Design of a Decision Support System for an Electronic Medical Record System

Submitted by

Brian Mrkonjic

B00074113

Brian_Mrkonjic@hotmail.ca

This Internship was completed for

Concertia Technologies Inc.

1869 Upper Water St

Halifax, Nova Scotia

B3J 1S9

In partial fulfillment of the requirements of the Master of Health Informatics Program

Dalhousie University

Internship for the period May 17th to September 13th, 2013

Date Submitted Sept 15, 2013

Acknowledgement

This internship report was written by Brian Mrkonjic in partial fulfillment of the requirements for the Master of Health Informatics Program, Dalhousie University. This report has not received any previous academic credit at Dalhousie University or any other institution.

I would like to thank Mr. Ross MacDougall, President, Concertia Technologies Inc., for providing me with the opportunity to complete this internship. Mr MacDougall was responsible for securing this opportunity in response to a request for proposals from the Nova Scotia Department of Health and Wellness and provided project management expertise throughout the course of the project.

I would like to thank Mr. John MacCallum, Senior Project Manager, Nova Scotia Department of Health and Wellness, and Mingxiu Li, Report Manager, Primary Healthcare Information Management (PHIM) for their guidance and direction in the areas of project scope, business requirements and data analysis.

.

Executive Summary

The intent of the internship was to attempt the design and implementation of a Decision Support System (DSS) in support of the Primary Healthcare Information Management (PHIM) program, an initiative within the Nova Scotia Department of Health and Wellness. The PHIM program supports the deployment of an Electronic Medical Record (EMR) system to primary healthcare practitioners throughout the province of Nova Scotia. The EMR system is used to capture information that is related to patient encounters with primary healthcare providers and community based specialists. The EMR data store represents a rich source of primary healthcare data that is sought after by many healthcare stakeholder groups across the province. The primary goal of the internship was to attempt to design and implement a DSS using primary healthcare data extracted from the PHIM. Implementation of the DSS would facilitate improved access to primary healthcare data which would provide unique opportunities to improve efficiency and decision-making in a wide range of areas including public and population health (eg chronic disease monitoring), clinical research, resource utilization and public health planning.

This internship was completed in co-ordination with Concertia Technologies Inc (Concertia), which was contracted by the Nova Scotia Department of Health and Wellness to design and implement the DSS for the PHIM program. All work was completed within the offices of Concertia in Halifax, Nova Scotia. The author was responsible for completing the following tasks:

1. Understanding the business requirements of the PHIM program as they pertained to primary care reporting requirements
2. Reverse engineering of the internal data structures of the Nightingale EMR data store
3. Development of a data mapping document
4. Completion of a high-level technical design for a DSS

5. Implementation of some or all of the DSS

The design and implementation of the DSS was only partially completed during the course of internship engagement due to unforeseen constraints beyond the control of the author and the project team. Specifically, the project team was not able to gain access to the EMR production database or a representative thereof, which prevented the author from performing a definitive data analysis, validation of the design of the Change Data Capture (CDC) mechanism and stress test of the proposed DSS schema design. As a result, the internship was limited to the design and partial implementation of the Primary Operational Decision Support System (PODSS) and to research into the approach for de-identification of patient data for eventual transfer to the Health System Decision Support System (HSDSS).

Upon completion of the internship, the student was able to gain knowledge and experience in the following areas of health informatics:

1. Electronic Medical Record systems.
2. Data Warehouse Design (Normalized versus Dimensional).
3. Data anonymization.

1. Table of Contents

2. Table of Figures

3. Introduction

In 2003, the Province of Nova Scotia began a process to implement the Nightingale On Demand™ Electronic Medical Record system (Nightingale) in support of the province's primary healthcare providers. The Primary Healthcare Information Management Program (PHIM), which is hosted within the Nova Scotia Department of Health and Wellness (DOHW), was given responsibility for the implementation and support of the EMR initiative. Today, approximately 460 of the province's primary healthcare physicians are actively using the Nightingale system. In addition to the Nightingale system, there are several other EMR systems in limited use within the province of Nova Scotia, including Dymaxion's Practimax™, Clinicare (QHR) and several other client-server based EMR systems.

Currently, PHIM has achieved an EMR adoption rate of approximately 60% of supported practitioners (primary healthcare physicians and community based specialists), out of an approximate total of 2000 within the province, and expects to achieve a 70% adoption rate of all primary health care providers and 20% of community based specialists in 2013-14. Nova Scotia's EMR adoption rate compares well with other Canadian provinces and the fact that a majority of primary healthcare providers will be suing one EMR platform is somewhat unique within the Canadian primary healthcare system as most provinces employ a wider range of EMR vendor solutions which leaves the resulting EMR data stores in a more fragmented state, thus making it much more challenging to collect and integrate primary healthcare data into a single pan-provincial or pan-Canadian view.

The Nightingale EMR data store represents a rich source of primary healthcare data that is sought after by many healthcare stakeholder groups across the province. It is thought that increased access to this data store could provide the opportunity to improve planning and decision making in areas ranging from studying the long-term

health benefits of the primary healthcare provided to the province's patient population, to identification of new ways to improve the quality and efficiency of health services delivery. Regardless of the intended use of the data, the desire for increased access to the EMR data is consistent across EMR stakeholders as they believe access to the data will improve their ability to develop and support evidence based decision making.

In 2010, the DOHW conducted an EMR DSS Strategy review. The review identified several key stakeholder communities that would benefit from the secondary use of primary healthcare EMR data. The review defined a high-level technical architecture for a DSS and identified several issues pertaining to data governance. A recommendation was made to proceed with the development of a Decision Support System (DSS).

In late 2012, DOHW partnered with Concertia to design and implement a DSS for the Nightingale EMR instance hosted by PHIM. Project tasks that were defined to be in scope for the project included:

1. Identification of a subset of EMR data attributes that would satisfy a significant portion of the operational reporting needs of the PHIM user community, including immediate users of the system such as primary healthcare providers as well as other stakeholders within the DOHW such as public health epidemiologists and healthcare economists
2. Reverse-engineer the Nightingale On Demand™ database structure to allow for extraction of the subset of clinical data
3. Design and implement a database Repository to allow for reporting and data analysis functions against the subset of clinical data

The author was given responsibility for completing the following tasks:

1. Determination of the business requirements of the PHIM program as they pertained to primary care reporting.
2. Reverse engineering of the Nightingale EMR data store.

3. Development of a technical design for a DSS.
4. Implementation of some or all of the DSS.

4. Description of Organization

This internship was completed under the direction of Concertia in Halifax, Nova Scotia, as part of a professional engagement with the DOHW. Concertia is a consulting company that provides information technology and business solutions to clients located within the Maritime Provinces.

The DOHW is a province of Nova Scotia government department which is responsible for the provision and delivery of healthcare services to the province. Management and delivery of healthcare services is provisioned by nine district health authorities (DHA) and the Izaac Walton Killam hospital (IWK). The health authorities are responsible for delivery of health care services to individual residents and are also responsible for hospitals, community health services, mental health services and public health programs. The PHIM program exists within the DOHW and is responsible for the roll-out of the Nightingale system to the end-user community. A key goal of the PHIM program is to enhance patient care by providing the ability to healthcare providers to access patient clinical information in real time. In addition, benefits of the EMR system include accessible patient healthcare data at all points of care, more effective and comprehensive sharing of patient healthcare data, especially within community health settings and a reduction of medical errors by improving the accuracy and reliability of patient records.

The Nightingale On Demand™ system is a web-based ASP EMR solution produced by the Nightingale Informatix Corporation. The application is centrally hosted within the province by Health Information Technology Services Nova Scotia (HITS-NS). HITS-NS is a provincial agency mandated by the DOHW to provide IT technical support services to interested stakeholders, such as the PHIM program. The

Nightingale database supports over 1700 tables, some of which contain tens of millions of records and the database currently requires 1.5 terabytes of storage.

5. Discussion of Work Performed

The author was responsible for completing the following tasks during the internship:

1. Determine the business requirements of the PHIM program as they pertained to primary care reporting and develop an initial design of the DSS.
2. Reverse engineer the Nightingale EMR data store.
3. Develop a technical design for a DSS.
4. Implement the DSS.

5.1 Business Requirements

The EMR DSS project was originally intended to be designed based on the work of the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) project. CPCSSN project is a pan-Canadian collaborative effort between the Public Health Agency of Canada (PHAC), the College of Family Physicians of Canada and the Canadian Institute of Health Information (CIHI) and its goal is to design and implement a multi-disease electronic medical record surveillance system focused on chronic disease management. Its primary goal is to create and maintain a pan-Canadian epidemiological database using longitudinal patient primary care data that is collected and validated from EMRs in operation across Canada. The resultant database will be utilized to allow for the monitoring of chronic care across Canada.

The CPCSSN database stores de-identified patient health information, collected under agreement from practitioners, which is related to eight chronic diseases and neurologic conditions:

- chronic obstructive pulmonary disease
- depression
- Diabetes

- Hypertension
- Osteoarthritis
- Alzheimer disease and related dementias
- Epilepsy
- Parkinson disease

Currently, the data is collected quarterly from eight provinces and nine EMR vendor platforms.

Critical issues:

- Validation of patient information is complex. For example, is a patient that has been identified as a diabetic really a diabetic?
- Validation of individual data points (weight, blood pressure) is complicated due to the use of many different EMR vendors, each using proprietary systems to record this information.

Potential benefits:

- Provide surveillance reports for stakeholders including policy makers, health planners, health economists and government agencies and practitioners.
- Feedback reporting can be targeted to individual practitioners that will allow them to compare their patient outcomes to national averages.
- Integration with Canadian Primary Care Sentinel Surveillance Network data sources is out of scope for this project.

The DOWH's original intention was to leverage the work done by CPCSSN project to develop a broader DSS solution that would utilize the data captured by the PHIM program and provide for a wider variety of end-uses. Unfortunately, our project was not able to secure an information sharing agreement with CPCSSN within the EMR DSS project time-frame that would allow the EMR DSS team to access CPCSSN deliverables. As a result of discussions conducted by the author with DOHW staff, it was determined that the design of the DSS should satisfy the following high-level requirements:

1. Provide a subset of Nightingale data attributes that could be used to satisfy the majority of operational reporting needs of the current end-user community (ie primary healthcare providers).

2. Provide a de-identified set of primary healthcare data attributes that could be made accessible to users within the DOHW to improve the delivery of primary care services across the province. Such uses might include chronic disease monitoring, epidemiology, health economics, etc.

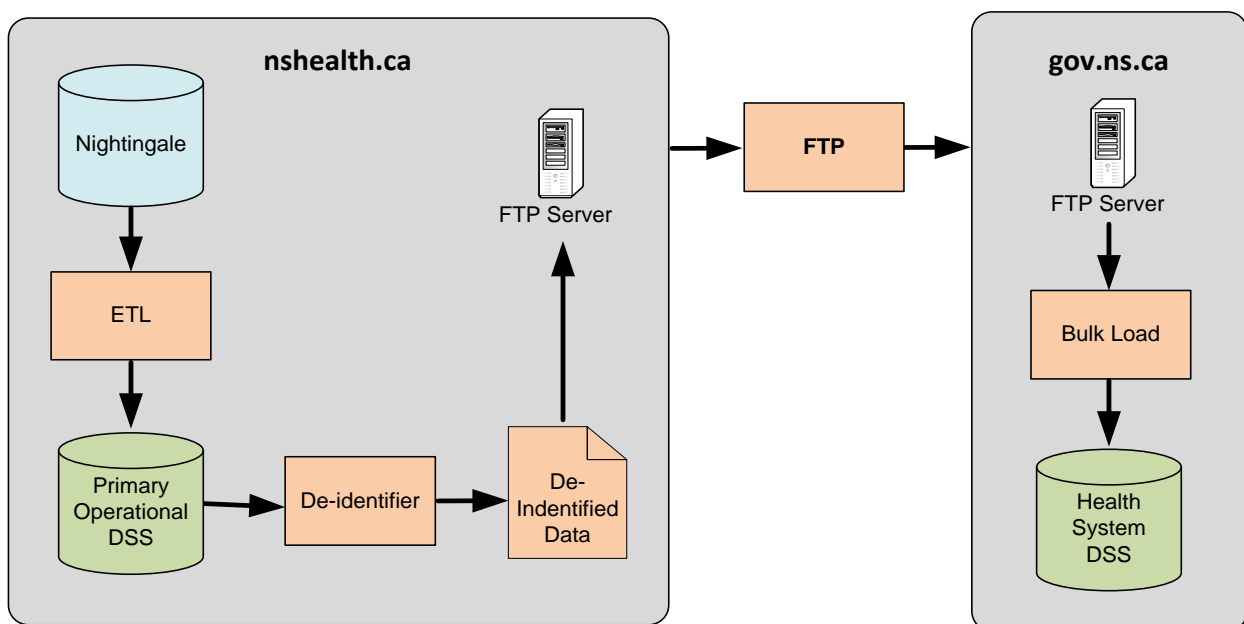A high-level architecture diagram of the design was developed and is shown below:

Figure 1 – EMR DSS High Level Architecture

The design envisioned the implementation of two reporting databases:

1. Primary Operational Decision Support System (PODSS)
   The PODSS database would be hosted within the HITSNS network and would provide Nightingale end-users with a set of Nightingale data attributes that would allow them to satisfy their operational reporting needs.

2. Health System Decision Support System (HSDSS)

    The HSDSS database would be hosted within the DOHW network and would provide DOHW staff with a de-identified subset of PODSS data attributes.

The PODSS would serve as the data source for the HSDSS. The PODSS would be populated by an Extraction-Transformation-Load (ETL) process that would execute on a periodic (ie batch) basis. Next, a process would be used to extract and de-identify selected data from the PODSS for subsequent transfer to the HSDSS via flat file format. These files would then be loaded into the HSDSS via a bulk load process.

Due to unforeseen technical issues, the project team was not able to complete implementation of the processes responsible for the de-identification and loading of data from the PODSS to the HSDSS.

Lastly, the author worked with DOHW staff to identify a set of approximately ten initial reports intended for use against the PODSS.

5.2 Reverse Engineering of the Nightingale Database

DOHW staff provided a listing of data attributes which represented a subset of the Nightingale EMR dataset which had been deemed sufficient to allow Nightingale end-users to produce the required reports to meet ongoing operational needs. The Nightingale system is a proprietary system and very little technical documentation was available to allow the EMR DSS team to gain a clear understanding on the system's data structures and data processes. The author was responsible for developing a strategy to reverse engineer the Nightingale database to allow for extraction of the data attributes that had been identified by DOHW staff as candidates for the DSS. The author utilized two main approaches to reverse engineer the Nightingale database:

1. Inspection of pre-existing Nightingale database objects (views and stored procedures) to identify selection and join logic that could be utilized to extract targeted data attributes.

2. Creation of test patient data using the Nightingale application followed by manual inspection of the database to determine the storage location of inputted data attributes.

These efforts led to the development of a data mapping document which listed over 200 Nightingale data attributes along with their storage location within the Nightingale database schema and instructions for extraction.

5.3 DSS Design

Once the data mapping document had been completed, the project proceeded to the design phase of the DSS. Research was conducted to determine the best approach to the design of the data schema in order to meet the following requirements:

1. Efficient storage of data.
2. Fast execution of queries.
3. Ability to support multi-dimensional reporting (ie slice and dice).
4. Ability to secure access to data such that DSS end users should only be able to view data to which they are authorized.

Research was conducted that examined the merits of two generally accepted approaches to data warehouse design; the "Immon" method which espouses the use of a staging area to prepare data for eventual loading into a third normal form relational data schema and the "Kimball" method which outline a four step methodology to implement star schema data marts.

Further work was conducted to research the subject of de-identification of patient data. Patient data was to be de-identified before being loaded into the HSDSS. This work was related to the implementation of the HSDSS which fell out of scope for the internship due to project time and cost constrains.

5.4 DSS Implementation

Due to unforeseen technical issues, implementation was limited to the design and implementation of the PODSS along with a set of data extracts which were used to test the accuracy of the data mapping deliverable. Technical issues which prevented the author from completing his assigned tasks were:

1. Inability to gain access to the EMR production database or a representative thereof, which prevented the author from performing a definitive data analysis, validation of the design of the Change Data Capture (CDC) mechanism and stress test of the proposed DSS schema design.

2. Provisioning of a development environment was significantly delayed which reduced the time allowed for implementation of the ETL modules, DSS Repository schema, and CDC logic.

6. Relating the Internship to Health Informatics

As a result of work conducted during the internship, the author was able to achieve a deeper understanding of the following subject areas related to health informatics:

1. EMR Systems
2. Data Warehouse Design
3. Data De-identification

6.1 EMR Systems and Data Quality

One of the primary challenges of any electronic systems is to ensure that data is captured in a clear, consistent and reliable manner. In general terms, data can be maintained in an electronic information system in one of two ways: structured or unstructured:

1. Structured

    Structured data is defined by a known set of rules and constraints which unambiguously describe its state and characteristics. These rules can be

used to validate the data as it is captured and thus ensure its quality.

2. Unstructured Data

   Unstructured data, generally in the form of free-form text, typically utilizes only simple rules for validation and, as a result, data quality is generally of an inconsistent and poor quality. Examples of unstructured data would be clinical notes, and to a certain extent, Lab and DI Reports as they are stored as XML documents within the Nightingale EMR.

After performing an analysis of the Nightingale data schema, the author learned that the EMR utilized unstructured fields to capture a significant portion of patient healthcare data. The use of free-form data fields allows for a wide variance in the way that users might choose to enter patient data. For example, some users may prefer to avoid the use of discrete data fields (pick-lists, validated fields, etc), and instead, choose to enter all or most of their data into one or more "note" fields. In addition, clinical notes, which are typically captured as unstructured text, can contain misspellings, alternate terminology, abbreviations, shorthand annotations, etc. This user behavior can have a significant impact on the ability of the DSS to provide usable data for reporting purposes.

Many EMR systems rely heavily on unstructured data fields for data capture due to the complex nature of clinical data which usually requires complicated data entry validation processes. It is also much easier to build EMR systems that use unstructured fields to capture data due to the reduced need to validate data and build complex data capture screens.

Advanced data processing techniques such as Natural Language Processors (NLP), such as IBM's Content Analytics solution, can potentially be used to convert free-text

data into discrete components; however, this approach was deemed out of scope for the current project phase.

6.2 Data Warehouse Design

Although not strictly related to health informatics, the design of the data warehouse was researched by the author as part of this internship. The author learned that there are two generally accepted approaches to data warehouse design:

1. Immon (normalized)

    The Immon approach specifies that the data warehouse schema design should be modeled such that data entities are categorized and organized in 3rd normal relational form. The advantage of this approach to schema design is that it is relatively easy to add or modify data to the data warehouse as a result of changes in source systems. The disadvantage of this approach is that the resulting data schema is difficult for end users to understand and data queries must typically join together many tables to form desired result sets which can be inefficient in large enterprise databases.

2. Kimball (dimensional)

    The Kimball approach categorizes data as either a Fact or a Dimension. A Fact represents data that can be counted or tabulated. A Dimension gives perspective to, or describes, the fact data. Facts and Dimensions are organized into data marts using a Star or Snowflake data schema. Data marts tend to be focused on providing query support to a narrowly defined set of business requirements. The advantage of this approach to schema design is that the resulting schemas tend to be more easily understood by end-users and the queries that are implemented to work with the schemas are very efficient because there are typically few joins involving large data tables. A potential disadvantage of the Kimball method is that the individual data marts may be difficult to modify in response to a change in business requirements or underlying processes.

A critical feature task of the data warehouse design was the development of the CDC mechanism. CDC is the process that is used to track data changes within source systems such that the data changes can be incorporated into the data warehouse. It was important to minimize any load to be placed on the Nightingale database server during the execution of the CDC mechanism.

6.3 Data De-identification

The initial high-level technical design for the DSS included a process to extract and de-identify patient data from the PODSS for eventual incremental load into the HSDSS where the data could be used to understand primary healthcare across the province. The subject of de-identification of personal data has gained significant importance in the healthcare sector due to the sensitive nature of patient data and the statutory requirement to protect patient privacy, and to the emergence of "big data" which describes technological advances that allow for the merging and data mining of large distributed databases for the purposes of producing individualized personal profiles of users, essentially de-identifying these individuals.

The process of de-identification or data anonymization is a critical step in the attempt to make valuable healthcare more accessible to interest stakeholders. It is generally understood that it is impossible to completely de-identify a dataset without losing most, if not all, of its information utility. When de-identifying a data set, a trade-off must be considered, wherein the higher the degree of anonymization that is achieved, the lower the information content or utility of the data set.

There are two general techniques for de-identifying a data set:

1. Data Masking

    Data masking is a process that is applied to direct identifiers to eliminate their ability to be used to identify a single instance. Examples of direct identifiers are an individual's name, social insurance number, employee number, etc.

Common techniques for masking direct identifiers include substitution whereby a direct identifier value is replaced with a substitute key value, deletion (or nullification) whereby the data value is removed from the dataset entirely and encryption whereby the value is encrypted.

2. Statistical De-identification

   Statistical de-identification involves the analysis and manipulation of indirect identifiers. Indirect identifiers such as a person's Date of Birth, postal code or date of medical appointment do not directly identify an individual on their own, but, when combined together can potentially re-identify a single instance. To remediate indirect identifiers, an analysis is performed and one or more data transformations are applied to reduce the ability to use the identifiers to re-identify an individual while still maintaining an acceptable level of information utility. For example, a Date of Birth value might be converted into a range value, thus maintaining an appropriate level of utility of the information conveyed by an individual's age but at the same time reducing the ability to use the value, in combination with other indirect identifiers to re-identify the individual.

Another important factor to consider when attempting to de-identify a data set, is to consider the target consumers of the data set. If the consumers are highly trusted, the level of de-identification might be reduced thus increasing the utility of the information contained in the data set.

7. Analysis and Solution of a Health Informatics Problem

The health informatics problem that the internship attempted to address was how a DSS could be designed and implemented to make clinical healthcare data that resided in a proprietary EMR system more accessible to the EMR end users and to interested stakeholders within the DOHW.

7.1 Data Warehouse Design for Patient Data

The project team decided to implement the PODSS based upon the Immon method as this made the resulting schema more robust in terms of reacting to changes in the source data schema which ultimately is subject to change by the system vendor. Although, the HSDSS was not implemented, it was thought that the Kimball approach would be the optimal schema design due to its ability to support highly efficient queries against large data tables and its more intuitive schema design would be beneficial to DOHW end users who may not be familiar with the Nightingale system.

The final design of the PODSS data schema was driven by the data mapping document which listed the minimal subset of Nightingale data attributes that had to be supported in the DSS. A process of design was commenced wherein the data mapping attributes were organized and categorized into entities within the data schema. Next, an ETL module was designed and implemented. This work was completed in stages and at the end of each stage, a data validation script was written to test the ability of the schema to support the reporting needs of the end users. This was done by populating test data into the PHIM Training environment, executing the appropriate ETL module to populate corresponding target tables in the DSS and then executing the data validation scripts against the DSS.

Clinical data residing in the Nightingale database was moved into the DSS using a three step process commonly referred to as Extraction-Transformation-Load (ETL). During the Extraction step, target data is retrieved from the source system and stored in Stage tables which were implemented in the DSS database. The Stage tables were designed to mimic the source Nightingale tables and minimal effort made to cleanse the data as it was loaded. This approach to Stage table design simplified the extraction process and minimized the interaction between the extraction process and the Nightingale host database server.

There are two general approaches for loading data into Stage tables:

1. Full Load (also commonly referred to as "Trunc and Load")
   In this approach, all pre-existing data in a Stage table is deleted before the data is loaded, in its entirety, from the source system (ie Nightingale). This method is relatively easy to implement; however, it is generally not suitable when the Stage tables must accommodate large amounts of data due to the need to transfer large amounts of data over a network

2. Incremental Load
   In this approach, only the records from the source system which have undergone change since the previous execution of the extraction process are extracted and merged with the data that currently exists in the Stage table. The incremental load approach generally requires less time to execute since it is extracting only a subset of data; however, this method usually requires more complex code to implement since it must interpret changes to source data (insert\update\delete) and apply them to the pre-existing staged data.

Once the raw data was loaded into Stage tables, a CDC process had to be designed. Due to an inability to access a representative copy of production data, the author was not able to determine a final design for the CDC mechanism; however, the author was able complete an initial analysis of potential design options.

Change Data Capture describes the mechanism that is used to detect and track changes to data over time. The CDC mechanism typically creates\updates meta-data, such as status flags and date time stamps, such that follow-up processes are able to integrate the data changes with a target data store. In the DSS, CDC will be utilized to detect data changes in the Nightingale database and ETL modules will use the CDC metadata to integrate the data changes into the DSS Repository.

There are several possible approaches to consider for CDC design:

1. Brute Force

   This method utilizes a record-by-record comparison of columns between the source and destination data stores to detect data changes. It is computational expensive and may not be suitable for processing the larger tables in Nightingale. An optimization of this method utilizes Checksums to avoid column-by-column comparisons. Checksums are calculated for each record based on the columns that are of interest and the Checksums are stored in the Stage table. During the Extraction process, the records from the source system are retrieved and a new Checksum is calculated and compared to the Staged version. If they differ, a change has been detected. Data changes arising from record insertions and deletions are detected by performing table lookups based on primary key values.

2. SQL Merge

   The SQL Merge command was introduced by database vendors to facilitate common data warehouse tasks such as CDC. The command can be used to compare two results sets and generate SQL Insert, Update and Delete statements based on the results of the comparison. The command is useful for synchronizing tables at destination.

3. MS SQL Server CDC and Change Tracking

   SQL Server provides an out-of-the-box CDC solution (Enterprise, Developer editions). This feature must be configured for each target table. Once configured, a "shadow" CDC table is created which effectively mimics the source table in structure with the addition of several columns which contain metadata that track the nature of data changes made to the source table. Each time a data change is made against a record in the source table, a record in created in the corresponding CDC table that essentially documents the nature of the data change. This CDC method is costly in terms of database space requirements (at source) bit provides a rich source of

metadata that describes the exact nature of data changes to each record in a table.

4. Third Party Vendor

   Several third party vendors offer CDC solutions. Typically, these tools will process the transaction log of the source data store to generate a list of insert\update\delete metadata that an ETL process can then consume.

The project team was not able to design and test CDC solutions due to the lack of access to a representative copy of Nightingale production data. An initial investigation has determined that a CDC solution involving SQL Server Change Tracking to facilitate an incremental data load of Stage tables and then use of the SQL MERGE command to apply data changes from the Stage tables to the DSS Repository tables using a set based approach might be the most promising CDC solution to investigate

8. Conclusions

The author gained valuable insight into the challenges of working with patient healthcare data and some of the unique issues that are involved when working with patient data. Specifically, as a result of working with the Nightingale EMR, the author learned about potential deficiencies of modern EMR systems in regards to their ability to capture complex medical data in a way that preserves high data quality. The author also gained knowledge in the subject domain of data de-identification. The challenge of data de-identification, especially as it relates to patient healthcare data, is the trade-off that must be made between preserving a patient privacy and the information value of the data.

9. Recommendations

The author recommends the following follow-up actions in support of an ongoing initiative to complete the implementation of the PODSS and HSDSS systems:

1. Once access has been provided to the project team to a representative copy of the Nightingale production database, a data analysis should be conducted to better understand the overall data quality of the primary healthcare data that is stored within the Nightingale database. Specifically, more understanding needs to be gained in the following areas:

    a. Unstructured data elements – what is the nature of the data that is stored within these fields and how can the data be processed to make it more easily reportable? Is there potential to use NPL technology to increase data quality of the DSS especially as it pertains to data that is captured in unstructured, free-form text fields within the Nightingale EMR?

    b. Can end user behavior in regards to data entry be modified such that greater use is made of structured data capture fields within Nightingale? Changing end user behavior has great potential to improve overall data quality and simplify data extraction from Nightingale into the PODSS.

2. The DOHW should perform an analysis of its departmental business requirements for a standardized approach to patient data de-identification. The department should consider an evidence based methodology and toolset to ensure that it can share valuable data to a variety of interested third parties in a secure and safe manner that ensures the integrity and utility of healthcare data while at the same time ensuring the privacy of patients.

10. References

1. Immon, H., W., (2002). Building the Data Warehouse. New York: Wiley Computer Publishing.
2. Kimball, R., Ross, M., (2002). The Data Warehouse Toolkit, The Complete Guide to Dimensional Modeling. New York: Wiley Computer Publishing.

3. Privacy Analytics Inc., (2013). Product information regarding the PAROT application. Retrieved from http://www.privacyanalytics.ca/software/parat/

4. Aitken, J., Gardner, N., Alteen, T., Booth, P., Hansen, C., Giokas, D., … Ross, V., (2012) Draft Pan-Canadian Primary Health Care Electronic Medical Record Content Standard, Version 2.0 - Business View. Canadian Institute for Health Informatics.

5. Kimball, R., Caserta, J., (2004). The Data Warehouse ETL Toolkit, Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data. New York: Wiley Computer Publishing.

6. Fraser, R., Willison, D., (2009). Tools for De-Identification of Personal Health Information. Pan Canadian Health Information Privacy (HIP) Group.

7. Cavoukian, A., El Emam , K., (2011). Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy. Ottawa. Information & Privacy Commissioner.