Toward an Architecture for mHealth Web Data Choreography

D. Cenk Erdil and Saranya Radhakrishnan
School of Computer Science and Mathematics
Marist College
Poughkeepsie, NY 12601
{cenk.erdil, saranya.xx1}@marist.edu

Abstract

Web information systems have been increasingly used in medical informatics, with one particular focus on mobile healh (mHealth). Mobile computing devices, when used together with web information systems and integrated communication infrastructure, prove extremely useful in certain medical tasks, such as health data collection. In medical informatics, the term bioinformatics has generally been used to refer to two particular fields of study: (i) genomics, and (ii) computational biology. However, as many health sciences use computers and other computational resources, bioinformatics also has spanned other fields such as clinical informatics. In clinical informatics, one core activity is *clinical research*, which heavily relies on collecting data, mostly patient-level, as electronic health records (EHRs). In addition to being a data repository for patients, EHRs also have rapidly become a potential information repository, with recent advances in data analysis and other forms of informatics support for clinical research.

Clinical informatics tasks, such as decision support, and information access, have become core methods of research, and means for intervention mechanisms for medical professionals. Many fields of bioinformatics and health informatics have started to intersect at select fields of study, such as clinical genetic variation correlation, gene expression comparison, and genetic results in EHRs. Moreover, bioinformatics research can also benefit from common big data tasks such as data mining, knowledge representation, and computer modeling of diseases.

Advances in many medical sciences allow certain patient-level indicators to help determine the level of severity of certain epidemics and pandemics at the population-level. These indicators, called *biomarkers*, are recent focus of interest. For example the United States Centers for Disease Control and Prevention (CDC) has used biomarkers as one core measure. This approach identifies patients who are in a *critical stage*, i.e., the period shortly after incident infection. Using such biomarkers, essential population-level indicators, such as incidence rates, can be easily determined. Moreover, biomarkers produce unbiased estimates of essential indicators, and hence reduces error rates in estimates.

With the increasing use of data in health sciences, collecting data has also become a major activity in mHealth. Mobile communication devices such as smartphones, tablets, as well as highly specialized, network-connected advanced medical devices (e.g., cardiometabolic analyzers, blood samplers, and so on) are used together on the field to collect, organize, and store critical patient-level data, which then becomes main source for certain clinical informatics tasks.

mHealth has recently been gaining popularity in developing countries, mainly due to increasing availability of mobile networking infrastructure, and abundance of cellular phones. One important aspect of this observation, from the point of medical practice and health informatics, is increased access to healthcare (for patients), and health-related information (for clinical researchers) in key populations. Thus, mHealth tools and technologies help collect both patient-level and population-level health data. Patient-level data simply refers to any health-related data collected about individuals, and population-level data refers to aggregation of the same data, over certain health indicators deemed important by public health policy implementers as well as clinical researchers.

In this environment, there is a significant need to orchestrate collection of data from the field, which is hard due to many factors, including autonomous nature of data collection activities. To address this, we propose a data adaptation module, that performs an early cleaning and analysis of data, and also serves as a replication point for collected data. This data adaptation module allows flexible storage alternatives for patient-level and population-level data, while not increasing the complexity of existing data collection activities. Furthermore, our framework can be placed in existing data architectures without any alteration in existing system design. We define our preliminary design of the data adaptation module, and show how it can handle several essential data collection tasks. One of our immediate next steps is to expand our work with more details on the Choreography Layer, and devise real-world cases for immediate use. We also plan to release our code for the data adaptation module, and contribute into the open-source Open mHealth Project.

Key Words: Cloud computing, big data, web clusters, data architecture, mHealth data choreography, workflow orchestration.