Presented at *HIC 2011 Conference By Health Informatics Society of Australia (HISA)*. Brisbane, Australia.

Clinical Data Integration Approach Using SAS Clinical Data Integration Server (CDI) Tools

Saliya Nugawela,[1] Tony Sahama
Computer Science Discipline, Faculty of Science and Technology
Queensland University of Technology (QUT)
Brisbane, Australia

Abstract. The decisions people make about medical treatments have a great impact on their lives. Health care practitioners, providers and patients often make decisions about medical treatments without complete understanding of the circumstances. The main reason for this is that medical data are available in fragmented, disparate and heterogeneous data silos. Without a centralised data warehouse structure to integrate these data silos, it is highly unlikely and impractical for the users to get all the information required on time to make a correct decision. In this research paper, a clinical data integration approach using SAS Clinical Data Integration Server tools is presented.

Keywords. Clinical data integration, clinical data warehouse, SAS data integration.

Introduction

Informed decision making in health care is vital to provide timely, accurate and relevant advice to the right person, to reduce the cost of health care and to improve the overall quality of health care services. Since medical decisions are very complex, making choices about medical decision-making processes, procedures and treatments can be overwhelming [1].

One of the major Information Technology (IT) challenge in clinical practice is how to integrate several disparate, standalone information repositories into a single logical repository to create a single version of truth for all users [2]. A massive amount of health records, related documents and medical images created by clinical diagnostic equipments are generated daily [3]. Medical documents are owned by different hospitals, departments, doctors, technicians, nurses and patients. These valuable data are stored in various medical information systems such as HIS (Hospital Information System), RIS (Radiology Information System), PACS (Picture Archiving and Communications System) and etc., in various hospitals, departments and laboratories being primary locations [3]. These medical information systems are distributed and heterogeneous (utlising various software and hardware platforms including several configurations). Such processes and data flows have been reported by Zheng *et al*., (2008) [3].

"*All medical data are located in different hospitals or different departments of single hospital. Every unit may use different hardware platforms, different operating systems, different information management systems or different network protocols. Medical data is also in various formats. There are not only a tremendous volume of imaging files (unstructured data), but also many medical information such as medical records, diagnosis reports and cases with different definitions and structures in information system (structured data).* "[3].

This causes Clinical Data Stores (CDS) with isolated information islands across various hospitals, departments, laboratories and related administrative processes, which are time consuming and demanding reliable integration [4]. Data required to make informed medical decisions are trapped within fragmented, disparate and heterogeneous clinical and administrative systems that are not properly integrated or fully utilised. Ultimately, health care begins to suffer because medical practitioners and health care providers are unable to access and use this information to perform activities such as diagnostics, prognostics and treatment optimisation to improve patient care.

In this paper, the use of the latest data integration tools of one of the leading business analytics and business intelligence software to integrate disparate and heterogeneous medical resources are presented.

[1] Corresponding Author.

The rest of this paper is organised as follows: proceeding section introduces the related works on integrating data into a Clinical Data Warehouse (CDW). Section 2 presents use of SAS Enterprise Guide 4.3 to create a CDW. How the latest SAS Clinical Data Integration 2.2 and SAS Clinical Data Standards Toolkit 1.3 can be of a help in creating a data warehouse structure for medical resources are described in Section 3. Finally the conclusion is given in Section 4.

Integrating Data into a Clinical Data Warehouse (CDW)

"*Given its sensitive nature, diverse storage formats, and inherent privacy issues, healthcare data can benefit greatly from a data warehouse that integrates the data and ensures its correctness*." [5].

Data integration deals with incorporating all types of organisational data that are scattered throughout into a unified data warehouse. According to the SAS, New Data Integration Landscape [6], a comprehensive universal data integration solution should successfully complete following different programs or business initiatives:

- **Data cleansing and enrichment:** A data integration solution should have the capability to cleanse and enrich the data in order to ensure the completeness and accuracy of data.
- **Data warehousing/marts (ETL):** Should provide the capability to build and maintain data warehouses/data marts via the ETL (extract, transform and load) process.
- **Cross system data consistency (data synchronisation):** Should reflect changes made between systems across the enterprise.
- **Data migration/consolidation:** Should provide the capability to migrate data from multiple existing systems to one or more new or existing systems.
- **Master Data Management (MDM):** MDM is the practice of creating a single "perceived" truth by mapping multiple disparate definitions of items, such as names of patients and medicines, which are held in various systems.

According to SAS Solution Overview [7], SAS for health care, the main challenges in integrating clinical data into a Clinical Data Warehouse are as follows:

Siloed departments: Different departments within health care organisations use different clinical and administrative systems that do not communicate with each other.

Data overload: Health care organisations have huge volumes of patient and organisational data, but no way to make sense of it all.

Manual processes: The manual processes often involved with accessing disparate data sources are time-consuming and error-prone, taking up valuable time that could be better spent doing actual analysis.

Compliance mandates: The adoption of the electronic medical record (EMR) has been a slow transition, with data volumes growing exponentially at a time when compliance, regulations and increasing types of media-rich files are creating added pressure on organisations.

Building a CDW Using SAS Enterprise Guide 4.3

In order to demonstrate the use of SAS Enterprise Guide to create a CDW, a hypothetical data scenario for diabetes treatment was constructed. Figure 1 shows how the different clinical data silos related to diabetes treatment exist in various departments within the healthcare system. Table structures for the data scenario were adopted using the tables available on BioGrid Australia [8]. The tables were populated with hypothetical data and using sample data files of PASW Statistics software.

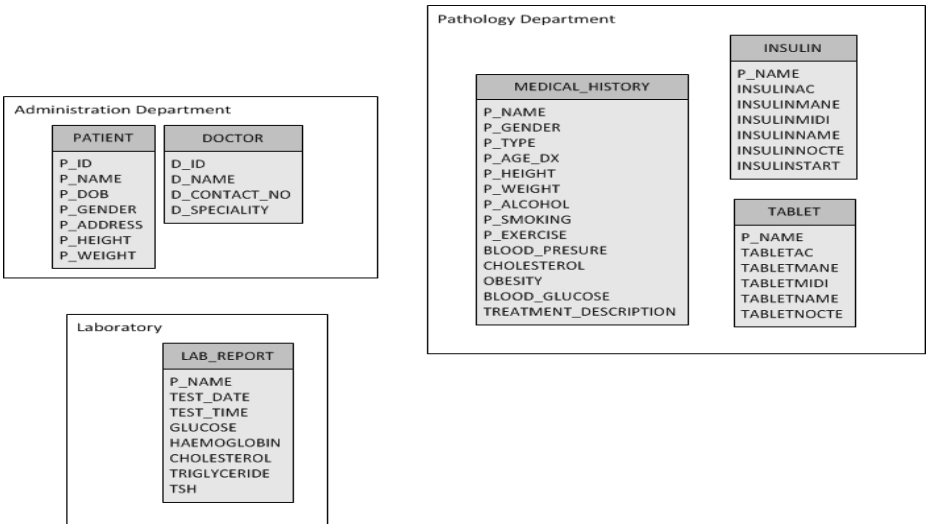

Figure 1. Clinical Data Silos Related to Diabetes Treatment (Source: BioGrid, 2010).

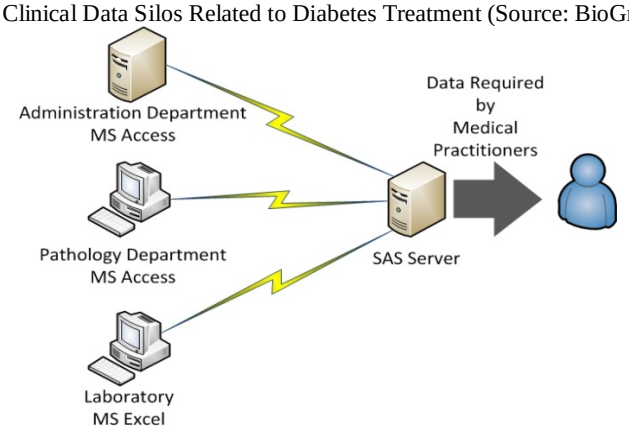

Figure 2. Integrating Fragmented Data Silos with SAS Server.

Each department database was implemented under separate Database Management System or a Spread Sheets application. Administration Department and Pathology Department data in separate Microsoft Access databases and Laboratory data in Microsoft Excel.

After experimenting with all the data warehousing architectures shown in Figure 3, we finally selected the distributed data warehouse architecture due to its high suitability to the chosen data scenario. The data from three data sources were extracted, transformed and loaded into SAS Server to create a distributed data warehouse structure.

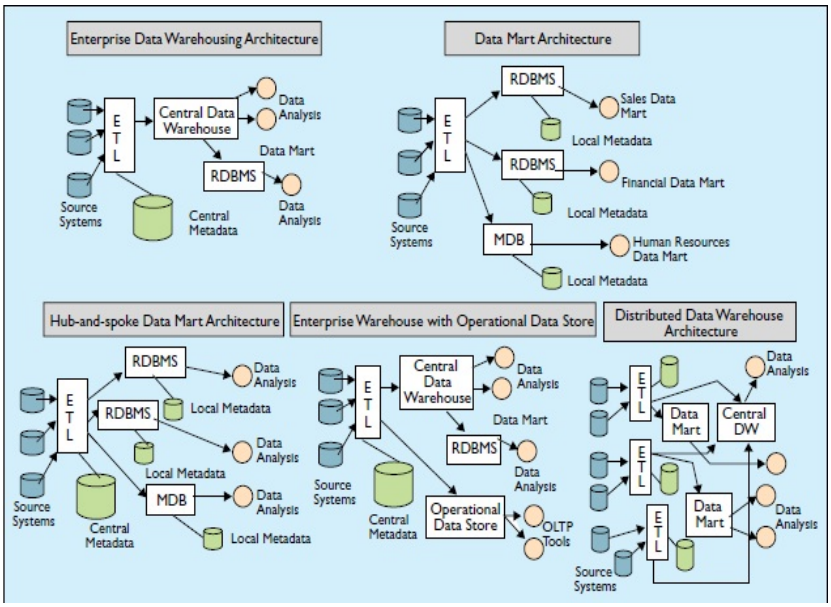

Figure 3. Different Types of DW Architectures (source: Sen & Sinha, 2005).

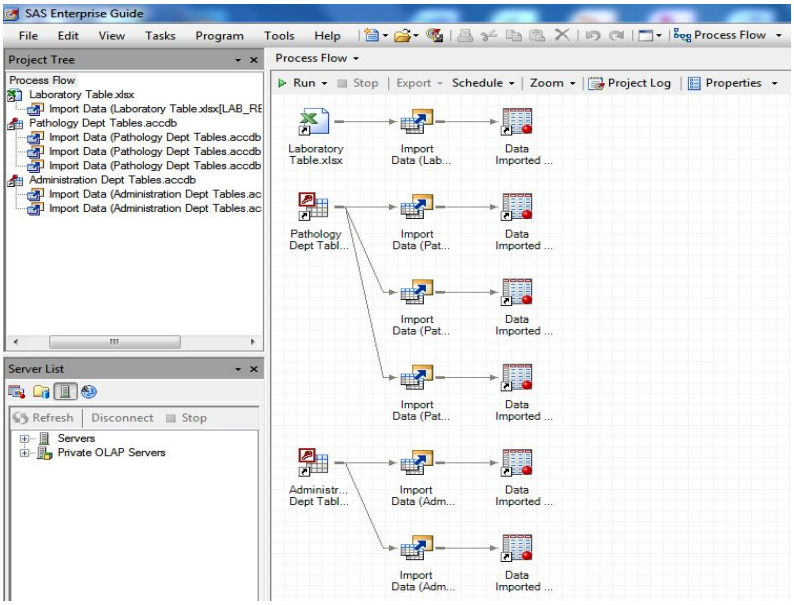

Figure 4. SAS Enterprise Guide Process Flow after ETL Process.

During the data cleansing and enrichment process additional data fields to establish relationships among tables and computed fields (e.g.: computed BMI using Weight and Height) were added. Figure 5 shows the logical ERD of the table structure in SAS server.

After constructing the CDW, SAS Enterprise Guide can be used to query data, apply data analytical functions, generate reports and share data over a network.

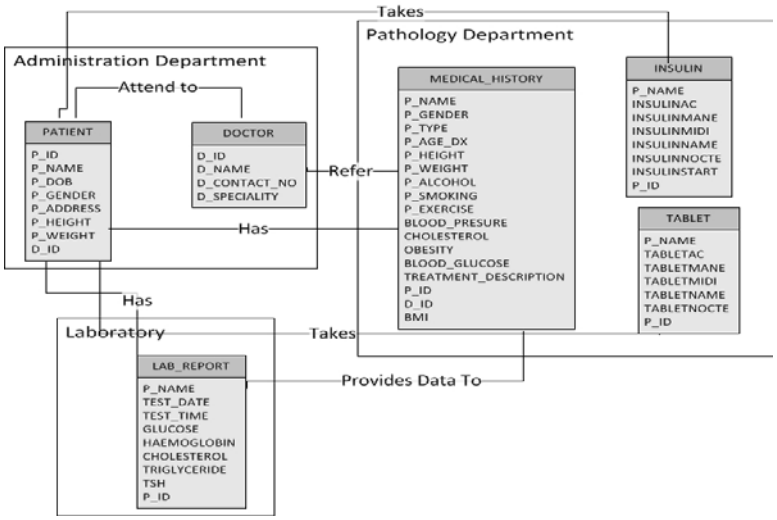

Figure 5. Logical ERD of the Diabetes Treatment Data Scenario.

Uses of SAS Clinical Data Integration when Building a CDW

SAS Clinical Data Integration mainly consists of SAS Clinical Data Integration 2.2 and SAS Clinical Data Standards Toolkit 1.3. SAS Clinical Data Integration is mainly used by clinical research professionals to improve efficiency, quality and speed in collecting, managing, analysing, reporting and assessing data from clinical trials.

SAS Clinical Data Integration implements CDISC (Clinical Data Interchange Standards Consortium) standards. This is helpful in data cleansing, enrichment process and to maintain the consistency of data. SAS can be used to gain both speed and efficiency by automating repeatable clinical data integration tasks and to deliver cleaner, more standard data for analysis.

Conclusion

Health care information is complex and must be accessed by health care practitioners, providers, researchers and patients. Some professionals involved in this processes are with minimal medical or information technology related training (or seldom no training). For all the health care information users, accessing medical data from fragmented, disparate and heterogeneous data silos is time consuming and expensive. Therefore medical data must be available in a centralised data warehouses equipped with proper tools and mechanisms to integrate data timely approach. As the amount of data and the number of systems involved increases rapidly, efficient and accurate data integration approach is required to create a single version of truth for all users.

References

Demetriades, J. E., Kolodner, R. M., & Christopherson, G. A. (2005). *Person Centered Health Records. Towards HealthePeople*. USA: Health Informatics Series, Springer.

Shepherd, M. (2007). *Challenges in Health Informatics*. IEEE. Proceedings of the 40th Hawaii International Conference on System Sciences.

Zheng, R., Jin, H., Zhang, Q., Liu, Y., & Chu, P. (2008). *Heterogeneous Medical Data Share and Integration on Grid*. IEEE. International Conference on BioMedical Engineering and Informatics.

Sahama, T. R., & Croll, P.R. (2007). *A Data Warehouse Architecture for Clinical Data Warehousing*. First Australasian Workshop on Health Knowledge Management and Discovery (pp. 227-232), Ballarat, Australia.

Berndt, D. J., Fisher, J.W., Hevner, A.R., & Studnicki, J. (2001). *Healthcare Data Warehousing and Quality Assurance*. IEEE.

SAS®: *The New Data Integration Landscape: Moving beyond ad-hoc ETL to an enterprise data integration strategy*, SAS Institute Inc., USA.

SAS®: *Solution Overview: SAS® for Health Care*, SAS Institute Inc., USA.

BioGrid Australia. Retrieved February 15, 2011, from http://www.biogrid.org.au/wps/portal

SAS Clinical Data Integration. Retrieved February 18, 2011, from http://www.sas.com/industry/pharma/cdi/index.html

SAS®: *Clinical Data Integration 2.2 User's Guide*, SAS Institute Inc., USA.

Sen, A. and Sinha, A. P. (2005): *A Comparison of Data warehousing Methodologies*, Communication of the ACM, 48(3), 79-84.

McDaniel, S., & Hemedinger, C. (2007). *SAS® For Dummies®*. Indianapolis, Indiana: Wiley Publishing, Inc.

Pedersen, T. B., & Jensen, C. S. (1998). *Research Issues in Clinical Data Warehousing*. Proceeding SSDBM '98 Proceedings of the 10th International Conference on Scientific and Statistical Database Management, IEEE Computer Society Washington, DC, USA.