Enhancing SAGES to support multiple data sources

Submitted by Shraddha Patel on January 18, 2018 - 2:05pm
Last revised by Web Producer on June 21, 2018 - 3:09pm.
Proposal Status: Review Complete

Executive Summary:

Digital Square is a partnership of digital health experts and organizations working together to create digital health systems in developing countries. The aim of Digital Square is to help strengthen national digital health systems by coordinating investments into scalable technology solutions and creating an environment in which they can be sustained for maximum impact. This proposal is in response to a funding opportunity put forth by Digital Square to invest in digital health global goods (software that is usually free and open source). The Johns Hopkins Applied Physics Laboratory (JHU/APL) has several years of experience in developing and deploying digital health open source software. JHU/APL has developed the Suite for Automated Global Electronic bioSurveillance (SAGES), which is a set of open source digital tools designed to meet the challenges of electronic disease surveillance in resource-limited settings. It is freely available to download, and the source code is available on GitHub. The suite of tools enables users to perform data collection, as well as spatial and temporal analysis and visualization of health-related data. Furthermore, the tools include detection algorithms to alert an end user to an anomaly in a data stream over time – for example, higher than expected cases of fever in a given population or geographic area.

SAGES tools are built in a modular nature, which allows the user to select one or more tools to enhance an existing surveillance system, or to use the tools en masse for an end-to-end electronic disease surveillance capability. The philosophy behind the design and development of SAGES was to create a lightweight system that end users can easily configure, use, and maintain on their own without the need for specialized information technology or software skills for a prolonged period of time.

If the proposal is granted, funding will be used to support the development of new functionality that will enable SAGES users to easily incorporate data from multiple sources into a single SAGES analysis and visualization platform.

Consortium Team:

JHU/APL has several years of experience in developing and deploying digital health open source software. As a division of The Johns Hopkins University, we plan to engage with our counterparts across various divisions of the University to inform our work. We also plan to engage the existing SAGES user community as well as other stakeholders that we work with in the global health community.

Project Description:

Typical SAGES deployments have involved health data from one or more clinic settings, based on clinical encounters with patients. The data structure is based on a single data source (analogous to the clinic logbook). JHU/APL experts believe that bringing together disparate data sources onto a single platform that supports spatial and temporal visualizations and signal detection algorithms will provide health program managers at local, district, provincial and national levels an improved ability to use data to guide implementation of timely and targeted public health interventions. Therefore, JHU/APL proposes to develop multi-data source functionality within SAGES. For example, a SAGES malaria surveillance system may include: data from clinic visits, laboratory data, vector data, environmental data, and other programmatic indicators. Such a system can also be used in the One Health context, bringing together epidemiological, clinical and laboratory information on infectious diseases in medical, veterinary and environmental sectors.

In keeping with the original design philosophy, the new functionality will enable end users to easily connect to various data sources from the user interface and enable them to be analyzed and visualized together. The user interface will be updated to enable users to manage multiple data sources, including the selection of multiple data sources in the existing analytical workbench functionality.

Data sources can also include data in other applications, such as DHIS 2 (District Health Information System 2), taking advantage of SAGES' support for comma separated value (CSV) file import and open-source application programming interface. The new functionality will enable

end users' to take advantage of SAGES' streamlined and easy-to-use analysis and visualization tools and detection capabilities on multiple data sources to create a more comprehensive capability.

Use Case:

We have learned from past SAGES engagements in Africa that there can be an ecosystem of disparate information systems and data sources in a given country. DHIS 2 is a national level system that is present in many countries in Africa. In these countries one often finds various levels of data collection and aggregation capabilities and technologies ranging from paper based, to mobile, to spreadsheet, to sophisticated, but isolated electronic systems. As an integrator for these countries, SAGES could ingest data from DHIS 2 via its REST API (Representational State Transfer application programming interface), and could be used as a disease surveillance platform side-by-side with the national system. In particular, the aberration detection algorithms integrated into the SAGES analytics platform could improve the early warning capabilities of the national level DHIS 2 installations.

For example, public health practitioners working on a malaria program in-country can utilize the dynamic dashboard capabilities within SAGES to visualize data from DHIS 2, as well as data from other sources that may be relevant to a particular effort (such as laboratory data and vector data for a malaria surveillance system). This enables data analysis and visualization to occur closer to the point of data collection.

Digital Health Technologies:

OpenESSENCE (OE) is the central analysis and visualization platform within the SAGES suite. OE is an AngularJS and Node.js based web application that uses Elasticsearch as its backend datastore. Data retrieval is executed through REST API (application programming interface) calls to the web application. The application features the ability to create dashboards and visualizations created from filterable datasets contained within the Elasticsearch datastore. A description of the technology development activities for this proposal are included below.

JHU/APL will enhance OE by providing new mechanisms for the creation and integration of data sources. These could be additional datasets uploaded via formatted CSV or Excel files, or created by connecting to other systems or databases via their exposed REST APIs and connectors. The target Digital Health systems are DHIS2, OpenMRS, and OpenLMIS. The primary interaction with these systems will be to target the common datasets offered by each and store the information inside of OE to enable the polling and replication of these datasets. Once inside OE, the data will be available for analysis and visualization.

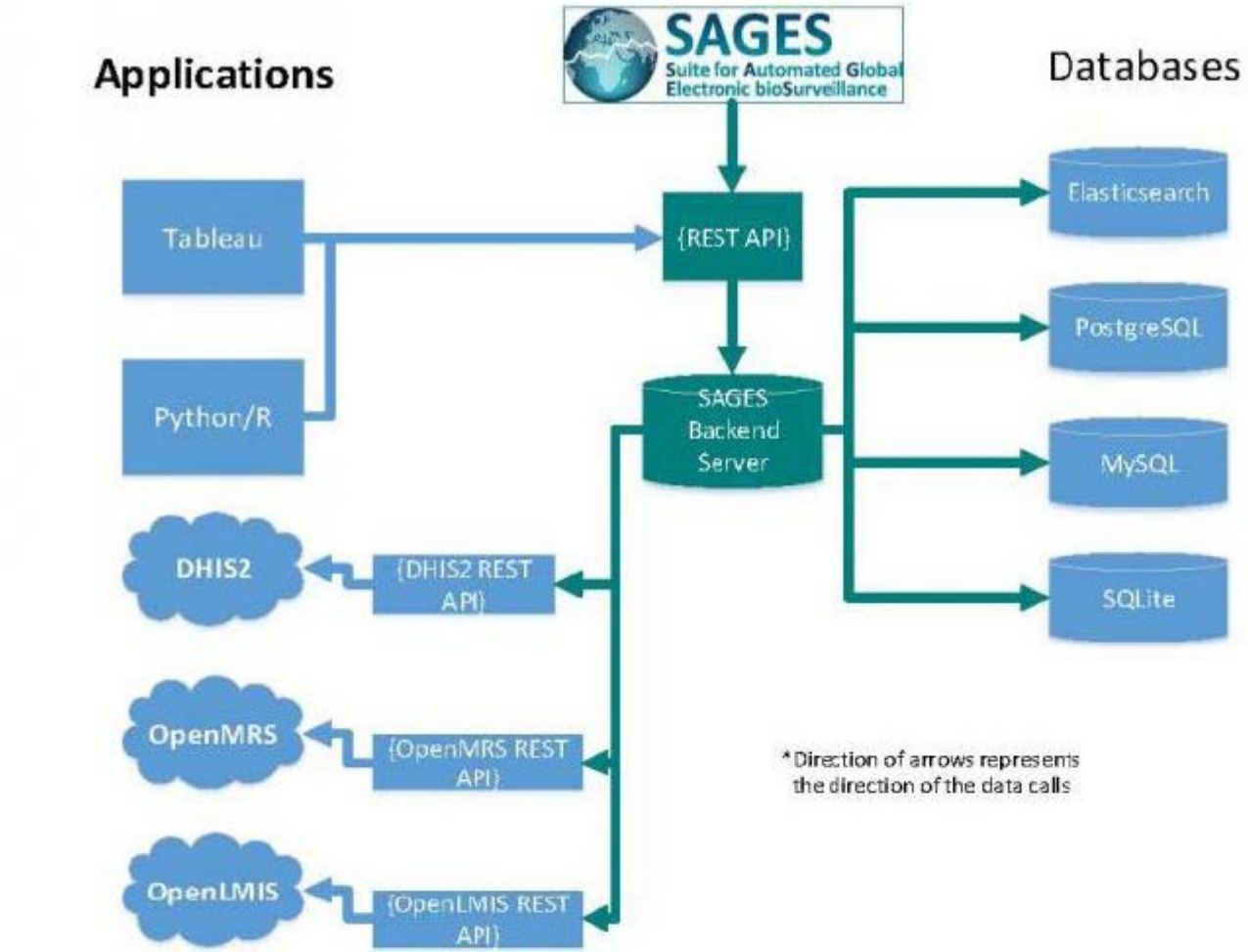
JHU/APL will also implement support for SQL-based databases (PostgreSQL, MySQL, SQLite). The default behavior for importing information from these databases will be to create unidirectional data flow from those systems to OE, with potential for bidirectional data flow depending on the target system's data flow process. The support for these databases will enable organizations with existing data collection infrastructure and storage to quickly import their data for analysis without creating intermediary files for import.

In order to streamline data import from existing systems, JHU/APL will add a new user interface (UI) workflow to OE that will assist administrators in connecting to existing databases and creating new data sources based on tables within those databases. A mechanism for the automated querying and importing of data will be created to help ease data management for administrators. The UI workflow will also include the option to select the target applications (DHIS2, OpenMRS, OpenLMIS) and guide the user through the process of integrating data.

JHU/APL will expand the REST API to enable querying of these new datasets, allowing not only OE to visualize and access the datasets, but also external tools like Tableau to connect to the SAGES database and produce additional visualizations from its store of datasets. The improved API will also help entities with proper access to OE to connect with tools in R or Python to perform more complex data analysis.

JHU/APL will also create and deliver End User, Administrator, and API documentation detailing the enhancements in Microsoft Word and/or PDF format.

A high level system diagram representing the multiple data source architecture is shown below:



Community Feedback:

JHU/APL will engage with the existing SAGES community as well as other subject matter experts with whom we have engaged across a broad spectrum of organizations (including government and academia) to help inform our efforts.

Global Maturity Model Self-Assessment:

See attachment *Digital Health Software Global Good Maturity Model – SAGES.pdf*

Work Plan, Project Deliverables & Schedule:

Product	Completion Schedule*
New version of OpenESSENCE that includes all enhancements	06/30/2019

(packaged as virtual machine)

Documentation – End User, Administrator, API	06/30/2019
Quarterly Progress Reports	Quarterly from 07/01/2018 start date
Final Report	06/30/2019

***Note these dates are based on a 01 July 2018 start date.**

Budget Narrative

Total Requested Budget Amount: $149,993

The Key Personnel for this proposal are listed below.

Shraddha Patel, Epidemiologist, Project Manager

Ms. Patel has several years of project management experience in electronic disease surveillance systems. She has participated in the implementation of these types of systems in several resource-limited settings in Asia, Latin America, and Africa.

Miles Stewart, Software Developer, Technical Lead

Mr. Stewart is the lead software developer for the SAGES project. He has extensive experience in developing and implementing electronic disease surveillance systems, including syndromic surveillance systems.

Martina Siwek, Health Surveillance Program Manager

Dr. Siwek has extensive experience in academia and government and is the Program Manager for Health Surveillance at Johns Hopkins Applied Physics Laboratory. She will be involved in engaging with the broader digital health community.

William Woodcock, Financial Manager

Mr. Woodcock provides JHU/APL's financial support by tracking project spending, preparing grant financial summary reports for internal JHU/APL use, responding to all financial questions from internal technical personnel and external sponsor requests and assisting with grant closeout activities and follow-on grant documentation.

Monika Punjabi, Program Management Assistant

Ms. Punjabi will provide administrative support for the project, including planning and logistics for meetings, conferences, and communications.

Supporting Documents: Global Goods Maturity Model Self Assessment - SAGES