Building a Longitudinal National Integrated Cardiovascular Database

— Lessons Learnt From SingCLOUD —

Khung Keong Yeo; Hean-Yee Ong; Terrance Chua; Zheng Jie Lim; Jonathan Yap; Hee Hwa Ho; Fazlur Jaufeerally, BSc; Khim-Leng Tong; Pipin Kojodjojo, PhD; Hwee-Bee Wong, PhD; Derrick Heng; Kelvin Bryan Tan, PhD; Arthur Mark Richards, MD, PhD; Kristine Leok-Kheng Teoh; Kenny Sin; Ngiap Chuan Tan; Simon Biing Ming Lee; Terence Lim, BSc; Andy Ta; Edwin Liok, BSc; Yee How Lau, BSc; Fei Gao, PhD; Christian Liman, BSc; Joydeep Sarkar; Anders Sahlén, MD, PhD; Tian Hai Koh; Mark Y. Chan, PhD
SingCLOUD utilizes the Ministry of Health’s Business Research Analytics and Information Network (BRAIN) platform to extract and integrate data.

**BRAIN System Architecture**

The Business Research Analytics and Information Network (BRAIN) pulls together data from different domains and gives authorized users the ability to access large amount of distributed data that are patient-linked and anonymised for study as well as analytical purposes. It comprises three key components: Business Research Analytics Insights Network (BRAIN) as the data federation platform, the National Electronic Health Record (NEHR) as the primary data source for cross cluster patient information and Singapore Cardiac Data Bank (SCDB) for cardiac specific data. Other than NEHR and SCDB, SingCLOUD also utilizes data from the National Health Identity System (NHIS), Enterprise Terminology System (ETS) MOH Data Hub and the two main data warehouses for public healthcare systems, EDW and eHINTS. BRAIN and each of its key data source is described as follows.

**A. BRAIN**

The central platform powering the SingCLOUD study is BRAIN, which was built using state of the art on-the-fly late-binding analytics architecture. The benefit of this system as a data federation platform is that it can link to data sources rapidly which are beyond what is stored within BRAIN – the system sits “on-top” of existing data sources, and it draws the most current data in a timely manner from these source systems. As data is automatically anonymized, this reduces the security requirements of the system. The BRAIN system work on a federated data grid approach to automate query execution to different data sources.

The results from the query is then subsequently linked and harmonized before anonymizing the data for analysis by the users. The process of data federation, linking, harmonizing and anonymizing are all done in one single step. The key components of BRAIN are as follows:

1. **Secure Access Channels**, enables BRAIN connection to the multiple data sources to pull and aggregate data from NEHR and different data warehouses.
2. **Patient Linking Service** utilizes the National Health Identification Service (NHIS), a patient master index that enables matching of patient records received from across the different health domain. NHIS enables BRAIN to uniquely identify people through a combination of factors—including the national identifier, name, date of birth, address and other demographic information.

3. **Enterprise Terminology Service**, BRAIN leverages on international standards for harmonizing data across multiple source systems.

4. **Anonymization Service**, supports the de-identification of patient data before presenting this to end users or prediction model. The anonymized data has been verified to be HIPAA Safe Harbor Privacy Rules compliant.

**B. National Electronic Health Record (NEHR)**
The main source of data is from NEHR. In April 2011, Singapore’s launched the first release of NEHR system that provides a common access point for medical information for its population. To date, NEHR collects near-real time medical data from all the public hospitals, specialty centers, polyclinics and selected community hospitals, nursing homes, private hospitals and general practice (GP) clinics. Daily, data from NEHR is extracted, harmonized and linked onto Business Research Analytics Insights Network (BRAIN), a federated data grid that provide a near real-time and secure environment of the development of predictive models. The types of data collected include patient demographics, immunisations, allergies, medical alerts, visits and visit summaries, appointments, referral notes, medical and dental screening data, diagnoses and problem lists, medications, procedures, laboratory, radiology and cardiac reports. NEHR was deployed in a gradual and phased manner with all public hospitals included by 2011.

**C. Singapore Cardiac Data Bank (SCDB) (via eHints Data Warehouse)**
SCDB is a clinician-led collaborative effort by all public cardiac departments in Singapore (Annex 1). It was started in 2000 as an audit and quality improvement registry program and maintains a registry of patients with cardiovascular disease and their treatment procedures. SCDB data modules include the following; acute myocardial infarction (SCDB collected data up to 2007; MOH collected data after 2007); Coronary Angiography and Percutaneous Coronary Intervention; Cardiac
Electrophysiology (including Ablation and Device Implantation); Cardiac Surgery (including Coronary Artery Bypass Graft (CABG) and Valve Surgery); and heart failure. SCDB utilizes data forms developed in consultation with the hospitals. Each hospital provides funding for their own clinical coordinators who are responsible for data collection. Data is then sent to the central analytics team where integration of data is performed and generated for annual national audit meetings of each domain. Clinical audit of the data collected on cardiac catheterization and coronary intervention is performed regularly by trained coordinators. The audit includes 300 fields from 300 cases or 3% of all records, whichever was lower in each of the participating hospitals. The last audit performed in 2016 showed an accuracy (concordance) rate ranging from 95 to 98%. Data from SCDB is transmitted to SingCLOUD via eHINTS.

D. Data Warehouses: eHINTS and EDW
From 2007 to 2014 the IT systems of all public hospitals placed information in 2 broad data warehouses serving the needs of the 2 original hospital clusters in Singapore (SingHealth and National Healthcare groups). These were known as eHINTS and EDW (Enterprise Data Warehouse) respectively. The data warehouses are managed by a government-controlled statutory board known as IHiS (Integrated Hospital Information Systems) under the control of the hospital chief information officer and chief medical information officer. Use of data from the warehouses requires approval by IHiS and the respective information officers. Singapore’s public healthcare system treats approximately 80% of the population and manages 96% of acute myocardial infarction patients.

E. The National Health Identity System (NHIS)
NHIS is the national body that consolidates identity in health records. This is necessary as identity may be changed as a result of changes in name, marital status, residency or citizenship status. The NHIS is critical in ensuring robust identification of patients in our healthcare system. A key component of identity is the National Registration Identity Card (NRIC), with a unique identification number. This unique NRIC number is used throughout Singapore’s public healthcare institutions, and allows tracking of outcomes across the country. The NRIC is also used in all
government services for administrative claims and the recording of births and deaths.

**F. Enterprise Terminology System (ETS)**

ETS is a platform for harmonizing terminology across the healthcare system. This includes medications, laboratory tests, diagnosis and procedural codes. However, it did not include highly specific specialty terms. For example, ETS does not harmonize specific nomenclature in echocardiography such as criteria for ejection fraction, valvular heart disease or grading of diastolic function.

**G. MOH Data Hub**

The MOH data hub collects and analyses administrative, financial and claims data from all public hospitals and from private institutions for central planning purposes. Data include diagnosis codes, procedural codes, Diagnosis-Related Grouping (DRG) codes, dates of admission, discharge and procedures, bill size, and type of housing.