruz, C. Byrne, K. Show, and E. Redman. 2012. *Mobilize Project Evaluation*. Fall 2012 Report. National Center for Research (CRESST).

Christine Ong, Noelle Griffin, Mark Hansen, A. Madni, K. Binning, C. Byrne, E. Redman, and N. Shin. 2013. *Mobilize Project Evaluation*. Fall 2013 Report. National Center for Research (CRESST).

John P. Pollak, Phil Adams, and Geri Gay. 2011. PAM: A photographic affect meter for frequent, in situ measurement of affect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, New York, NY, 725–734.

Nithya Ramanathan, Dallas Swendeman, W. Scott Comulada, Deborah Estrin, and Mary Jane Rotherman-Borus. 2013. Identifying preferences for mobile health applications for self-monitoring and self-management: Focus group findings from HIV-positive persons and young mothers. *International Journal of Medical Informatics* 82, 4, e38–e46.

Rajib Kumar Rana, Chun Tung Chou, Salil S. Kanhere, Nirupama Bulusu, and Wen Hu. 2010. Ear-phone: An end-to-end participatory urban noise mapping system. In *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN'10)*. ACM, New York, NY, 105–116.

Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*. IEEE, Los Alamitos, CA, 336–343.

Kate Starbird and Leysia Palen. 2011. "Voluntweeters": Self-organizing by digital volunteers in times of crisis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*. ACM, New York, NY, 1071–1080.

Evan Welbourne, Pang Wu, Xuan Bao, and Emmanuel Munguia-Tapia. 2014. Crowdsourced mobile data collection: Lessons learned from a new study methodology. In *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications (HotMobile'14)*. ACM, New York, NY, Article No. 2.

Pengfei Zhou, Yuanqing Zheng, and Mo Li. 2012. How long to wait? Predicting bus arrival time with mobile phone based participatory sensing. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys'12)*. ACM, New York, NY, 379–392.

Received December 2013; revised November 2014; accepted January 2015