# Development of Mobile Health Capabilities for Remote Data Collection in Resource-Limited Settings

*Adjoa M. Poku and Aaron T. Katz*

*Since 2007, experts with the Johns Hopkins University Applied Physics Laboratory (APL) and the SAGES (Suite for Automated Global Electronic bioSurveillance) program have been developing mobile technologies focused on improving electronic disease surveillance. Key findings from a 2007 assessment of a short message service (SMS)–based data collection system in the Philippines helped shape development efforts for future SAGES mobile systems. Subsequent industry and trade studies helped the team to identify suitable communications channels and essential requirements for a data collection system targeting resource-limited settings. The resulting open-source framework provides form-based data entry on Android devices, encrypted SMS data submissions, and automated data processing at a remote reporting site. The SAGES development team at APL continues to improve components of mobile systems in response to changes in the mobile technology ecosystem and the needs of the SAGES implementing partners.*

## INTRODUCTION

Disease surveillance systems are important tools for characterizing patterns of disease occurrence and detecting anomalous disease activity, aiding in disease prevention and control efforts.[1] The effectiveness of a disease surveillance system is limited by many factors, including the timeliness of reporting and the quality of the data provided.[1] Technological advances in the area of electronic disease surveillance have been shown to have a positive impact on overall reporting timeliness and data quality relative to paper-based reporting,[2] and it is in this context that recent advances in mobile technologies have the potential to advance the field further.

Since 2007, experts with the Johns Hopkins University Applied Physics Laboratory (APL) and the SAGES (Suite for Automated Global Electronic bioSurveillance) program have been engaged in the development of mobile technologies focused on improving electronic disease surveillance. This article provides a brief history of these efforts and an overview of the current state of the mobile technology efforts for SAGES.

## SAGES OVERVIEW

SAGES was developed by APL in collaboration with the Global Emerging Infections Surveillance and Response System, a division of the U.S. Armed Forces Health Surveillance Center (AFHSC-GEIS), to provide an end-to-end electronic disease surveillance capability in resource-limited settings. SAGES tools can be configured to customize a surveillance system that is designed around the needs and constraints of a particular installation and partnering organization. SAGES is composed of three primary categories of tools: data collection; data warehousing; and data analysis and visualization (Fig. 1). The overall objective of the mobile components of SAGES is to provide options for data collection in surveillance systems that are well matched to available communication and computing infrastructure at the data collection sites.

## THE mHEALTH LANDSCAPE

The growing field of mobile health systems, termed mHealth, is recognized as a distinct subfield of electronic health, or eHealth. No universally accepted definition of mHealth exists, but for the purposes of this article we will use the definition provided by the World Health Organization Global Observatory for eHealth: mHealth is "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices."[3]

The communication infrastructure needed to support mHealth systems is growing significantly. The International Telecommunications Union estimated that by the end of 2013, the global mobile-cellular penetration rate, based on individuals,[3] was 96%, with an 89% penetration rate in the subset of developing countries.[4,5] By the end of 2012 the global mobile broadband penetration rate stood at 22.1%, with a prediction of 30% by the end of 2013.[5]

The adoption of capable mobile platforms is growing in concert with this increasing mobile penetration. In 2013, sales of smartphones exceeded sales of feature phones, accounting for 53.6% of sales globally.[6] In August 2013 the bestselling smartphone internationally, the Samsung Galaxy S4 International Version, sported a GPS sensor, a 1.9-GHz quad-core processor, and 2 GB of RAM.[7,8] Although it is not possible to make a direct comparison, these specifications are roughly comparable to those provided by the 2007 2.0-GHz MacBook.[9]

Given the convergence of increasingly capable mobile platforms for health applications,[10] and increasing communication channel availability, the stage has been set for increased mHealth system realization. To this end, mHealth systems have seen a significant adoption globally, with the World Health Organization reporting that as of 2011, 87% of high-income countries and 77% of low-income countries responding to their global mHealth survey are participating in at least one mHealth initiative.[3]

It is in this context of growing capacity for mHealth systems and the promise demonstrated by eHealth to support disease surveillance[2] that the SAGES program began developing mHealth solutions focused on disease surveillance.

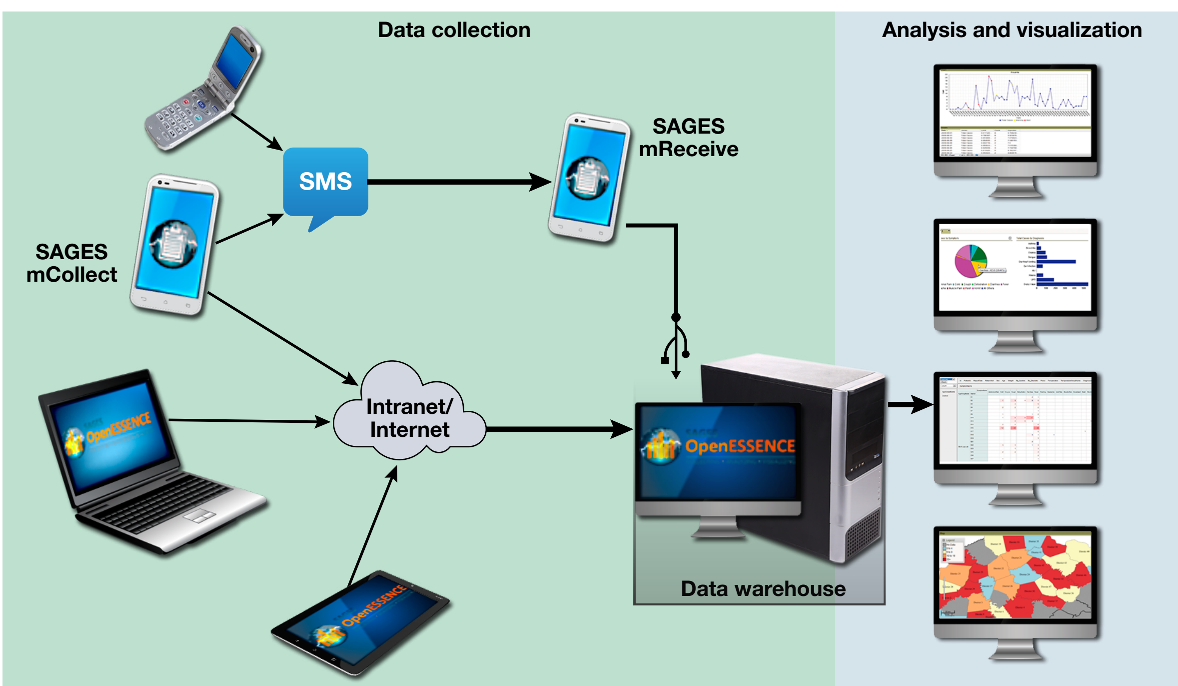


**Figure 1.** SAGES high-level architecture.

## SAGES mHEALTH TIME LINE

The SAGES program's involvement in mHealth dates back to 2007 with a collaborative effort between APL and Philippines-AFRIMS Virology Research Unit (PAVRU). PAVRU and the Republic of the Philippines Health Office, Cebu City (CHO) staff jointly developed a short message service (SMS)–based data reporting system to augment the existing paper-based reporting capabilities at clinics in Cebu City. APL was asked to perform a technical and public health assessment of the PAVRU system. Several key findings emerged from this effort that would help shape future SAGES mHealth development efforts, most notably:

- The sustainability of a mHealth system is constrained by the ability to develop and maintain the system using locally sourced materials and networks.
- The utility of a mHealth system is directly proportional to the validity of the data produced, and as such, robust data validation capabilities should be integrated and supported at as many levels as possible.
- The viability of a system is based not only on technical factors but also on the degree to which public health professionals, decision makers, funding organizations, and health workers embrace the concept.

The joint PAVRU–CHO system ultimately demonstrated the utility of integrating a close to real-time SMS channel of individual clinic data into a disease surveillance system to realize improvements in the timeliness of data reporting relative to paper-based systems.[11]

In 2009, APL undertook a study of global electronic information communication technologies, characterizing these technologies with respect to accessibility, cost, hardware requirements, and global availability. The results revealed a viable solution for mHealth in resource-limited settings.[12] The finding from this study, which is still supported by recent data,[5] is that although broadband communication methods are desirable for their speed and bandwidth, they do not have sufficient penetration in the developing world to serve as an underpinning for a mHealth system. In contrast, mobile-cellular communications have broad penetration (approximately 90%) in the developing world;[5] thus, leveraging the SMS protocol for data gathering in a mHealth system in the developing world appears to be the best path forward.

Mobile hardware evolves in tandem with the continual improvements and expansions of communications networks and protocols. Additionally, the availability of devices in different global markets often varies in terms of features and costs. In 2007, when SAGES initially entered the mHealth space, feature phones dominated the global markets. These handsets supported the Java 2 Micro Edition framework and offered a limited set of interactivity and a relatively small screen. By 2009, the technological landscape for mHealth system hardware had fundamentally shifted. With the release of the first generation of Apple iPhones in 2007 and the further proliferation of small-form-factor devices supporting Java 2 Micro Edition, there was a clear shift in the industry toward more interactive capabilities directly on the devices. Unfortunately, although these systems were highly capable, the proprietary nature of their application programming interfaces as well as the relatively high cost of the devices limited the utility of these systems in resource-limited settings.

With Google's open-source release of the Android Software Development Kit and the reference HTC Corporation's G1 hardware in 2008, the burgeoning smartphone space began to see the introduction of lower-cost but highly capable platforms with a myriad of sensors and touch screen interfaces. It is in this context that SAGES forged ahead to develop a more fully featured mHealth data collection capability that leveraged Android-equipped devices. Leveraging previous results from the APL–PAVRU effort, the system was architected with an aim to provide data validation at the point of collection, to provide data transmission via SMS, and to target lower-cost hardware sourced in the markets of foreign partners who were working with AFHSC-GEIS to develop an electronic disease surveillance capacity.

With this overarching mission in place, SAGES developers conducted a trade study to compare the suitability of developing a framework internally or leveraging existing frameworks. It was quickly determined that numerous suitable customizable frameworks existed in the marketplace, and leveraging an existing framework would be the most economical and expedient path forward.

Eight data entry frameworks were evaluated for the SAGES mobile platform (Table 1). The data entry frameworks were evaluated against seven separate criteria:

- Support for form-based data entry via a graphical user interface
- Support for transmission of data via SMS
- Support for Android-based devices
- Support for creation of forms with multiple data entry field types
- Open-source licensing for the framework
- Evidence that the community around the project is actively providing support and that the project is accepted as being beyond its beta release (i.e., a mature product)
- Free of licensing, maintenance, and service costs

**Table 1. Selection criteria for SAGES's client-side tool**

| Client-side tool | Form-based graphical user interface | SMS submission layer | Android support | Multiple field types | Open source | Mature product | Limited potential for operational costs |
|---|---|---|---|---|---|---|---|
| ODK Collect | X | | X | X | X | X | X |
| FrontlineSMS | X | X | | X | | X | X |
| JavaROSA | X | X | | X | X | X | X |
| EpiSurveyor | X | X | | X | X | | |
| GATHERdata | X | X | | X | X | | X |
| Mesh4X* | | | | | | | |
| eMOCHA | | | X | | | | X |
| Moca | | | X | | | | X |

Blue shading denotes selected candidate tool; taupe shading denotes candidate tool.
*Could not gain access to adequate information to validate any claims.

Based on these criteria, five data entry frameworks were selected as candidates: Open Data Kit (ODK) Collect, FrontlineSMS, JavaROSA, EpiSurveyor, and GATHERdata. After creating prototypes on these five frameworks, it was determined that the JavaROSA framework was too memory intensive for lower-end feature phones and that FrontlineSMS's proprietary code base would prohibit distribution under an open-source model. EpiSurveyor was eliminated because of the likelihood of third-party operational costs. GATHERdata was eliminated because of difficulty customizing the framework, time delays accessing the source code, and a lack of activity in the development community. The primary shortcoming of ODK Collect, namely its lack of built-in support for SMS transmission, could easily be overcome because of its open-source code base and developer-supportive application programming interfaces. Through this analysis, ODK Collect was selected as the basis for the SAGES data collection capability.

Nine data receiver frameworks were evaluated against seven criteria (Table 2):

- Support for receipt of data via SMS
- Local storage of transmitted data
- Compatible with the Microsoft Windows operating system
- Open-source licensing for the framework
- Setup process supported by an installer
- Evidence that the community around the project is actively providing support and that the project is accepted as being beyond its beta release (i.e., a mature product)
- Free of licensing, maintenance, or service costs

**Table 2. Selection criteria for SAGES's server-side tool**

| Server-side tool | SMS receipt layer | Data stored locally | Windows compatible | Open source | Installer-based setup | Mature product | Limited potential for operational costs |
|---|---|---|---|---|---|---|---|
| ODK Aggregate | | | n/a | X | X | X | |
| FrontlineSMS | X | X | X | | X | X | X |
| EpiSurveyor | X | | n/a | | X | X | |
| GATHERdata | X | X | X | X | | | X |
| RapidSMS | X | X | | X | | X | X |
| RapidAndroid | X | X | X | X | X | X | X |
| Mesh4X* | | | | | | | |
| eMOCHA | | X | n/a | X | | | X |
| Moca | | X | X | | | | X |

Blue shading denotes selected candidate tool; taupe shading denotes candidate tool.
*Could not gain access to adequate information to validate any claims.

Based on these criteria, three receiver frameworks were selected as candidates: RapidSMS, RapidAndroid, and FrontlineSMS. All three platforms were highly capable, but RapidSMS's lack of support for the Windows operating system and FrontlineSMS's proprietary licensing eliminated these frameworks. RapidAndroid met all seven criteria and was selected as the development base for the data receipt capability for SAGES.

## SAGES mHEALTH ARCHITECTURE

The SAGES mHealth architecture has been designed around a model of supporting full end-to-end cellular communications for data collection. It comprises three major components: a mobile data collection platform; a mobile data receiving platform; and an extract, transform, and load (ETL) process (Fig. 2).

The data collection platform, SAGES mCollect, is an Android application that has been developed on top of the open-source ODK Collect framework. It provides customized form-based data collection and thus can support the collection of both individual patient data and aggregate condition counts (e.g., the number of influenza cases seen at a clinic during a given week). The ODK brand offers tools that can be used to develop custom forms tailored to the needs of different reporting sites, and these forms can be shared among all of the users of SAGES mCollect as part of a SAGES installation. SAGES mCollect includes several mechanisms for data validation and quality assurance, which satisfies the requirement of performing data validation at the point of data entry.

Because the ODK Collect framework does not inherently support all of the features necessary for a secure cellular (minimum 2G)-only data infrastructure, several additional capabilities were developed in realizing SAGES mCollect. Extensive work was done to create a multipart SMS protocol, which allows for sending data in quantities that would otherwise be too large to package into a single SMS message (typically limited to 140 characters). Because of privacy concerns related to open transmittal of sensitive health data, a capability for sending and receiving encrypted and multipart SMS messages was provided. The system currently supports Advanced Encryption Standard 128-bit symmetric key encryption (the 128-bit key size meets the National Institute of Standards and Technology's minimum approved value for providing security strength using Advanced Encryption Standard cryptographic keys).[13] All of these additions are transparent to the end user of the SAGES mCollect app.

The data receiver application, SAGES mReceive, has been designed to support installations where a minimum 2G cellular data connection is available. For this use case, it has been implemented as an app on top of the open-source RapidAndroid framework. mReceive runs on an Android phone that is tethered via universal serial bus to the OpenESSENCE workstation. (OpenESSENCE is the flagship SAGES web system that provides data warehousing, analysis, and visualization capabilities.) This framework was extended in a fashion similar to that used for the mCollect application, adding in the multipart SMS and data encryption capabilities. In addition, an automation suite was built into the app, enabling mReceive to repackage and export data received in a form suitable for consumption by the ETL process.

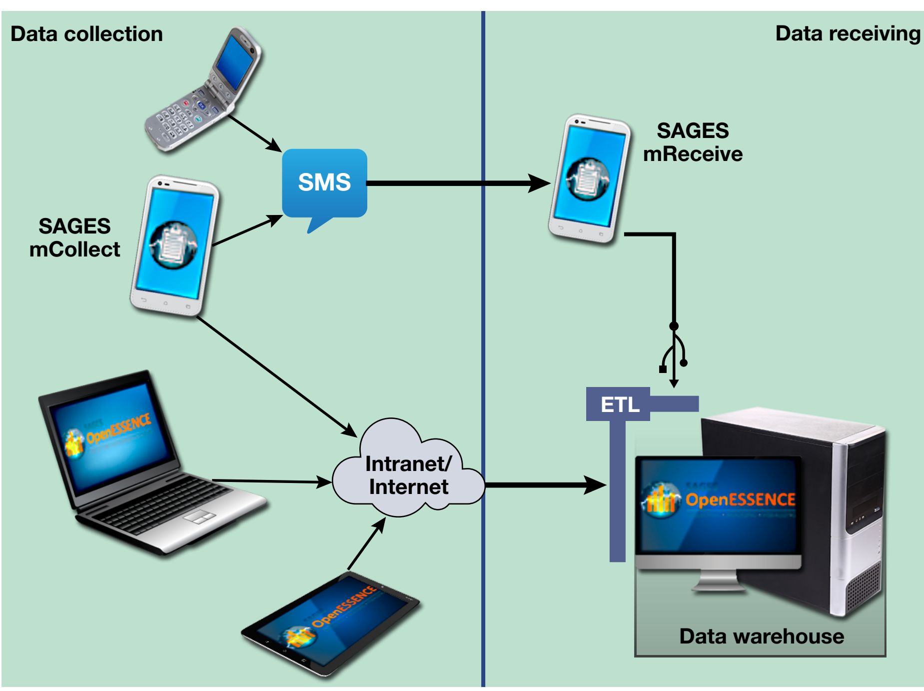


**Figure 2.** SAGES mobile components.

The ETL process has been implemented as a set of lightweight scripts, a Java application, and database-stored procedures that execute on a scheduled basis on the OpenESSENCE workstation. The mReceive application writes its data to a directory on the receiver phone, which is monitored by the ETL process. If data are available in this directory, the ETL process pulls the data from the phone, transforms them to the match the OpenESSENCE database schema, and pushes them into the OpenESSENCE database. A set of stored procedures in the OpenESSENCE database is triggered upon this load, serving as a final validation layer before the data enter the database.

## CONCLUSION

The SAGES mHealth architecture has been an integral component in most of the SAGES deployments to date. By supporting structured electronic data collection in remote point-of-care settings, it has the potential to improve data timeliness and validity, enabling more effective disease surveillance. It has been architected as a fully open-source system, on top of existing open-source frameworks, decreasing its cost burden and increasing its potential for modification and reuse. At the time of publication of this article, APL staff has partnered with public health professionals in sub-Saharan Africa to jointly tailor the architecture to support their unique data collection needs, with ownership and maintenance migrating fully to the partners in country.

The mHealth footprint for SAGES is expected to increase and become further refined in the future. SAGES mHealth components will continue to be suitable for support in resource-limited settings, but the leaders of the project plan to take advantage of recent advancements in mobile technologies to more broadly support better-resourced environments.

### REFERENCES


[1]Thacker, S. B., Choi, K., and Brachman, P., "The Surveillance of Infectious Diseases," *J. Am. Med. Assoc.* **249**(9), 1181–1185 (1983).

[2]Ward, M., Brandsema, P., van Straten, E., and Bosman, A., "Electronic Reporting Improves Timeliness and Completeness of Infectious Disease Notification, The Netherlands, 2003," *Eurosurveillance* **10**(1–3), 27–30 (2005).

[3]World Health Organization, *mHealth: New Horizons for Health through Mobile Technologies: Based on the Findings of the Second Global Survey on eHealth*, Global Observatory for eHealth Series, Vol. 3, World Health Organization, Geneva (2011).

[4]International Telecommunications Union Telecommunication Development Bureau, *ITU BDT Regions, Corresponding to the Regional Grouping of the ITU's Telecommunication Development Bureau*, http://www.itu.int/ITU-D/ict/definitions/regions/ (accessed 20 Nov 2013).

[5]The Broadband Commission for Digital Development, *The World in 2013: ICT Facts and Figures*, International Telecommunications Union, Geneva (2013).

[6]Gartner, "Gartner Says Annual Smartphone Sales Surpassed Sales of Feature Phones for the First Time in 2013," press release, http://www.gartner.com/newsroom/id/2665715 (13 Feb 2014).

[7]Whitney, L., "Galaxy S4 Slumps But Still Top-Selling Phone, Research Firm Says," *CNET*, http://news.cnet.com/8301-1035_3-57607739-94/galaxy-s4-slumps-but-still-top-selling-phone-research-firm-says/ (16 Oct 2013).

[8]Samsung Electronics, "Samsung GALAXY S4 Specifications," http://www.samsung.com/global/microsite/galaxys4/ (accessed 5 Mar 2014).

[9]Apple Computer Corporation, "MacBook (Mid 2007) – Technical Specifications," http://support.apple.com/kb/sp18 (accessed 5 Mar 2014).

[10]PricewaterhouseCoopers, *Emerging mHealth: Paths for Growth*, http://www.pwc.com/en_GX/gx/healthcare/mhealth/assets/pwc-emerging-mhealth-full.pdf (2012).

[11]Velasco, J. M. S., Coberly, J. S., Wojcik, R. A., Gibbons, R. V., Alera, M. T. P., et al., "Applications of the ESSENCE Desktop Software in the Analysis of Philippine National Dengue Data," *Am. J. Trop. Med. Hyg.* **85**(6 Suppl), 95–96 (2011).

[12]Ashar, R., Lewis, S., Blazes, D. L., and Chretien, J. P., "Applying Information and Communications Technologies to Collect health Data," *J. Biomed. Inform.* **43**(2), 332–341 (2010).

[13]Barker, E., and Roginsky, A., *Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths*, NIST Special Publication 800-131A, National Institute of Standards and Technology (2011).


## The Authors

**Adjoa M. Poku** is a software engineer in the Bio-Threat Awareness Systems Group in the Asymmetric Operations Sector (AOS). She is currently working on disease surveillance and modeling systems, including SAGES. **Aaron T. Katz** is the Assistant Group Supervisor and a project manager in the Bio-Threat Awareness Systems Group in AOS. For more information on the work reported here, contact Adjoa Poku. Her e-mail address is adjoa.poku@jhuapl.edu.

The *Johns Hopkins APL Technical Digest* can be accessed electronically at **www.jhuapl.edu/techdigest**.