Implementation Issues of Enterprise data Warehousing and Business Intelligence in the Healthcare Industry

Edward T. Chen

University of Massachusetts Lowell

Follow this and additional works at: http://scholarworks.lib.csusb.edu/ciima

Recommended Citation
Available at: http://scholarworks.lib.csusb.edu/ciima/vol12/iss2/3

This Article is brought to you for free and open access by CSUSB ScholarWorks. It has been accepted for inclusion in Communications of the IIMA by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.
Implementation Issues of Enterprise Data Warehousing and Business Intelligence in the Healthcare Industry

Edward T. Chen
University of Massachusetts Lowell, USA
edward_chen@uml.edu

ABSTRACT

The healthcare industry is following the lead of other industries and finding value in enterprise data warehousing (EDW) and business intelligence (BI) tools. Healthcare organizations are leveraging these tools to provide a plethora of benefits realized through enhanced business operations and performance. The EDW combines data from multiple source systems across an enterprise, and BI tools extract the data in meaningful ways to enable managers to make the best informed decisions. As with all management information systems, there are technical issues to be considered that impact the design, build, implementation, and support of the system. These benefits and challenges are explored, as well as special considerations necessary for the healthcare industry compared to other industries utilizing data warehousing and business intelligence. This paper investigates these critical issues and provides suggestions to harness the implementation of EDW and BI in the healthcare industry.

Keywords: Enterprise data warehouse, business intelligence, knowledge management, data mining, on-line analytical processing, HIPAA

INTRODUCTION

It seems common to many patients that have gone to different healthcare providers in the same organization and needed to give the same registration information such as name, date of birth, insurance information, and other demographics such as race, religion, marital status, and language. At times it can be frustrating why patients need to give all this same type of information again when they just gave it to the provider office next door. Beyond frustration, it may even seem irrelevant. Why does the provider need to know things such as race, religion, marital status, and language? All of this data is entered into the administrative system and is often fed to a central data mine, along with data from patient visits such as diagnosis and procedure codes that is fed to the same central data mine from a clinical system. As explored in this paper, data is mined at all levels for decision-making. This central data mine is referred to throughout this paper and corresponding literature as an Enterprise Data Warehouse (EDW), Information Warehouse (IW), or simply data warehouse.

The use of enterprise data warehousing and data mining through business intelligence tools is widespread and well established in industries outside of healthcare. March and Hevner (2007) describe the concept of data warehousing as the “integration of data from disparate sources into one coherent repository of information,” and business intelligence as “interpreting that data with...
respect to a business task (contextualization).” Just the compilation of data within the warehouse is a complex process and highly specialized within healthcare. The dimensions of this complexity is explored and compared with other industries.

Endless data is useless if there are no tools to form meaningful reporting for users. Expanding on the concept of contextualizing data to a business task, Mach and Salem (2010) describe BI as a way to use information to drive business activity and derive value. As healthcare organizations find more value in data warehousing and BI tools, this technology is experiencing significant growth also spurred by external factors such as governmental policy and deadlines. Along with rapid technology growth comes challenges and evolving design, build, and support issues.

WHY USE EDW AND BI IN HEALTHCARE

As technology advances and government regulations require increased application and “meaningful use” of healthcare data, management information systems (MIS) are being implemented within healthcare organizations at a very fast pace. As more MIS are built, more data is collected and available. A main benefit of data warehousing is the compilation of clinical and financial data from separate silo systems, and storing them together in compatible formats (Ferranti, Langman, Tanaka, McCall, & Ahmad, 2010). The EDW simply collects the data from various disparate systems and centrally stores it in a way that enables BI tools to easily extract it.

As data quickly builds, healthcare organizations can use this data to drive business strategy by turning it into meaningful information. Abidi (2001) refers to healthcare organizations as ‘data rich’ but ‘knowledge poor,’ and asserts that Knowledge Management (KM) through BI tools and Data Mining (DM) of the warehouse can lead to strategic decision-making and planning. The impact of data mining and use of BI tools is far reaching across many dimensions and departments within healthcare organizations. Kamal, et al. (2010) note that a diverse set of applications is impacted including business applications, clinical applications, research applications, and inter-organizational applications which allow collaboration with external entities. Not only is the data used to enhance internal operations, but it is also aggregated with external organizations to drive higher level change such as insurance standardization and public health policy (Kamal et al., 2010).

A great deal of research and analyses report a long list of benefits that healthcare organizations realize through the application of EDW and BI tools (Ferranti et al., 2010; Mach & Salem, 2010; Stodder, 2010). The benefits of data warehousing and business intelligence include the following:

- monitor quality and improve outcomes
- improve patient safety
- manage patients as they move through the organization
- gain insight into the effectiveness of patient treatments
- improve workflow efficiency
- develop best practices
- optimize insurance procedures
- uncover patterns of increased expenditures
• develop and monitor key indicators to improve performance and quality
• increase efficiency and accuracy of planning, budgeting, forecasting
• staff and equipment resource optimization
• enhance human resource management
• better manage supply chain and logistics
• cost containment
• manage financial performance
• streamline billing
• uncover revenue opportunities
• move toward quality healthcare objectives and pay-for-performance metrics

In short, data warehousing and business intelligence tools can be used to deliver a better product through improved patient care, monitor quality assurance, reduce costs and improve operational efficiencies, and improve financial performance utilizing benchmarked key performance indicators. This long list of benefits can only be realized through the design, build, and application of useful BI tools.

WHO USES EDW AND BI AND HOW

Referring to the long and diverse list of benefits that can be realized through the application of data warehousing and BI tools, it should be no surprise that the users who access these tools are just as diverse. Those who have an interest in these data range from data analysts and researchers up through the organizational hierarchy all the way to executive level decision-making. This diversity requires flexible tools that can be used by all roles. Additionally, this easy access to data brings both pros and cons. Lastly, the Information Systems Strategy Triangle is witnessing changes as data warehousing and business intelligence becomes more prevalent in healthcare organizations (Pearlson & Saunders, 2010). The implementation of an EDW will most likely impact the organizational structure by requiring additional resources to build and support the system as well as provide the BI analytics. These topics are further depicted as the details of the business intelligence tools that extract the data from the enterprise data warehouses are explored.

Business Intelligence Tools

Without adequate BI tools to extract the data and form meaningful reports, data warehouses sit underutilized and untapped. Common healthcare BI tools that have been copied and modified from other industries include reporting tools such as dashboards and scorecards. To reach the diverse set of end users, many healthcare organizations have implemented Online Analytical Processing (OLAP) tools to provide an easy yet sophisticated end user interface.

OLAP tools provide flexibility through multidimensional tables and data cubes. Tremblay, et al. note the capability of OLAP to provide data at multiple levels of details for different perspectives, allowing analysts to manipulate the data as desired and perform complex analysis (Tremblay, Fuller, Berndt, & Studnicki, 2007). Parmanto, Scotch, and Ahmad (2005) also focus on OLAP tools that allow the user to roll up to a higher summary level of data or drill down to more detailed transaction specific data. Users have the ability to access the data online whenever
they want, take bits and pieces, and slice and dice the data in a way that is most meaningful for their decision-making needs.

Business Intelligence tools take data a step further by providing meaning and transforming the data into information and knowledge that can be used to drive business strategy. To fully optimize warehouse data, Mettler and Vimarlund (2009) assert that BI tools need to be process-centric versus data-centric. By integrating process analysis with data analysis through BI tools, business operations can be further enhanced. This argument for integration of data and process analysis is strengthened by reviewing the list of benefits from section one that includes many process improvements.

**Users of Business Intelligence Tools**

The building of an enterprise data warehouse and the availability of all of its data can be a double-edged sword. There is no doubt that there are many benefits that come with EDW and BI. The data can be valid, the BI tool can be useful, but this is all negated if the end user is not skilled enough to adequately leverage these tools. As with all data, the misuse and misinterpretation are concerns to the system designers as well as the healthcare management.

There are three categories of data warehouse users as identified by Spil, Stegwee, and Teitink (2002). The different users of a data warehouse include those that just want basic standard reports, those who complete further analysis on the standard reports, and business analysts who use the data for analysis, trending, and forecasting. Those who just utilize or analyze standard reports are typically satisfied with static designed tools such as dashboards or scorecards that retrieve refreshed data. The data output changes as it pulls from the data warehouse, but the format, columns, headings, etc. remain the same. Those who perform more sophisticated analysis, often by leveraging OLAP, may change fields of data extraction each time they access the data warehouse depending on the business need. As the skill set becomes more advanced, a new breed of professionals is developing in business analytics.

**Impact on the Information Systems Strategy Triangle**

The Information Systems Strategy Triangle is a framework as depicted by Pearlson and Saunders (2010). The IS Strategy Triangle is formed by organizational strategy and information strategy supporting and driving overarching business strategy. A piece of this triangle cannot change without impacting the other two components. When evaluating the impact of the implementation of data warehousing and business intelligence systems, this triangle is changed in several ways. It is easy to see how EDW and BI support and drive business strategy through the list of benefits that the organization enjoys. But management needs to know how the organizational strategy is impacted by these systems (Pearlson & Saunders, 2010).

As previously noted, business analysts are emerging as a new breed of professionals. While business analysis has always existed, the role is becoming more specialized and requires more distinctive skills. Tremblay, et al. (2007) note that the role of those who in charge of extracting the data has evolved from just pulling data to interpreting and analyzing data, resulting in more elevated status of these employees. These business analysts are also playing a more active role in
the design, development, and implementation of tools such as OLAP. As this new tier or departmental group of business analysts develops, the organizational structure and strategy needs to adjust accordingly. Weighing function versus specialty, healthcare organizations need to decide if business analysts fall under operations departments, information technology units, or they may form their own consultative like departments.

Another impact on the organizational structure and strategy is the change required to the governance structure. Mettler and Vimarlund (2009) note the importance of BI governance and decision-making regarding system structure. Considering the dynamic nature of the healthcare industry compounded with the constant additional and changes to EDW, the build of a strong and decisive governance structure is imperative for success.

**TECHNICAL ISSUES EDW AND BI**

As with all information systems, there are managerial considerations that must be taken into account. Pearlson and Saunders (2010) identify potential technical issues of adaptability, scalability, standardization, maintainability, and security. Each of these technical issues is evaluated in terms of the design, build, implementation, and support of data warehousing and business intelligence tools within healthcare organizations. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the Recovery Act are also discussed since the use of technology has become so prevalent in the healthcare industry. Patients have a right to their healthcare information and how the information can be used to protect their privacy.

**Adaptability**

The architecture and build of a sustainable and flexible enterprise data warehouse up front is critical for healthcare organizations. New clinical and administrative systems are constantly being implemented and the EDW needs to be able to accommodate each new data source as they are added to the network. March and Hevner (2007) point out that the evolution of the EDW is critical to capture the results of the changes and to support on-going future cycles of improved management decision support systems. With the build of system networks constantly in flux, fully integrating and keeping up with all of the changes is a major task for the EDW. Wood (2000) points out some design decisions that can be considered to enhance the adaptability of a data warehouse; including decision of the primary key such as medical record number (MRN), which needs to account for “child” or “surrogate” keys; designing a flexible “top-down” or “bottom-up” approach to data flow; and the decision of how often to feed data from source systems to data warehouse. Building in as much flexibility up front will provide significant cost savings of resources through various system implementations.

Identification of a primary key is fairly standard in interface design such as the feed of a source system into the EDW. It seems natural that a healthcare organization would use a medical record number (MRN) as the primary key. However, since many organizations have multiple systems, each system may maintain its own record numbering wheel. This is where the “child” or “surrogate” keys may come into play, linking to the primary key. The EDW needs to be able to keep linking between the keys, and ensure they are easily referenced upstream and downstream.
Some organizations are solving this issue by implementing an Enterprise Master Patient Index (EMPI), which indexes all record numbers and the primary key becomes the EMPI number.

Designing data flow from “top-down” or “bottom-up” is a difficult decision that often requires analysis of the data to determine which direction will provide the highest level of data integrity. Some organizations identify one system as the overarching “source of truth” that feeds all downstream systems. Other organizations allow various systems to feed one another through bi-directional interface feeds. Sometimes this decision is quickly revisited once implemented as downstream systems realize critical data is wiped out or updated with outdated information by the EDW.

**Scalability**

Scalability becomes an issue regarding the decision of the frequency of data feed from source system to the data warehouse. This can usually be changed relatively easily as long as the format remains the same. Typically operations managers want data refreshed as often as possible, ideally in real-time. However, this can bring another host of issues due to timing of when data is updated. Depending on the means of data transmission, updating the EDW may entail batch file transfers, manual effort of loading, and/or the build of a complex real-time interface.

The main concern of increasing the frequency of data transmission is usually regarding storage or disk space and system performance. The addition or elimination of source systems requires the EDW to be able to respond to increased and decreased demands. Underutilization of resources can be as much of an issue as not enough resources such as storage space. As some healthcare organizations are implementing more comprehensive systems that fulfill the role of multiple systems such as administrative, scheduling, and clinical functions, the older legacy systems are phased out and there may be less need to integration and system space resources.

**Standardization**

Standardization is of utmost importance throughout all phases of the system life cycle for the EDW and BI tools. Spil, et al. (2002) stress the need for common definitions of terminology. The label of a field in one system can have the exact same label of a field in another system, but the data stored in those fields may hold completely different meanings. Standardization is needed to enhance interoperability, and it can be achieved through established governance. Chute, Beck, Fish, and Mohr (2010) argue that a well-designed and standardized data warehouse allows easy retrieval of information that enables transformation at a system level rather than at the specific process level (Chute et al., 2010). Ideally, one of the goals of implementing EDW and BI tools is to extend beyond the process level and drive business strategy transformation at the system level. Standardization of fields, communication protocols, definitions of fields, and record matching logic enables a smoother integration of system platforms and also ultimately lowers future resource costs that may have been spent in fixing errors or redesigning formats.
Maintainability and Data Errors

The ongoing support of an enterprise data warehouse and business intelligence tools needs to be manageable so IT resources can maintain quality systems. Standardizing and integrating all systems across departments should help create an environment that makes it easier to troubleshoot issues and maintain all integrated systems. Stodder (2010) notes that to leverage BI in clinical decision-making, healthcare organizations need to conquer the silo problem and improve integration and information sharing across groups. Information sharing can also enhance the skills of the users which can help the maintainability of the EDW. If end users are more skilled with the BI tools available to them, they may be less likely to request system modifications and customizations.

Information sharing across groups may also enhance the data integrity by ensuring accurate data entry, and thus the maintainability of the system. Rohloff (2011) notes that a BI system is only as good as the data it contains and the skill of the analysts using it. To realize the full benefits of a BI tool, a hospital must first optimize the entering, integration, and analysis of data. This again points back to the important growing field of business analysts in the healthcare industry, and how the organizational structure may need to be reconsidered.

Several researchers have specified the existence of various errors that occur with the integration of data warehouses. These errors have a direct impact on the maintainability of these systems. Berndt, Fisher, Hevner, and Studnicki, (2001) identify five different errors that can compromise data integrity in EDW: (1) Design or configuration errors, (2) Collection errors, (3) Staging errors, (4) Data integration errors, and (5) Query errors (Berndt et al., 2001). Kamal, et al. (2010) and Rohloff (2011) indicate that managing interface error logs help provide additional insight and recommendations for resolution of these errors.

Data or configuration errors occur when there is a field formatting issue such as a special character may have been entered into a field that only accepts numeric characters. These errors can be corrected by fixing the error in the source system. Although this seems easy, the process requires data quality feedback to loop back to the appropriate manager who supervises the person who entered the data incorrectly.

Collection errors occur when incorrect information is entered. These errors may be difficult to catch because there is often nothing that flags the incorrect information because it passes formatting requirements. However, once the error is discovered it should loop through the same feedback loop to be corrected in the source system.

Staging errors occur during loading process. They may be due to an error in the extraction query, such as incomplete queries or only partial extractions of the data. These errors can often be fixed by manipulating the incoming file or record, or by making changes to the loading process.

Data integration errors occur when information files into the wrong field. This may occur because the source system started using a field that was not previously used, or may have changed the use of a field. These errors may also occur due to data mapping mistakes and can usually easily be fixed through cross-mapping or field mapping.
The last category, query errors, occurs when the BI tool has pulled the wrong data. This is a broad category and can entail various scenarios such as user error in application of the BI tool or error in data selection. These errors are usually evident when the reporting tool produces output that does not match the expected results. They are often remedied by going back, tweaking the BI tool, and re-running the data extraction.

All of these categories of data errors highlight how easily something can go wrong with the data flowing in and out of the data warehouse. One other scenario that can compromise data integrity is the dangerous omission of data. End users often do not know what they do not know. Maintaining data integrity has a significant and direct impact on the maintainability of the EDW. Additionally, information sharing can help the maintainability when uninformed or unskilled users may think there is something wrong and request modifications to the system, when the issue may really be user error.

**Security**

The responsibility for healthcare professionals to keep patient information confidential dates all the way back to the Hippocratic Oath taken by doctors that was written in the 5th Century (Murray, Calhoun, & Philipsen, 2011). The Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the Recovery Act reinforce that and cover more modern day situations since the use of technology has become so prevalent in our lives. The Recovery Act says that patients have a right to their healthcare information, and that information is to be protected.

HIPAA sets rules and guidelines that must be followed when it comes to what and how healthcare information can be used. The physical privacy and security of housing health information and the transmittal of information is protected and enforced by Title II under HIPAA (Murray et al., 2011). Title II encourages the use of EMRs while enforcing the need for protection from fraud and abuse. HIPAA also protects information on adolescents, which is a complex matter in healthcare even without the use of technology. One of the ethical questions posed on using EMR is who is going to have access to patient information? How will data sharing through EMR affect the patient? How is the security of patient information being handled? As we have seen in our society today, despite having the best firewall, passwords, or encryptions hackers are still getting into networks and wreaking havoc for people for their own personal use (Kopala & Mitchell, 2011).

Healthcare data carries a heightened sense of privacy and security issues. In addition to basic network access security checks, some EDW systems define security and access by roles and rights. This helps ensure that users only have access to parts of the data warehouse that they need to perform their jobs.

**EDW AND BI IN THE HEALTHCARE INDUSTRY**

Adoption of enterprise data warehouse and business intelligence technology within the healthcare industry lags well behind other industries. This section explores the unique challenges
and factors that have contributed to this slower adoption rate within healthcare, as well as the prospects for the future including document management technology.

Complexity of Healthcare Data

Multiple sources cite the complexity and heterogeneity of healthcare data as the primary reason for the slower adoption rate of EDW and BI. Chute, et al. (2010) blame the slower adoption rate of the data warehouse by the healthcare industry on the complexity and heterogeneity of biomedical, operational, and clinical data. Parmanto, et al. (2005) point out that since healthcare data is primarily encounter-based, the design and build of a healthcare data warehouse must consider patient data as a value circle rather than linear model in other industries where a product moves through a line. This complexity is addressed in the multi-dimensional OLAP BI tools.

In addition to the complexity of the actual data, the decentralized multiple source systems add another layer of complexity to healthcare data. They add complexity through incompatible platforms, varying definitions, and standards. Connell and Young (2007) note that the heterogeneity of healthcare systems and the complex network of internal and external source systems require more person-to-person knowledge transfer than other industries. This requirement of more person-to-person knowledge transfer involves a slower implementation cycle than other industries, which may dampen enthusiasm as systems are deployed.

Madsen (2011) comments on the differentiation of the healthcare industry by citing additional factors including: (1) Regulation and compliance factors, (2) Risk that entails a patient’s well-being, (3) Complex relationships and interactions between patients, insurance, providers, departments, etc., (4) Inclusion of non-standard, subjective, contextual data, and (5) Differing perspectives of data may produce different definitions for the same thing, highlighting the importance of standardization. Berndt, et al. (2001) further expand on the complexity of the healthcare industry by pointing out that not only does data come in from multiple sources; it also goes out to diverse set of stakeholders. To further compound the complexity healthcare data is the sensitivity of healthcare data carries security and privacy concerns. All of these factors contribute to create an environment with multiple layers of complexity in the healthcare data world.

Future Prospects

Nevertheless, to remain current, relevant, and competitive; healthcare organizations must embrace the EDW and BI technology to fully realize and implement intelligent IS, organizational, and business strategy. Mettler and Vimarlund (2009) assert that to move forward, the main issues that must be addressed are policy issues, emerging trends in the field of BI, as well as applications that are currently overlooked but could be applied to healthcare analytics. Policy forces driving implementations do not allow unlimited time for design and development which may bring another set of issues. Sometimes costly concessions are made that jeopardize further adoption of the technology. Emerging trends and other analytical applications can be evaluated, but the slow adoption rate compounds the difficulty of assessing these alternative solutions. Additionally, providers are seen as the “revenue producers,” but organizations will need to evaluate the value added by EDW and BI and determine what technologies to implement.
Within healthcare IT, all resources are facing fierce competition and are limited. Resources include funding, operational support, implementation and integration specialists, and IT employees.

**Document Management**

One area that lacks extensive research and experience is document warehousing. Tseng and Chou (2006) point out that most information warehouses focus on numerical and easily indexed data even though a lot of important information is stored in textual based documents. Healthcare vendors are starting to tap into this area, such as GE Healthcare Centricity Business’ Enterprise Document Management (EDM) module. This module is built to integrate with other modules in the Centricity Business suite of applications. The challenge for document warehousing will be the transmission, indexing, and storing of these documents. Healthcare still depends a great deal on textual based documents such as emails and word documents, so development and integration of document management will need to advance to provide managers with truly comprehensive knowledge.

**CASES OF EDW AND BI**

There are several case studies that describe the outcomes after implementation of an enterprise data warehouse and business intelligence tools within healthcare organizations. The three case studies cited here include the Mayo Clinic, Duke University Health System (DUHS), and Ohio State University Medical Center (OSUMC). The implementation of EDW and BI technology at the Mayo Clinic brought success to cancer center projects, infection analytics project, as well as referral analysis, balanced scorecard reporting, quality dashboards, and ad hoc reports (Chute et al., 2010).

DUHS also cited success with the technology to support patient safety, financial effectiveness, and public health issues through BI tools such as dashboards and reporting that allowed real-time access (Ferranti et al., 2010). OSUMC’s system was originally designed to improve efficiency of patient care delivery and financial operations, but was found to provide useful data for research efforts (Kamal, et al., 2010). These cases illustrate how far reaching the benefits of EDW and BI technology can penetrate across multiple facets of a healthcare organization.

**CONCLUSION**

Even though there is a great deal of evidence that enterprise data warehousing and business intelligence tools can help healthcare organizations support business strategy, securing scarce resources to implement these systems remains a challenge. Many factors contribute to the slower adoption rate of these technologies in the healthcare industry, but the most influential is the complex nature of the data – what it is, where it comes from, how it is stored and passed, and where and to whom it goes. There are many technical issues that a manager must consider when implementing EDW and BI, but an adaptable and scalable system will support most future changes. Finally, how the implementation of these systems impacts the organizational structure...
must be addressed. The roles of business analysts and the structure of governance to determine standards are essential for a successful implementation and adoption of EDW and BI tools.

The influence that technology has in the healthcare field is undeniable. The involvement of the federal government such as HIPAA compliance has made it mandatory to adopt new systems and change the business processes in healthcare agencies. The time, money, and work that it is going to take to unify our healthcare system to enhance communication, reduce errors, and improve decision-making will prove its worth in time.

REFERENCES


