Data, Information, Knowledge: A Healthcare Enterprise Case Study

Save to myBoK

by Sorin Gudea

Abstract

An efficient, integrated health services delivery enterprise requires the ability to coordinate service delivery across the provider network and avoid duplication of services. It must be able to associate relevant clinical information with patients regardless of which facility delivered the services. There are significant challenges in collecting, organizing, and extracting value from data collected in the course of providing healthcare. This paper follows a large urban public healthcare enterprise in its attempts to address some of these challenges. Using a case-study methodology, the paper shows how information technology (IT) can help a healthcare organization derive improved information and generate knowledge from data stored in disjoint systems.

Introduction

Data are the fundamental element of cognition, the common denominator on which all constructs are based, and are stored in information systems. Derived from data, and positioned along a continuum that eventually leads to wisdom, are information and knowledge. Data pertain to facts and given attributes, such as name, gender, birth date, address, phone number, temperature, and so forth. Attaching meaning to data transforms them into semantic data, or information. Knowledge, at the next level, implies contextualized information, which is information interpreted by the receiver and from the perspective of the receiver. The highest level on this continuum—wisdom—pertains to a state of refined, sublimated knowledge that affords the receiver the potential to optimize interaction with the environment.[1]

From Data to Information and Knowledge

In many instances the distinction between information and knowledge is rather ambiguous. Certain users may interpret one set of data as information, while for others it is knowledge.[2]

To minimize equivocation, an information system uses a database to store data and metadata, which are data about data. Metadata help interpret and transform data into information. Large organizations often store the same data in different systems. For every given system, one must take into account metadata when attempting to interpret data. Sometimes, additional data elements must be considered together (for example, a patient's name may be stored in three different data fields, as last name, first name, and middle name) before information can be derived from data. For healthcare enterprises, the complexities involved in accessing patient information across different software applications and across organizational boundaries are significant, yielding a potential for costly errors.

Managing Information as a Resource

A healthcare enterprise has different facilities and systems that are used for processing patient data. A patient may receive care at more than one physical facility affiliated with the same organization, sometimes by random occurrence and sometimes as part of the medical management process. In such an instance, and when no data are exchanged between the facilities' systems or incomplete data exchange occurs between disjoint systems, the same patient may end up with two different identifiers or medical record numbers within one healthcare information system (HIS). The issue of dealing with duplicate data has pestered data warehouse initiatives for quite a while. Such situations require data cleansing, record cleanup, and record linkage efforts.

Data cleansing and record cleanup involves detection, removal, and correction of incorrect, incomplete, or poorly formatted data and deals with duplicate records within each system Record linkage relies on powerful algorithms to "match" records in disjoint systems to a unique enterprise-wide identifier.

Information Technology Solutions for Enterprise Data

There are numerous technologies available to the system integration practitioner. System integration solutions for disjoint systems involve data exchanges among the participant systems. Electronic transmission of data among disjoint systems across organizational boundaries can be ensured by relying on standards for information content, format, and exchange mechanisms.[3]

Middleware has been around since the mid-1980s, assisting companies in extracting data and exchanging data among different platforms. With the advent of the Internet, open data transmission and access standards such as HTTP, FTP, XML, and SOAP started to find their way into the middleware realm. Web services complement other middleware technologies rather than replacing them completely.[4]

In healthcare, various vendors offer enterprise-wide systems, commonly known as Enterprise Master Person Index (EMPI) systems. The master index is in fact a global identifier for the patient. Its purpose is to provide a common, unique patient identifier that supports data integration across multiple yet similar systems. While this approach works well in an enterprise environment, system integration across a heterogeneous environment nonetheless encounters many difficulties stemming from the diversity of operating systems, databases, and HIM applications used.

Health Level Seven (HL7) is a standard for electronic exchange of patient medical record information supported by the National Committee on Vital and Health Statistics (NCVHS). HL7 is developed by a standards body accredited by the American National Standards Institute. It is a messaging standard that allows software applications to exchange information across platforms in a manner that preserves the meaning of the information transmitted. This requires that data be transmitted in a specific, predefined form.[5]

The HCO Case Study

During the spring of 2004 the researcher had the opportunity to study a project team at a large public healthcare enterprise on the West Coast. The team was responsible for an information system project that would integrate patient demographics across the enterprise and offer access to improved demographics and clinical information. In order to maintain the privacy of the participants, this enterprise will be referred to as Health Care Organization (HCO).

The Challenge
HCO is a public healthcare enterprise consisting of networks of clinics and hospitals delivering medical services over a large geographical area. Maintaining correct patient identification data, reliably linking patients to the correct medical record information, and making health information available across its vast network of providers are the issues known to be a problem in this organization. The information system team, using information technology, is trying to alleviate the problems stemming from inconsistent data. Low-data quality is often a source of medical errors and can lead to increased costs.[6]

The Baseline Environment
Why is patient identification such an issue for HCO? The main answer resides in the fact that the vast majority of the population served by HCO is not enrolled in a medical benefit plan. In the case of private-sector healthcare organizations, members are identified through health plan numbers, social security numbers, or employee numbers. Moreover, enrolled members of a medical benefit plan have an incentive to cooperate with the identification process, in order to ensure access to benefits. These factors are not true in the case of HCO. At the same time, just as in the private sector, access to complete and accurate patient medical records is important and has value both for supporting the quality of care and for reimbursement. Making health information available across the network of providers is an issue in HCO's situation because many of its provider sites are not linked by common information systems. Furthermore, some of the clinics and hospitals have their own, independently implemented healthcare information system and their own medical record system. In this situation, a clinician providing care at one facility has to rely on the patient to report what other medical interventions were provided to him or her at another facility. At HCO, as many as 60 percent of patients are new to the system every year. Some empirical evidence indicates that identification error rates (due to duplicate records and overlays) are around 4 percent, and perhaps as high as 14

⏎

percent in some systems. The high patient turnover rate makes the task of the information system team even more challenging. The solution must help improve HCO's operational efficiency and quality of care (Figure 1).

Figure 1

HIM Improvement[7]
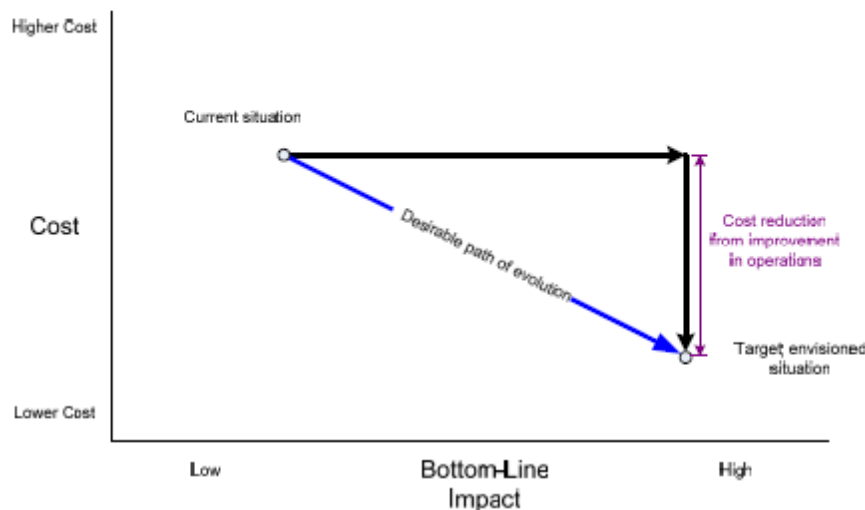
The data processing architecture at HCO contains a number of disjoint, or separate and distinct, applications and application-specific databases. It is a complex environment with many vertical applications, each with its own database. This makes aggregation of data into information very difficult (Figure 2). In the literature, this is referred to as the "islands of data" environment.[7]

Figure 2

Disjoint Systems
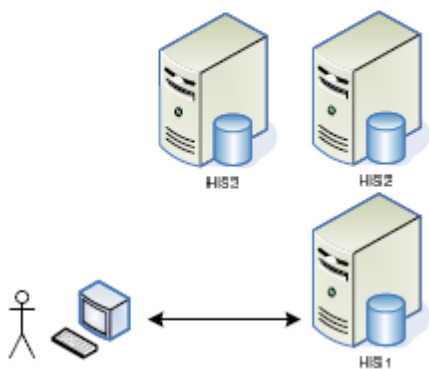

Alternatives and Considerations
The solution identified by the project team relies on an enterprise messaging architecture. Middleware messaging comes in as a possible solution for the following reasons:

- It uses a standards-based messaging interface.
- It does not disrupt existing data processing routines.
- It offers transparency.
- It causes no disruption to user activities and business processes.

More important, it is a loosely coupled architecture that allows systems to evolve independently and allows applications to be added, removed, or changed more easily. Demographics processed at each patient-clinician encounter generate HL7 messages

(Admission, Discharge, and Transfer) that are processed by messaging gateways (one at each site) and sent to a central messaging hub, where they are stored in a data depository built on a relational database. Each of the participating messaging gateways takes into account data formats, data domains, and the originating system (Figure 3). Gateways use the metadata available to them in order to gather data from the local systems and transform them into information that gets stored in the data repository. As stringent state laws and regulations must be observed when accessing healthcare information, secure access and role-based security rules take into account all of the legal constraints that apply (e.g., HIPAA).

Figure 3

Connected Systems
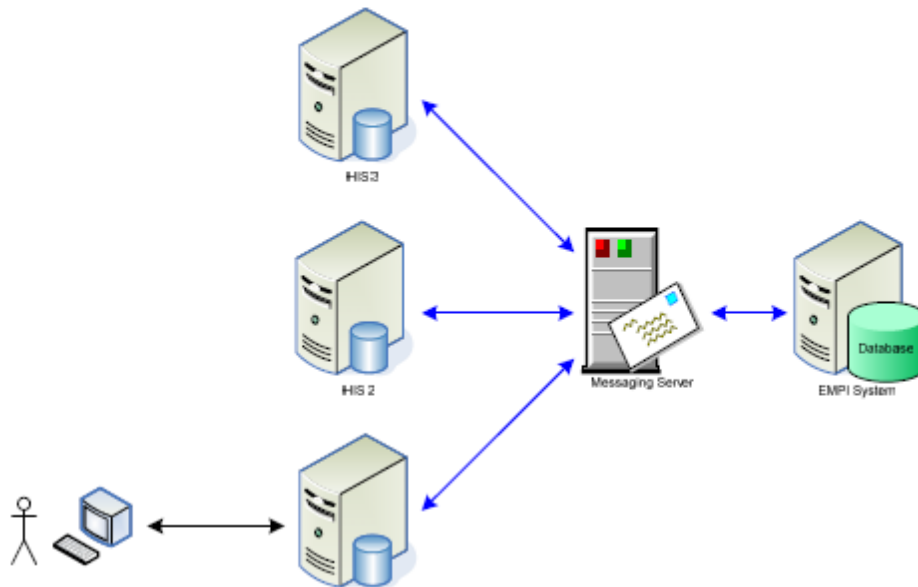
Across all of the data in the repository, a probabilistic record linkage/matching algorithm identifies records that are automatic matches (based on predefined thresholds), nonmatches, and possible matches. An enterprise-wide patient identifier is associated with the stored records that belong to a patient, regardless of the medical record number or local identifier used in the originating system.

Human intervention is required to resolve the possible matches as appropriate. While generic EMPI systems offer this type of functionality across one enterprise-wide system, the solution sought by HCO expands the functionality to a number of dissimilar systems that do not exchange data. The solution favored by HCO supports an incremental deployment strategy with minimal disruptions to operations, a nontrivial endeavor.

Team deliberations were lively, yet managed to address the numerous issues raised by the project stakeholders. The system, as designed, does not change the way patients are internally identified in these participating systems. Each continues to use its own internal identifier. The system serves as a master index to all of these identifiers, and the enterprise ID links records that pertain to one patient.

Registration and clinical personnel have the ability to query the data repository and, based on medical record number or local patient identifier, find information about patient encounters that may have taken place at other facilities. Because more complete information can now be associated with each patient, improved medical care is being provided. More is *known* about the patient, akin to knowledge creation. Knowledge is now available that pertains to the patient (Figure 4).

Figure 4

Data Transformation
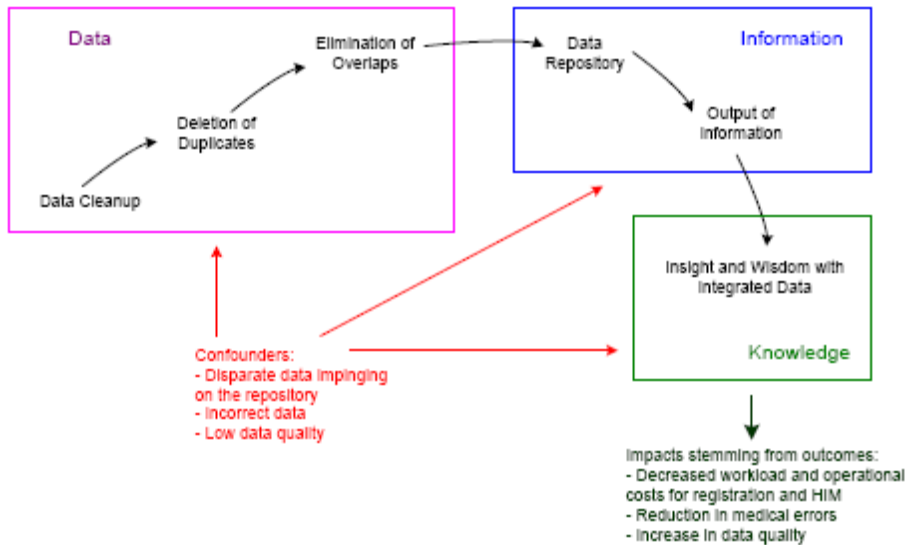
More complete demographical information is derived for the financial side of the patient encounter. In addition to registration staff, billing clerks have access to more complete information for each patient. Again, as more is *known* about the patient, the financial side of the clinical interactions becomes more accurate, leading to increased effectiveness and efficiency.

The Solution
Given the complexity of the existing environment and the large number of disparate systems involved, this loosely coupled system integration architecture made the most sense for the project team because the enterprise solution allows systems to be integrated one by one. Among all the choices considered by the project team, the most effective one in terms of cost and functionality appears to be the purchase of an off-the-shelf solution, with minimum modifications to accommodate local requirements. While the above may seem self-evident, the complexities of the messaging architecture are still significant. Recognizing this, HCO has decided to approach the project implementation in a phased manner. For the initial scope of the project, only selected demographics and clinical data are collected from each of the facilities and stored in the central data repository system. Still, the system, as envisioned, will help HCO derive information from its data and generate knowledge that can be disseminated as needed.

Expected Outcomes
As the implementation proceeds, once the initial data cleanup is conducted, it is expected that the system will lead to a significant reduction in the number of duplicates and overlays in the patient databases, an indicator of the solution's effectiveness. In time this will help reduce workload and operational costs for both registration and HIM staff.

Sorin Gudea teaches technology courses in the College of Information Systems and Technology at University of Phoenix, in the Southern California Campus. He is a PhD candidate in information science at Claremont Graduate University, in Claremont, CA.

Acknowledgments

Notes

1. Nonaka, Ikujiro. "The Knowledge Worker." In Peter F. Drucker (Editor), *Harvard Business Review on Knowledge Management.* Boston, MA: Harvard Business School Press, 1998.
2. Davenport, Thomas H., and Laurence Prusak. *Working Knowledge.* Boston, MA: Harvard Business School Press, 2000.

3. Klein, Jim, and Ned Frey. *Requirements for Integration Brokers in Healthcare: Strategic Analysis Report.* Stamford, CT: The Gartner Group, September 24, 2001.
4. Gibbons, Paul L. "Proprietary to Open: Middleware Evolves." *CIO Magazine*, June 15, 2004, 89–93.
5. Tracy, Wayne R., and Michelle Dougherty. "HL7 Standard Shapes Content, Exchange of Patient Information." *Journal of AHIMA* 73, no. 8 (2002): 49–51.
6. Rippen, Helga E., and William A. Yasnoff. "Building the National Health Information Infrastructure." *Journal of AHIMA* 75, no. 5 (2004): 20–26.
7. Inmon, William H. *Building the Data Warehouse*. Hoboken, NJ: Wiley, 1993.

References

Christen, Peter, et al. "Probabilistic Name and Address Cleaning and Standardization." Presented at the Australasian Data Mining Workshop, Canberra, Australia, December 2002. Available at http://datamining.anu.edu.au/publications/2002/adm2002-cleaning.pdf (accessed July 6, 2004) [no longer active].

Graham, Gail, et al. "Information Everywhere: How the EHR Transformed Care at VHA." *Journal of AHIMA* 74, no. 3 (2003): 20–24.

Gu, Lifang, et al. "Record Linkage: Current Practice and Future Directions." Technical Report 03/83, CSIRO Mathematical and Information Sciences, Canberra, Australia, April 2003. Available at http://www.act.cmis.csiro.au/rohanb/PAPERS/record_linkage.pdf (accessed July 6, 2004) [no longer active].

HL7 Protocol Committee. *Health Level Seven Implementation Support Guide, Final Version—6/98 for HL7 Standard Version 2.3*. Ann Arbor, MI: Health Level Seven, Inc., 1998. Available at http://www.hl7.org/library/standards_non1.htm (accessed June 21, 2004).

Wing, Paul, and Margaret H. Langelier. "The Future of HIM: Employer Insights into the Coming Decade of Rapid Change." *Journal of AHIMA* 75, no. 6 (2004): 28–32.

Article citation:
Gudea, Sorin. "Data, Information, Knowledge: A Healthcare Enterprise Case Study." *Perspectives in Health Information Management* 2005, 2:8 (November 8, 2005).