LOD Approach to Creating Health-Sensor datasets built on Big Data and Graph storage technology

Neil Benn, Roger Menday, Takahide Matsutsuka*

1 Background

At Fujitsu Laboratories of Europe (FLE) we are researching and developing a technological platform that combines core components of Big Data and RDF Graph technologies. Our aim is to enable rapid development of applications – built on top of our underlying technological platform – that make insightful use of various sources of data aggregated within the underlying platform, sources of data including Linked Open Data as well as more application-specific data sources.

Within FLE we have been exploring the use of Big Data and Linked Data technologies in the domain of healthcare, all as part of our overall corporate emphasis on Human-Centric Technologies.

However, when it comes to dealing with data in the healthcare domain we encounter a number of challenges, one of which finding efficient and scalable ways to capture, store, aggregate, analyse, and visualise the variety of data from multiple data sources. Types of data include real-time patient monitoring data, electronic medical records, medical best practices, drug efficacy data, and clinical trial data, just to name a few.

FLE have recently begun a collaboration with the National University of Singapore (NUS) to establish a new joint research programme – Body Sensor Network for Disease Management and Prevention-Oriented Healthcare – focusing on in-patient and home-based monitoring solutions for disease management and prevention.

Over the programme duration, the initial objectives involve developing a wireless body sensor network, designed for continuous patient monitoring, both in hospital and in the home, and connected to a "Health Cloud"

*{Neil.Benn, Roger.Menday, Takahide.Matsutsuka}@uk.fujistsu.com

for remote processing by computers and healthcare providers. The wireless biomedical sensors form a wireless body sensor network around the body, enabling a patient's vital signs to be collected and transmitted in a context-sensitive manner. The recorded physiological data are stored and processed in the Health Cloud and, in the case of any abnormalities being detected, the Health Cloud creates flags to alert healthcare providers.

The collaboration will be deploying research and development innovations jointly developed by FLE, including the cloud data platform and a range of associated services. NUS will contribute its experience in tackling medical challenges with innovative devices, exemplified by the ultra-low powered electrocardiography (ECG) digital self-adhesive plaster technology – for comfortable and continuous ECG monitoring – they have developed.

Ultimately, it is envisaged that the research project could deliver a number of benefits for both patients and healthcare professionals. The wireless body sensor network enables the patients to remain mobile and provides them uninterrupted rest as the need for nurses physically to take regular measurements at fixed locations is removed. For healthcare professionals, wireless monitoring and automated analysis of a variety of complex and customised parameters significantly reduces their workload, enabling them to make more informed and better decisions for their patients.

2 Discussion points

This recently started research has led us to explore the possibilities of Open Data with respect to the potential benefits of having large datasets of health-sensor data.

One of the promises of open health data is that it will enable citizens to make better decisions about their health and well-being, and it will enable policy-makers and health-care providers to make better policy-decisions. This has tended to cause open health data to be skewed towards datasets that expose mainly administrative (as opposed to medical) data. And while administrative health data such as healthcare costs, hospital rankings, insurance plan data, etc. is undoubtedly important in making our healthcare organisations work better for patients as well as clinicians, the Open Health Data movement should also equally emphasise datasets that drive medical knowledge and scientific research. Opening up this kind of data would encourage more of the kinds of analysis performed by biostatisticians and epidemiologists as they try to create new medical insights.

Examples of this kind of publicly available medical data are the databases

available at the PhysioBank archives[1]. These archives contain some 4 terabytes of data available for download. Some of the most impactful datasets in the archive are the ECG datasets that have actually been curated in the late-70s and early-80s. For example the MIT-BIH Arrhytmia Database has been quite influential in this respect, making available "48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 patients studied by the BIH Arrhythmia Laboratory" One of the main purposes of this database, along with other databases such as the American Heart Associations ECG database is to enable basic research into cardiac dynamics (allowing e.g. researchers to evaluate their methods for automatically identifying arrhythmias).

We take inspiration from this kind of work and are now exploring what a Linked Open Data approach to these kinds of physiological, health-sensor datasets (suitably anonymised to ensure data privacy) would look like. Already even at this early conceptual stage we identify some of the technical challenges associated with a project like the one proposed here include developing Open APIs for consuming this kind of data as well as Open Formats for representing this data. Also, if we want to adopt a LOD approach we have to consider some of the difficulties of handling temporal data (such as long-term ECG recordings) in current representation formats.

Some of the social/cultural/political challenges include a general rebranding of Open Health Data as being more than just about hospital administrative and performance data. As well as supporting managers, financial planners, and policy-makers we need to be supporting analysts such as biostatisticians and epidemiologists. Recent movements to promote open clinical trial data are definitely a step in this right direction[2], as well as the EveryHeartBeat project[3].

FLE would like to promote the widespread use of open, structured, and linked medical data. Specifically, we would like to continue promoting and contributing to the development of standards for APIs and data formats for Open Healthcare Data.

[1] http://physionet.org/physiobank/database/

[2] See for example http://clinicaltrials.gov/ and http://linkedct.org/

[3] http://www.everyheartbeat.org/