IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITB12777

This paper appears in the book, *Emerging Trends and Challenges in Information Technology Management, Volume 1 and Volume 2* edited by Mehdi Khosrow-Pour © 2006, Idea Group Inc.

Data Management Challenges for U.S. Healthcare Providers

Steven B. Dolins & Robert E. Kero
Bradley University and Health E-Discovery, Bradley University, Dept of Computer Science, 1501 W. Bradley Ave., Peoria, IL 61625
P: (309)677-3284, sdolins@bradley.edu

ABSTRACT

This paper identifies four major data management problems related to healthcare information systems: providers have complex application portfolios, system heterogeneity, data integration issues, and numerous, complex financial contracts to manage. Two emerging trends to solve some of these data management problems are described: electronic medical records (EMR) and national data repositories. This paper offers insight on how to make sure these types of systems are successfully developed and deployed.

INTRODUCTION

In 2004 President George W. Bush signed an Executive Order "to provide leadership for the development and nationwide implementation of an interoperable health information technology infrastructure to improve the quality and efficiency of health care". [1] The order further stated that this infrastructure:

"(a) Ensures that appropriate information to guide medical decisions is available at the time and place of care;

(b) Improves health care quality, reduces medical errors, and advances the delivery of appropriate, evidence-based medical care;

(c) Reduces health care costs resulting from inefficiency, medical errors, inappropriate care, and incomplete information;

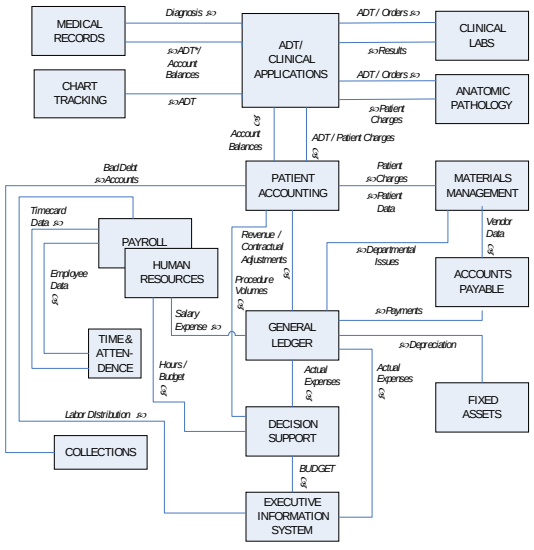
(d) Promotes a more effective marketplace, greater competition, and increased choice through the wider availability of accurate information on health care costs, quality, and outcomes;

(e) Improves the coordination of care and information among hospitals, laboratories, physician offices, and other ambulatory care providers through an effective infrastructure for the secure and authorized exchange of health care information; and

(f) Ensures that patients' individually identifiable health information is secure and protected."

In summary, this executive order lists data requirements for healthcare information technology (IT): the need for appropriate, timely, complete, accurate, and accessible information. This paper describes some of the data management challenges facing the IT departments at health care provider organizations, i.e., why these requirements are difficult to meet, and emerging trends and solutions.

Figure 1. Sample hospital information system data flow



**ADT – Admissions Discharges Transfers*
Source: The Zero-Base Hospital – Survival and Success in America's Evolving Healthcare System , P.33

CURRENT PROVIDER ENVIRONMENT

Data management is a very complex and costly undertaking for healthcare providers. A portfolio of software components that define a "typical" provider environment is shown in Figure 1. [2] A data driven system interaction in this environment could occur as follows: A patient enters the hospital and personal and payer information is collected, when applicable a hospital room assignment is made, diagnoses are rendered, procedures are conducted, services are utilized, tracked, and recorded, and discharge information is captured. Many hospitals have older, more brittle, less integrated legacy systems that support these business functions; this is why patients are consistently asked to provide all of their personal information during each visit. Throughout the remainder of the paper the term provider will be synonymous with hospital.

At the heart of the provider business applications is the patient accounting system, which is often a product from a leading software supplier based on proprietary data architecture. While many of the software components in the provider environment are designed to track patient care, the patient accounting system is used to keep track of all billable services utilized.

Although not designed for patient care, billing data contains both clinical and financial data which can be further integrated and analyzed to support patient care and improve provider performance. For example, we can perform segmentation analysis to see how utilized services are associated by Diagnosis Related Group (DRG), compare DRG-based service utilization patterns across physician populations, and evaluate patient population care patterns against outcomes. [3] [4] A DRG is a code used to represent clinical diagnoses. There are some researchers

who take a more pessimistic view and claim these data items are not adequate for patient care improvement analysis. [5 pg. 25]

The portfolio of interrelated, legacy, transaction-based applications can be quite costly and difficult to leverage data from, but the problem should not be seen as insurmountable. Numerous vendors and provider IT groups have built sophisticated mechanisms to harvest and further leverage these silo data assets. [6] [7] [8] [9] [10] [11] [12] [13] Often data repositories and warehouses are developed to cleanse, integrate, enrich and manage this data, and to provide a strategic platform for report generation and analysis. So why *aren't* data repositories and warehouses considered the silver bullet needed to solve data management issues and to improve the delivery of care?

DATA MANAGEMENT CHALLENGES

Despite the technological advances in the past thirty years and the significant investments in IT by the medical community, computer technology has produced little innovation in the organization and delivery of care. [14, pg. 106] There are many social, economic, political, and technological reasons for this but one important technological reason is that data has been largely inaccessible. This is partially due to the structure of existing application portfolios.

The provider environment is complex; it has many interrelated components. Therefore, it is difficult to extract data. Consequently, this makes it difficult to integrate heterogeneous sources of data.

Software has been successfully used to solve many technical and scientific problems in complex domains; NASA has computerized robotic vehicles that can navigate Mars, nuclear power plants have automated real-time systems for control, etc. So what makes the provider environment so uniquely complex? Besides the obvious budgetary differences, the provider environment consists of many interrelated yet virtually autonomous departments, and each department frequently manages its own computer systems and data requirements. The systems are often legacy investments that do not easily lend themselves to flexible data extraction or history capture. Data from each of these systems is difficult to share because they regularly follow proprietary data formats.

While the typical IT environment for a provider organization is made up of many interrelated systems, the problem complexity is further compounded through hospital mergers, acquisitions, and Integrated Care Delivery Networks. For example, a hospital network can often have different patient accounting systems in each of its affiliated hospitals.

Integrating data requires more than just a mapping of the data sources from functionally-bound stove-pipe applications, such as patient accounting, admissions, laboratories, etc.; multiple vendors may store semantically different data that has to be integrated as well.

The level of system heterogeneity directly affects the complexity of the data integration problem. Hence, the longitudinal tracking of patients across multiple hospital visits even within a network is difficult. For example, data integration problems are due to data incompatibilities in representation and data incompatibilities in entity identification, e.g., unique patient identifiers, coding, etc. across the different applications or application suites. [15 pp. 59-60] While data cleansing and editing, transforming, and loading (ETL) tools have matured, providers using ETL tools are still challenged based upon the enormity of the integration effort. The discovery of data quality issues, e.g., data entry mistakes made during patient admissions, greatly complicates the data integration effort and lowers subsequent decision-making confidence.

Capturing information about the health of a patient can only be captured in part through hospital encounters. Critical patient information is collected during primary care visit(s) prior to and often after a hospital stay. This information may offer insight about the severity of illness for a patient and also their eventual length of stay for a specific healthcare incident.

Information is unavailable. Physicians responsible for the care of a patient during a hospital visit can not access a complete patient history or the data collected at previous primary care visits, e.g., by an internist, and specialty care visits, e.g., by a cardiologist or orthopedist. That is, data is not easily shared or exchanged.

Many historical paper-based processes have not been evolved to collect patient visit data electronically. The state-of-the-art practices collect electronic medical records (EMR) but do not attempt to integrate data with the provider.

Financial data is also difficult to collect in the provider systems. There are different types of payment, e.g., per case basis, fee for service, different combinations, etc. and payment is received from different sources, e.g., government, payers, secondary payers, self insureds, different combinations, etc. Providers have many complex contracts with each payer and they face a mix of incentives and rewards.

Patient accounting and general ledger systems can not easily handle complex and frequently changing financial contracts across large heterogeneous payor mixes.

This section listed data management challenges. Unfortunately, IT vendors are not effectively helping address these technical problems. This is possibly because there is no quick fix, and there is more profit in selling new, killer applications, like surgical simulation, and innovative, trend setting quality improvement software, like six sigma. At the same time, hospital IT departments, with shrinking budgets, are focusing on how to just keep the legacy systems operational. The result is a lack of computer systems that can provide sufficient clinical and financial information to physicians, patients, and administrators to support informed decision making. [14 pg. 15]

TRENDS, PITFALLS, AND REMEDIES

The last section identified current shortcomings of IT systems applied to health care. Some of these shortcomings can be remedied with multiyear commitments for full scale reengineering efforts. These kinds of IT projects are impractical and risky because they require sizable capital investments and the likelihood of commercial success is uncertain[1]. What practical steps can be taken so that IT systems can better meet health care needs, that is, to make patient information complete, accurate, and accessible? This paper identifies two emerging trends to solve data management problems and offers insight on how to make sure these trends are successfully developed and deployed. The emerging trends in healthcare IT are electronic medical records (EMR) and national data repositories.

Both trends will now be introduced, potential pitfalls to their successful implementation will be discussed, and remedies for the pitfalls will be recommended.

Designing Seamless, Complete EMR Systems Based Upon Standards

Many vendors have already successfully implemented and deployed EMR systems. For example, one web page lists over thirty companies with EMR products. [16] The desired result is to remove waste from the process, i.e., reduce paper and costs and improve quality of care. Unfortunately, EMR's are not the panacea that everyone would like to believe because they uncover new levels of data integration and data quality issues. For example:

- Vendors are building their own specific proprietary EMR's rather than using agreed upon standards; this will make integration with other IT systems difficult. [17 pg. 45] Without standards, there will be no simple way to relate data together,
- There are competing medical classifications for some concepts and no agreed upon classification to use in the EMR. For example, should a DRG, ICD-9, CCS (clinical classification software), MDC (major diagnosis category), or some combination for diagnoses be stored in the record, and
- How should subjective and inexact data related to a patient visit be stored? Traditional databases are excellent tools for storing numeric data, such as blood pressure, weight, height, etc. and

alphanumeric data such as drugs prescribed for patients. However, how do we store inexact information such as the size, location, and color of a rash on a patient's neck?

These problems are compounded because patient data is usually collected from several different sources: primary care physician office, specialist physician(s) office, laboratory, provider(s), pharmacy, and payer(s). This creates serious data integration, management, and security issues.

In order for the medical information of a patient to be totally integrated, all information has to be portable in one common, canonical form. That is, information about every patient visit to a physician and hospital, drug prescribed, treatment received, and test performed should be captured, collected, and stored in one EMR that is accessible and consumable.

Hospitals, physician offices, laboratories, pharmacies, payers, and IT vendors need to agree on a common, canonical form. We should be asking the following questions: What data items make up the EMR? Are these the data items that physicians universally collect and how may they change as medicine evolves? How will data from external sources be stored in the EMR, e.g., a common laboratory blood test? Real collaboration between payors, providers, and physicians is required to make EMR's accessible and consumable.

Even in the best case scenario where collaboration between providers and physician office management systems exist, there will always be complex data integration issues; the current set of commercially available tools, e.g. ETL, would be challenged to solve them. However, these tools should continue to be enhanced to specifically address medical data, i.e., intelligent capabilities for matching patient records from different sources, for matching different coding standards, for finding subsets of patients for further analysis, for removing outlier data from further analysis, e.g., patients with severe symptoms. The next generation ETL tools should be specifically designed for handling EMR's.

Building national data repositories.
Governmental agencies and private enterprises are currently collecting national sources of data. For example, there are a number of state level registries that keep track of immunizations for children. [18] There are also national sources of aggregated provider data available, e.g., Agency for Health Care Research and Quality (AHRQ) and Solucient. [19] [20] [21]

One example is the Nationwide Inpatient Sample (NIS) which contains inpatient data from providers, i.e., clinical and resource information available from discharge abstracts. NIS is a part of the Healthcare Cost and Utilization Project (HCUP) sponsored by AHRQ. NIS contains all-payer data and it is designed to approximate a 20% sample of US hospitals drawn from 22 states. The universe of U.S. community hospitals is divided into strata using five hospital characteristics: ownership / control, bed size, teaching status, urban/rural location and U.S. region. Every year, a sample is drawn by a procedure that retains most of the hospitals in the previous year, while allowing new hospitals in the sample. NIS is publicly available for a small fee.

National sources of data are useful for building applications for research and policy analysis and can be used to support analyses such as

- Use and cost of hospital services,
- Hospital financial distress,
- Effectiveness of medical treatments,
- Utilization of health services by special populations,
- Impact of proposed cost-containment, and
- Medical practice variation.

Additionally, if physicians and administrators could easily leverage this information it would allow them to better understand evolving practice patterns for treating specific illnesses. Software applications would be needed to display and clearly explain information, such as:

- For a specific DRG, how many services, service types, and service days are typically utilized in the course of treatment by a physician and his/her peer group organizationally, regionally and nationally, or
- For a specific DRG, what outcomes are typically realized by a physician and his/her peer group organizationally, regionally and nationally.

From this health information, physicians would be better able to determine whether significant variations in practice or outcomes exist. This information would help support better evidence based decisions and identify training needs. Furthermore, the quality and consistency of care would be improved which would directly impact patient outcomes, levels of safety, effectiveness, timeliness, efficiency, and equity.

National sets of data present their own data integration issues. For example, how should the data be stored and formatted so that it can be easily disseminated to different types of users? These data sets are typically large and stored as files of raw data; this makes the data inaccessible to a large number of potential users. There is no standard language for retrieving data for different user needs.

An alternative solution is to define a canonical reference model for health care delivery and map each data source to this model. The reference model would be an abstract representation of the business of health care delivery as practiced by providers, and of the automation that supports that business, both described in a common language intended to standardize and clarify discussion of health care delivery. Specifically, the reference model would be a collection of concepts, terms, and models: each of which would provide 'head-start' material for a particular aspect of the business – IT interaction. For example, one artifact would be a logical data model that would consist of an entity relationship diagram of the relational tables, Structured Query Language (SQL) or data definition language code for the creation of relational tables, and a data dictionary. The SQL language is industry standard and the code would work with all compliant RDBMS.

This approach would enable data to be easily retrieved for research queries, ad hoc queries, and for collaborative research efforts via extract programs. Other tools could be used to further model, query, analyze, and present information.

An advantage of this solution is that by using industry standard tools and a single, logical data model, collaboration between researchers is facilitated. Another advantage of this environment is that researchers can focus on research problems rather than data issues. Both of these advantages will facilitate higher quality research.

CONCLUSION

Despite technological advances in the past thirty years and significant investments in IT healthcare applications, critical information is not reaching patients, hospital administrators, and physicians. This can be attributed to many economic, legal, organizational, cultural, and technological problems. One technological problem is data management in the provider environment.

Four major data management problems have been identified in this paper. The problems are integrating data from many interrelated departments, integrating data in a network environment and from multiple vendors, integrating primary care and provider data, and handling frequently changing financial contracts.

Two emerging trends in the healthcare industry are beginning to address these data management challenges: EMR products and collections of national sources of data. Unfortunately, these new products and sources of data alone do not solve the data management challenges. EMR products need one common, canonical form so that hospitals, physician offices, laboratories, pharmacies, and payers can easily exchange data. IT vendors need to build EMR systems with formats that facilitate data integration. They also need to drive the next generation ETL tools with special purpose functions, e.g., for intelligently matching patient records across different sources, to facilitate integration and make EMR accessible and consumable.

The second trend cited is the collection of national sources of data. These data sources are useful for building applications for research and policy analysis. To support sharing and diverse usage requirements these data sources need to be designed based on a canonical logical data model.

ENDNOTE

[1]These activities would compete with direct clinical needs that are seen as the true investment for the provider business.

REFERENCES

[1] Executive Order - Incentives for the Use of Health Information Technology and Establishing the Position of the National Health Information Technology Coordinator, http://www.whitehouse.gov/news/releases/2004/04/20040427-4.html

[2] M. Person, *The Zero-Base Hospital Survival and Success in America's Evolving Healthcare System*. Chicago, IL: Health Administration Press, 1997, pg. 33.

[3] A. Sumner and C. Moreland, The potential impact of diagnosis related group medical management on hospital utilization and profitability, *Health Care Management Review*, Spring, 1995, Volume 20, No. 2, pp.92-100.

[4] M. Silver, T. Sakata, H.C. Su, C. Herman, S.B. Dolins, M. O'Shea, "Case study: how to apply data mining techniques in a healthcare data warehouse", *Journal of Healthcare Information Management*, Vol. 15, No. 2, pp. 155-164, Summer 2001.

[5] J. D. Kleinke, Release 0.0: Clinical information technology in the real world, *Health Affairs*, Volume 17, No. 6, Nov/Dec 1998, pp. 23-38.

[6] R. Scheese, Data warehousing as a healthcare business solution, *Healthcare Financial Management*, February 1998, pp. 56-59.

[7] W. Herr, The benefits of data integration: HFMA study findings, *Healthcare Financial Management*, September 1996, pp. 52-56.

[8] A. Keegan, The need to integrate clinical and financial information, *Healthcare Financial Management*, July 1995, pg. 74.

[9] C. Breen and L, Rodrigues, Implementing a data warehouse at Inglis Innovative Services, *Journal of Healthcare Information Management*, Volume 15, Number 2, Summer 2001, pp. 87-97.

[10] D. Ramick, Data warehousing in disease management programs, *Journal of Healthcare Information Management*, Volume 15, Number 2, Summer 2001, pp. 99-105.

[11] R. Verma and J. Harper, Life cycle of a data warehousing project in healthcare, *Journal of Healthcare Information Management*, Volume 15, Number 2, Summer 2001, pp. 107-117.

[12] J. Einbinder, K. Scully, R. Pates, J. Schubart, and R. Reynolds, Case study: a data warehouse for an academic medical center, *Journal of Healthcare Information Management*, Volume 15, Number 2, Summer 2001, pp. 165-175.

[13] A. Cheriyan, Data warehousing, *Healthcare Informatics*, February 1999, pp. 102-103.

[14] Committee on Quality of Health Care in America, *Crossing the Quality Chasm A New Health System for the 21st Century*. Washington D.C.: National Academy Press, 2001.

[15] J. D. Kleinke, Vaporware.com: The failed promise of the health care Internet, *Health Affairs*, Volume 19, No. 6, Nov/Dec 2000, pp. 57-71.

[16] List of healthcare vendors including EMR vendors www.health-infosys-dir.com

[17] C. J. McDonald, Need for standards in health information, *Health Affairs*, Volume 17, No. 6, Nov/Dec 1998, pp. 44-46.

[18] Centers for Disease Control and Prevention lists state agencies collecting immunization registries www.cdc.gov/nip/registry

[19] Agency for Healthcare Quality and Research website www.ahrq.gov

[20] Agency for Healthcare Quality and Research website www.ahrq.gov/data/hcup

[21] Solucient website www.solucient.com/solutions