Data Warehousing
A New Focus in Healthcare Data Management
July 2009

Enterprise Information Systems Steering Committee
Data Warehousing Work Group

Introduction

In 2005, HIMSS Analytics™ developed the eight-stage EMR Adoption Model℠. The model reflects the results of more than 5,000 institutions surveyed on their level of clinical system implementations and progress toward a paperless environment. So far, only 15 hospitals have achieved the highest level, Stage 7 (Appendix 1). One of the criteria for this stage is the use of a data warehouse, which requires a sizeable commitment from an organization's financial and human resources. Today, healthcare institutions across the country are assessing their warehouses in order to address the evolving criteria of meaningful use broadly outlined in the American Recovery and Reinvestment Act of 2009 (ARRA).

This committee undertook a review of organizations with successful warehousing projects in order to develop a model for data warehouse maturity that would reflect the current state of data warehousing in healthcare. This paper presents the results of the case studies and the maturity model for data warehousing in healthcare.

Data Warehousing Overview

A data warehouse is a "subject-oriented, integrated, time-varying, non-volatile collection of data in support of the management's decision-making process."[1] It is a centralized repository that stores data from multiple sources and transforms them into a common, multi-dimensional data model for efficient querying and analysis. This definition of the data warehouse focuses on data storage. However, the means to retrieve and analyze data; to extract, transform and load data' and to manage the data dictionary also are considered essential components of data warehousing. Many references to data warehousing use this broader context. Thus, an expanded definition for data warehousing includes business intelligence tools (dashboard, scorecard, reporting, query, OLAP, alert, etc.); tools to extract, transform and load (ETL) data into the repository; and tools to manage and retrieve metadata.

The key components of the data warehouse include: data access of the operational systems, data staging area, data presentation area and data access tools. Each component serves unique function(s) to support the data warehouse.

- Data access of the operational systems: the source systems capture and process the organization's transactional data. Each operational system typically works independently and does share common data from other operational systems. The data acquired through these systems is uploaded to the staging area.
- Data staging area: the data staging area is both the storage area for the uploaded data and the platform for the ETL process. The ETL processes the raw

data into the data warehouse environment. First, the data is extracted from different sources (operational systems, flat files, manual input, etc.). Then the data is cleansed, formatted and calculated into a standard format and structure. Once the data is standardized, it is loaded into the presentation area.

- Data presentation area: a data presentation area is considered a set of integrated data marts. A data mart is a subset of the data warehouse to focus on a specific department (finance, clinical, etc.) or functional area. The data in a specific data mart is relevant to that specific area. The data contained in the data presentation area must be detailed and logically organized.

- Data access tools: once the data is ready in the data presentation area business users can use various tools to access the data for reporting and analysis. The typical tools include: dashboards, scorecards, reports, ad-hoc queries, OLAP (cube) tools and more.

Over the years, the data warehouse has evolved into a more mature technology and business process. Increasingly healthcare organizations have used the data warehouse as the platform to improve the decision-making for clinical, financial and operational purposes.

Morain and Norris[2] found the following common advantages of companies that implemented a data warehouse:

- One consistent data store for reporting, forecasting and analysis.
- Easier and timely access to data.
- Improved end-user productivity.
- Improved IS productivity.
- Reduced costs.
- Scalability.
- Flexibility.
- Reliability.
- Competitive advantage.

Case Study

The goal of the case studies was to identify and categorize common elements and success factors among organizations with successful use of data warehouses. A 19-item questionnaire was developed (Appendix 2). HIMSS members with demonstrated successful data warehousing projects were contacted**.** Ten organizations responded to the questionnaire and participated in telephone interviews. The results are presented below.

The organizations surveyed include:

1. Geisinger Health System, Danville, PA.
2. Kaiser Permanente, Sacramento, CA.
3. King Faisal Specialist Hospital and Research Center, Riyadh, Kingdom of Saudi Arabia.
4. Partners HealthCare, Boston.
5. Peace Health, Oregon.
6. Providence Health & Services, Northwest US (two facilities).
7. MD Anderson, Houston, TX.
8. Children's Hospital and Medical Center, Omaha, NE
9. Brigham and Women's Hospital, Boston.
10. University of Pittsburgh Medical Center (UPMC) Hospital and Community Services, Pittsburgh, PA.

Key Findings

Participants cited strong executive leadership as the most important factor for successful data warehousing.

Driving Force

- Executive leadership was most commonly provided by the Chief Medical Officer.
- Most cited a desire to better manage data, reporting that data from separate sources were difficult to coordinate.
- Most are implementing information-based clinical, financial or operational decision support.

User Community

Warehouses were used collate data from disparate systems. Users at different levels in the organization had needs for different data.

Business Need

Each organization had their own unique business need that started their efforts, but a common goal was consolidated data.

Standards

Surprisingly, none of the organizations profiled had a defined enterprise data standard. Most organizations relied on the source system to define the standard and implement quality measures.

Development

- Eight organizations purchased or contracted the original data warehouse. Only two respondents reported doing the development internally.
- There was no clear vendor of choice. Most used vendors other than their clinical, financial or operational systems vendors.
- Nearly half reported that they do not have a data model. The rest only had a data model for their data marts, but not the source systems.

Tools

- All reported using one of the major business intelligence (BI) vendors' tools—Brio/Hyperion, Business Objects, Cognos, Crystal Reports, and SAS—but there was no clear leader. Microsoft tools were surprisingly absent, except for Excel and MS Access.
- The question of database was not specifically asked, but several volunteered the RDBMS they were using. All the usual vendors were mentioned: DB2, Oracle and Sybase.

Content

All but one respondent reported that their warehouse contained both clinical and financial data. Each organization indicated they started with one kind of data then added the other.

Core Measures

Most respondents reported they use other tools to collect or report data for core measures.

Pay for Performance

Almost half of the respondents reported using data warehouses as a source of information for providing feedback to their physicians.

Meaningful Use for ARRA

Respondents are hoping for clarity in the meaningful use definition in order to utilize their warehouse for incentive payments and/or reimbursement.

Health Information Exchange (HIE)/Regional Health Information Organization (RHIO)

Only one respondent reported working with an external HIE, and that was a statewide HIE.

Return on Investment (ROI)

The general consensus is that it is hard to quantify a direct ROI for the data warehouse, but several cited improvements in quality, risk reduction or identification of opportunities for either cost-saving or clinical care improvement. Two respondents reported improvement in efficiency based on observations that they are able to do more with staff.

EMR Adoption Levels

Only two respondents reported an improvement in their perceived adoption level after data warehouse implementation.

The Data Warehouse Institute Data Warehouse Maturity Level

The majority of the respondents deemed their data warehouse implementation efforts at the Adult Stage. The rest considered themselves Teenagers or moving to Teenager. (See below.)

The Data Warehouse Maturity Model

The results of the questionnaire confirmed that the maturity model developed by The Data Warehousing Institute was relevant to the development of data warehouses in healthcare. The model consists of six stages: Prenatal, Infant, Child, Teenager, Adult and Sage. (Figs. 1 & 2) Organizations evolve at different rates through these six stages and may exhibit characteristics of multiple stages at a given time. Business value increases with each successive stage. There also are two major obstacles: the Gulf, which occurs between the Infant and Child stages; and the Chasm, which occurs between the Teenager and Adult stages. The primary way to overcome these obstacles is to change executive perceptions. To cross the Gulf, executives must recognize that DW/BI is more than a management reporting system and that the spreadsheets and desktop databases they rely on to run the business are actually undermining productivity and effectiveness. To cross the Chasm, executives must perceive the DW/BI environment as a mission-critical enterprise they own, direct and fund.

Fig. 1: Data Warehouse Maturity Model[3]

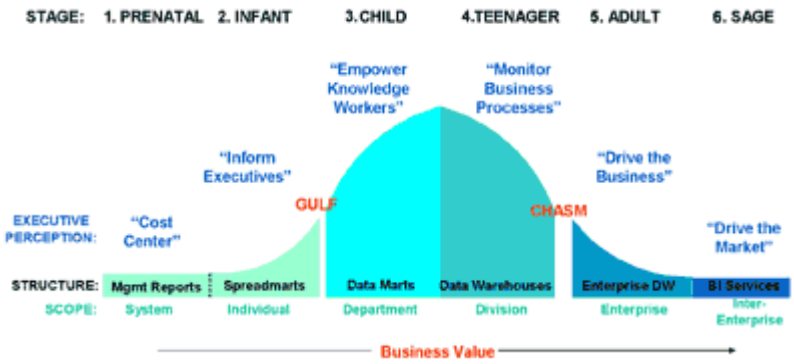
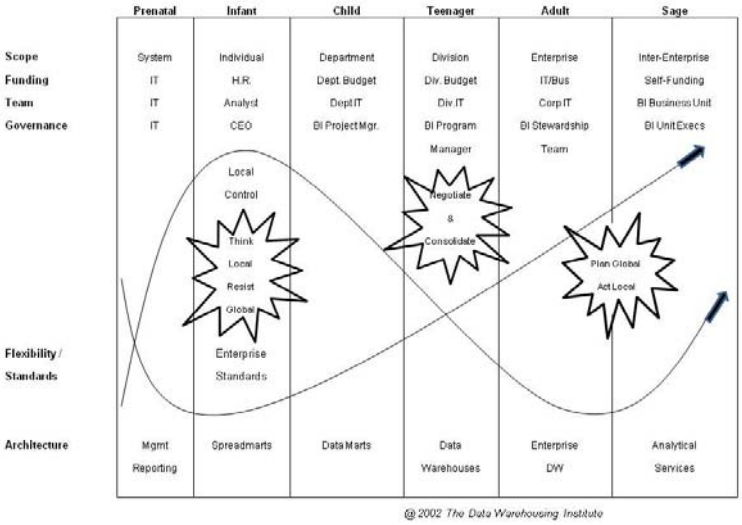


Fig. 2: Data Warehouse Maturity Model[4]



Prenatal Stage: Management Reporting

Once a healthcare organization has achieved Stage 1 in the EHR model, they have accumulated a portfolio of departmental applications, each with a set of standard reports particular to that system, i.e., laboratory, pharmacy, radiology, etc. Typically these reports are printed and distributed weekly, monthly or quarterly. Typically the reports are hand-coded against one specific system, and it is difficult for the IT department to keep up with the volume of requests for new and custom reports, which frustrates departmental users who need the data to do their jobs. Departmental users with IT skills begin to extract data from source systems and load it into spreadsheets or desktop databases so that they can use the data for management reporting. The participants in the case study also cited report request volume as a driving force for warehouse development.

Infant Stage: Spread marts

Because spreadsheets are cheap and easy to use, they spread quickly. Because they are developed by users working independently, many times the rules and data do not align with other spread marts, management reports or analytical systems. It is difficult or impossible for managers to get a clear and concise picture of the status of operations. Eliminating spread marts is difficult because they offer high local control at extremely

low costs. The first attempt to balance the need for local control with the desire for consistency is to move from spread marts to data marts.

Child Stage: Data Marts
For the purposes of this model, a data mart is defined as a shared, analytic structure that supports a single area, business process or department. A data mart is designed to meet the needs of knowledge workers in the department. Using interactive reporting tools, they are able to drill down or across a dimensional structure to follow trends to gain insight into the process or task they manage. Data marts exhibit many of the same fundamental weaknesses as the spread marts. Each data mart supports unique definitions and rules and extracts data directly from source systems. They may be adequate to support departmental needs, but data can't be aggregated to support cross-departmental analysis, such as the desire to track antibiotic administration times stored in the nursing system with surgery start times recorded in the surgery system. Once an organization has entered Stage 3 in the EMR Adoption Model, some sort of data mart will be needed in order to progress toward the goal of providing clinical decision support, as outlined in Stage 4.

Teenager Stage: Data Warehouses
In this stage, the organization realizes the need to standardize definitions, rules and dimensions to facilitate reporting across departments. Standardizing data marts can be done in a centralized or decentralized fashion. The most commonly encountered in our research was a hub-and-spoke strategy—a central data warehouse with logical, dependent data marts running in the same database as the data warehouse. This configuration facilitates deeper analysis across departmental boundaries. Many institutions also deploy a dashboard application to front-end the warehouse. Dashboards support alerting and drill-down paths to detailed reports. Data is available to users at all levels, not just the technical power users. The warehouse and business intelligence tools empower users with more information for fact-based decision making.

The results of the case study indicated that many organizations are still working toward developing data standards and data marts. Although many survey participants indicated that they believed that they were in the Adult Stage, the lack of data standards and implementation of data marts would indicate a Teenager Stage at best.

Adult Stage: The Enterprise Data Warehouse
In the penultimate stage, executives view data as a mission-critical asset supporting the business. The enterprise data warehouse (EDW) functions as a consolidation point for data from other structures. External data is integrated into the warehouse through the use of BI tools. A strong stewardship model is established so that business people own their data and guide the development of the EDW. At this point, scorecards are often used as a way to tie business results to departments.

Institutions with multiple facilities or that have merged or consolidated could be faced with the prospect of multiple data warehouses. In Stage 7 of the EMR Adoption Model, as organizations look to consolidated reporting of data from inpatient and ambulatory

environments, the potential for overlapping and inconsistent data requires that the organization establish a single source of truth and develop a true enterprise data warehouse.

Sage: Business Intelligence Services

As a data warehouse enters the Sage Stage, its value increases exponentially as its visibility declines. The warehouse becomes a piece of critical infrastructure. Institutions will be able to capitalize on their investments in statistical analysis and modeling as decision engines are embedded in the applications used by front-line staff. The expansion of clinical decision support into areas such as molecular diagnostics and genomics will be supported by the data warehouse in this final stage.

Benefits of Data Warehousing

The HIMSS survey revealed several common areas where organizations benefited from data warehouse implementation. They include:

1. Allowing end-users to self-service their data needs, freeing IT from being the sole provider of reports and data analysis.
2. Aggregation of data from clinical and financial systems into one repository.
3. Ability to provide standardized reports for core measures, internal quality improvements, pay-for-performance reporting and enterprise performance indicators.
4. Quicker identification of causal relationships of diseases.
5. Enterprise focus on common goals and objectives regarding data management.

Few of the organizations surveyed had documented return on investment (ROI) for their data warehouse initiative; however, all organizations indicated there were many benefits achieved by the implementation. This is consistent with findings documented in studies of other industries. Watson et al found that assessing ROI is difficult because the warehouse is "not usually liked directly to results. It is an investment in decision support architecture."[5] Although the warehouse is a required enabler, the benefits realized depend on the underlying business strategy and how well the warehouse is used to implement the strategy. Rather than a formal ROI, many warehouse projects are evaluated based on how well the implementation achieved established project objectives.[5] For healthcare organizations, RIO in clinical information systems is not achieved until these systems are used in support of clinical improvement. [6]

Organizations that benefited most from data warehousing were those that fell at or above the Adult stage of The Data Warehouse Institute Model, and were at or above Stage 4 of the EMR Adoption Model. Those organizations at lower stages of The Data Warehouse Institute Maturity Model and EMR Adoption Model did not have as many documented benefits or extensive implementation throughout the organization.

Bergeron notes that the common benefits of a properly implemented data warehouse are transparency to the underlying data, adherence to standards, adaptability and reduction of impact of reporting directly against production transactional systems.[7]

How upcoming legislation and ARRA will influence data warehousing efforts

One of the many topics that have been discussed extensively since the passage of the American Recovery and Reinvestment Act (ARRA) is the definition of meaningful use and the required objectives and measures set by the Office of the National Coordinator for 2011, 2013 and 2015.

The definitions of meaningful use proposed by the Health Information and Management System Society (HIMSS), the American Health Information Management Association (AHIMA) and the Markle Foundation share a high degree of consistency. The definitions are seen as a way to move the health IT industry forward in a manner that supports the requirements of ARRA and the needs of the medical community by addressing the key goals of improving quality, safety and efficiency and reducing health disparities; engaging patients and their families; improving care coordination; improving population and public health; ensuring adequate privacy and security protections for personal health information; and adopting advance-care measures while measuring and improving outcomes. The definitions seek an incremental approach that allows for ease of reporting with existing resources, while paving the way for better care and reporting in the future.

In cases where data warehouses contain significant clinical data and associated financial information, it is likely that the warehouse will allow the organization to address most of the meaningful use reporting required by ARRA. Investments that organizations have made in data warehouses should be able to be leveraged in many cases to help prove meaningful use. An organization may need to make some modifications to their data model to include other information needed for specific ARRA reporting, but if the basic data is available they should, if necessary, be able to piece the reports together using a manual process for a short time, leveraging the data and systems they already have.

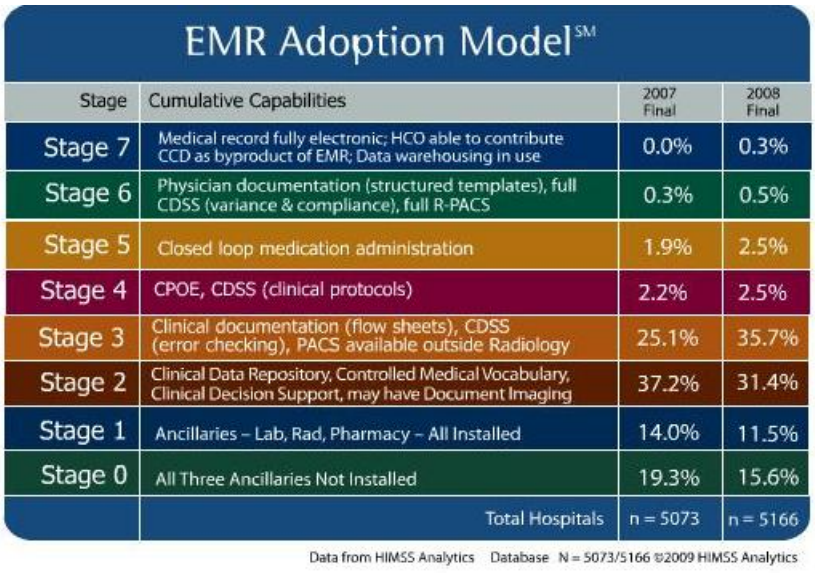
Below is a list of the key reporting requirements for ARRA 2011 measures. Although ARRA does not specify how these measures are to be produced, an organization with a data warehouse will be better positioned to provide the required measurements than organizations without a data warehouse. Without a data warehouse, an organization would need to rely on combination of options, including the comprehensiveness of its EHR, capabilities of in-house or contracted reporting staff, and/or the ability to purchase a system or systems to provide the needed data.

2011 ARRA Measures

- % diabetics with A1c under control [OP].
- % hypertensive patients with BP under control [OP].
- % of patients with LDL under control [OP].

* % of smokers offered smoking cessation counseling [OP, IP].
* % of patients with recorded BMI [OP].
* % eligible surgical patients who received VTE prophylaxis [IP].
* % of orders entered directly by physicians through CPOE Use of high-risk medications in the elderly [OP, IP].
* % of patients over 50 with annual colorectal cancer screenings [OP].

Appendix 1: HIMSS Analytics™ EMR Adoption Model℠



EMR Adoption Model℠

Stage	Cumulative Capabilities	2007 Final	2008 Final
Stage 7	Medical record fully electronic; HCO able to contribute CCD as byproduct of EMR; Data warehousing in use	0.0%	0.3%
Stage 6	Physician documentation (structured templates), full CDSS (variance & compliance), full R-PACS	0.3%	0.5%
Stage 5	Closed loop medication administration	1.9%	2.5%
Stage 4	CPOE, CDSS (clinical protocols)	2.2%	2.5%
Stage 3	Clinical documentation (flow sheets), CDSS (error checking), PACS available outside Radiology	25.1%	35.7%
Stage 2	Clinical Data Repository, Controlled Medical Vocabulary, Clinical Decision Support, may have Document Imaging	37.2%	31.4%
Stage 1	Ancillaries – Lab, Rad, Pharmacy – All Installed	14.0%	11.5%
Stage 0	All Three Ancillaries Not Installed	19.3%	15.6%
	Total Hospitals	n = 5073	n = 5166

Data from HIMSS Analytics Database N = 5073/5166 ©2009 HIMSS Analytics

In 2005, HIMSS Analytics developed the eight-stage EMR Adoption Model. The model reflects the results of more than 5,000 institutions surveyed about their level of clinical system implementation and progress toward a paperless EMR. All application capabilities within each stage must be operational before that stage can be achieved. All lower stages must have been achieved before a higher level will be considered achieved. A hospital can achieve Stages 3-6 if it has met all of the application requirements for a single patient care service (e.g., single nursing floor, cardiology service). Using the rules above, additional points are given for the implementation of applications in stages higher than the one fully achieved by the healthcare organization. In this fashion, implementation paths other than those prescribed by the stages can be taken into consideration for correlation with quality and financial research.

Stage 0: Some clinical automation may be present, but all three of the major ancillary systems—laboratory, pharmacy and radiology—are not implemented.

Stage 1: All three of the major ancillary clinical systems—laboratory, pharmacy and radiology—are installed.

Stage 2: Major ancillary clinical systems feed data to a clinical data repository (CDR), which provides physicians access for retrieving and reviewing results. The CDR contains a controlled medical vocabulary, and the clinical decision support/rules engine for rudimentary conflict checking. Information from document imaging systems may be linked to the CDR at this stage.

Stage 3: Clinical documentation (e.g., vital signs, flow sheets) is required; nursing notes, care plan charting and/or the electronic medication administration record (eMAR) system are scored with extra points, and are implemented and integrated with the CDR

for at least one service in the hospital. The first level of clinical decision support is implemented to conduct error-checking with order entry (i.e., drug/drug, drug/food, drug/lab conflict checking normally found in the pharmacy). Some level of medical image access from picture archive and communication systems (PACS) is available for access by physicians via the organizations intranet or other secure networks outside the radiology department.

Stage 4: Computerized provider order entry (CPOE) for use by clinicians is added to the nursing and CDR environment along with the second level of clinical decision support capabilities related to evidence-based medicine protocols. If one patient service area has implemented CPOE and completed the previous stages, then this stage has been achieved.

Stage 5: The closed loop medication administration environment is fully implemented in at least one patient care service area. The eMAR and bar coding or other auto identification technology, such as radio frequency identification (RFID), are implemented and integrated with CPOE and pharmacy to maximize point-of-care patient safety processes for medication administration.

Stage 6: Full physician documentation/charting (structured templates) is implemented for at least one patient care service area. Level three of clinical decision support provides guidance for all clinician activities related to protocols and outcomes in the form of variance and compliance alerts. A full complement of radiology PACS systems provides medical images to physicians via an intranet and displaces all film-based images.

Stage 7: The hospital has a paperless EMR environment. Clinical information can be readily shared via electronic transactions or exchange of electronic records with all entities within a regional health network (i.e., other hospitals, ambulatory clinics, sub-acute environments, employers, payors and patients). This stage allows the healthcare organization (HCO) to support the true electronic health record as envisioned in the ideal model.

Appendix 2 - *Questionnaire*

HIMSS Data Warehousing Questionnaire

The following questions are being presented to assist HIMSS in understanding the use of data warehouses at healthcare institutions. Answers provided to these questions will help assist HIMSS in developing utilities and resources to assist member institutions in the planning and development of data warehouses.

Category	Question	Response
TDWI Maturity Level	Based on the model, what stage is most representative of your current data warehouse effort?	Which level and why?
Driving Force	What was the primary force driving the development of a data warehouse at your organization?	Explain.
Driving Force	Was it a grassroots effort or executive mandate?	Explain.
User Community	Who are the primary users?	List job functions and how it is used
User Community	Who else interacts directly with the data warehouse?	List job functions and how it is used differently than primary users above.
Business Need	What business needs does it fulfill?	List needs.
Development	Was your data warehouse developed internally or purchased from a vendor? What vendor?	Vendor name and specific products. When purchased and installed.
Development	Do you have an enterprise data model?	
Standards	What level of data standards do you have?	
Tools	What Business Intelligence tools do you use?	List
Driving Force	Was data warehousing project driven primarily by IT, or was it part of a larger enterprise effort?	Describe
Content	Does the warehouse contain clinical as well as financial information?	List specific clinical info – ie orders, results, vital signs, nursing assessment, etc.
Core Measures	Do you use the data warehouse to submit core measure, quality reporting, state and regulatory reporting, or other registries?	Who is their core measure vendor? What registries?

Pay for Performance	How do you anticipate using the warehouse in pay for performance?	
ARRA	How do you anticipate that the warehouse will be used in demonstration of “Meaningful Use” as per ARRA?	
HIE / RHIO	Does the warehouse feed a larger HIE or RHIO initiative?	Describe interactions with RHIO.
ROI	What ROI have you achieved?	Did they have to prove ROI to get approval for funding the data warehouse project.
EMR Adoption Stage Before	Based on the HIMSS EMR Adoption Model, where would you rank your organization at levels 0-7.	
EMR Adoption Stage After	What level of the EMR adoption was your organization at when you initially implemented your data warehouse?	

Conclusion

This report discussed the concept and the benefits of data warehouses and the key findings from the survey. We used The Data Warehouse Institute's Data Warehouse Maturity Model as the guideline to evaluate the nature and the scope of the various data warehouse projects in the survey.

Data warehouse technology has evolved from the department level data marts to more enterprise-level platforms. Although the data warehouse is still in infancy for most of the healthcare organizations, we anticipate that most will accelerate the rate of the adoption of enterprise data warehouses due to ARRA requirements and other legislation.

Major enhancements of the data warehouse for healthcare organizations may include the following:

- Integration of more diverse source systems like EMR, financial, and other clinical systems, etc.
- Integration of external data sources (e.g., clinical and financial bench-mark data)
- Extended reporting capabilities like Core Measure and other mandatory reporting packages.
- More sophisticated and dynamic analytical capabilities to accommodate the ever changing market conditions and compliance requirements.

References:

1. Inmon WH. *Building the Data Warehouse*. 4th ed. New York City, New York: John Wiley & Sons, Inc.; 2005.
2. Morain SK, Norris DM. A Study of Perceived Post-Implementation Benefits of Data Warehousing the Data Warehouse Institute. *Bus Int J*. 2001;6(1).
3. Eckerson W. Gauge your data warehouse maturity. *Inform Manage Mag.* November 2004
4. TWDI. Available at: www.tdwi.org/publications/display.aspx?ID=7199.
5. Watson H, Thomas D, Preston D, Chen D, Abraham D. Data Warehousing ROI: Justifying and Assessing a Data Warehouse. http://www. bi-bestpractices. com/view/5780) 4780
6. Baker GR, MacIntosh-Murray C, Porcellato L, Dionne K, Stelmacovich, Born K. *High Performing Healthcare Systems: Delivering Quality by Design.* Toronto: Longwoods Publishing;2008.
7. Bergeron B, Hamad A, Hoque E, AlBawardi M, Alswailem O. *Developing a Data Warehouse for the Healthcare Enterprise: Lessons from the Trenches*. Chicago, Illinois: Healthcare Information and Management Systems Society;2007.

Contributors:

Data Warehousing Workgroup

William Bennett, BlueCross BlueShield, Dallas, TX

Edna Boone, MA, CPHIMS, HIMSS Staff Liaison, Washington, DC

David Parker, MBA, PMP, CPHIMS, Harris County Hospital District, Houston, TX

Amy Thorpe, BS, MBA, St Luke's Episcopal Health System, Houston, TX

Major Wang, MBA, President, Decision Source, Inc.

Paul T. White, Baylor Health Care Systems, Dallas, TX