Defining the Landscape:
Data Warehouse and Mining: Intelligence Continuum

A Work Product of the
HIMSS Enterprise Information Systems Steering Committee

Defining the Landscape:
Data Warehouse and Mining: Intelligence Continuum

Submitted by:

Nilmini Wickramasinghe, PhD, MBA
Associate Professor and Associate Director
Center for the Management of Medical Technology (CMMT)
Stuart School of Business
Illinois Institute of Technology
565 W. Adams St., Ste. 406
Chicago, IL 60661
Tel: 312-906-6578
Fax: 312-906-6549
nilmini@stuart.iit.edu

Richard A. Calman, CPHIMS
Deputy Chair, Enterprise IS Steering Committee
Senior Consultant, ACS Healthcare Solutions
Tel: 248–386–8300 ext. 4229
Cell: 908-463-8764
calmanra@umdnj.edu

Jonathan L. Schaffer, FHIMSS, MD MBA
Member, Enterprise IS Steering Committee
Managing Director, eCleveland Clinic
Information Technology Division
Cleveland Clinic
9500 Euclid Avenue, Desk A41
Cleveland, OH 44195
Tel: 216-444-8960
Fax: 216-444-9255
schaffj@ccf.org

Introduction

The care of patients requires the evaluation of a significant amount of data at the right time and place and in the correct context. Moreover, there is a significant amount of data hidden from the patient-care environment that helps to define and control specific events in healthcare. These clinical, administrative and operational sources of data are typically kept in separate and disparate operational repositories; a master set of data can be kept in a single data repository from which queries can be made that cross these specific disciplines.

Alternatively, virtual agents can search these separate data sets simultaneously, and combine at another level to provide a response to a query. Combining all the disparate data into a single repository—a data warehouse—will result in the creation of a store of data that can be used to make intelligent clinical and management decisions about healthcare and its delivery. This combination of data sets will lead to improved patient care through the harnessing and evaluation of this rich data content for a variety of healthcare-related improvement purposes, ranging from improving overall outcomes of care for patients and support for clinical research to economic issues, such as product-line cost and clinical productivity costs.

Given the advancement of the information tools and techniques of today's knowledge-based economy, it is imperative that they be appropriately utilized to enable and facilitate the identification and evaluation of pertinent information and relevant data about the efficiency and effectiveness of delivering healthcare. With the advent of the electronic health record, data warehouses will provide information at the point of care, and provide for a continuous learning environment in which lessons learned can provide updates to clinical, administrative and financial processes.

Knowledge Management

Knowledge management is an emerging management approach aimed at solving business challenges to increase efficiency and efficacy of core business processes while incorporating continuous innovation. Specifically, knowledge management through the use of various tools, processes and techniques combines germane organizational data, information and knowledge to create business value and enable an organization to capitalize on its intangible and human assets so that it can effectively achieve its primary business goals as well as maximize its core business competencies [1].

The need for knowledge management is based on a paradigm shift in the business environment, where knowledge is central to organizational performance [2]. Broadly speaking, knowledge management involves four key steps: creating/generating knowledge; representing/storing knowledge; accessing/using/re-using knowledge; and disseminating/transferring knowledge[1]

Knowledge management is particularly important to ensure that relevant data, pertinent information and germane knowledge permeate systems at all times, and the extant knowledge base continues to grow in a meaningful and useful fashion.

The Intelligence Continuum

Data permeate all areas of healthcare and permit the appropriate delivery of the correct services at the right time for each patient at the point of care. The application of Internet-based information communication technologies (ICT) to healthcare is a necessary but not sufficient solution to address today's healthcare challenges. To improve access and quality, and thereby realize the value proposition for healthcare, healthcare organizations must maximize the data and information that are generated by and flow through these ICTs. These goals can be reached by embracing the techniques of data mining and the strategies of knowledge management. When coupled with evidence-based medicine, and the introduction of various Internet-based ICTs such as electronic medical record systems and e-health initiatives, knowledge management and data mining create strategic imperatives for healthcare organizations to maximize both the clinical and administrative benefits from their application of Internet-based information and communication technologies to healthcare. As knowledge is created from the experiences of today, the techniques of data mining and strategies of knowledge management can be effectively and efficiently applied in healthcare to develop and apply new rules for the events of tomorrow, leading to a continuum of intelligence that is perpetual [3].

An understanding of the role of the intelligence continuum begins with an examination of a generic healthcare information system. The important aspects of this generic system include the socio-technical perspective; i.e. the people, processes and technology inputs required in conjunction with data as a key input. The combination of these elements comprises an information system, and in any organization multiple such systems could exist. To this generic system we add the healthcare challenges; i.e. the challenges of demographics, technology and finance.

To address these challenges a closer examination of the data generated by the information systems and stored in the larger data warehouses and/or smaller data marts is necessary. There are two philosophies that address data repository systems, that of Bill Inmon and that of Ralph Kimball.

Bill Inmon is considered the "father of the data warehouse." His philosophy maintains that enterprise data, although stored in disparate reporting systems known as data marts, must be tied to a larger repository (the data warehouse) for the sake of greater uniformity.

Conversely, Ralph Kimball (another data warehouse and business intelligence expert) espouses a methodology known as Dimensional Modeling. Under this method, individual data marts are logically connected using conformed dimensions. The warehouse is "virtual."

The data warehouse (Inmon) approach is costlier, takes longer to bring to fruition and will require organizations to have great patience in implementation. Large organizations with traditional data modeling skills in-house and understood will fare well here. The virtual warehouse, or data mart and Dimensional Modeling method (Kimball), is best suited in organizations where the required outcome to implementation must be fast and less expensive than the full blown warehouse.

In summary, the warehouse method is more top-down driven, while the data mart method is bottom-up in its approach. In either case, it is important to make decisions and invoke the intelligence continuum; apply the tools, techniques and processes of data mining, business intelligence/analytics and knowledge management, respectively. On applying these tools and techniques to the data generated from healthcare information systems, it is first possible to diagnose the "as is" or current state processes to make further decisions regarding how existing processes should be modified and thereby provide appropriate prescriptions to enable the achievement of a better future state; i.e. improve the respective inputs of the people, process, technology and data so that the system as a whole is significantly improved.

Components of the Intelligence Continuum

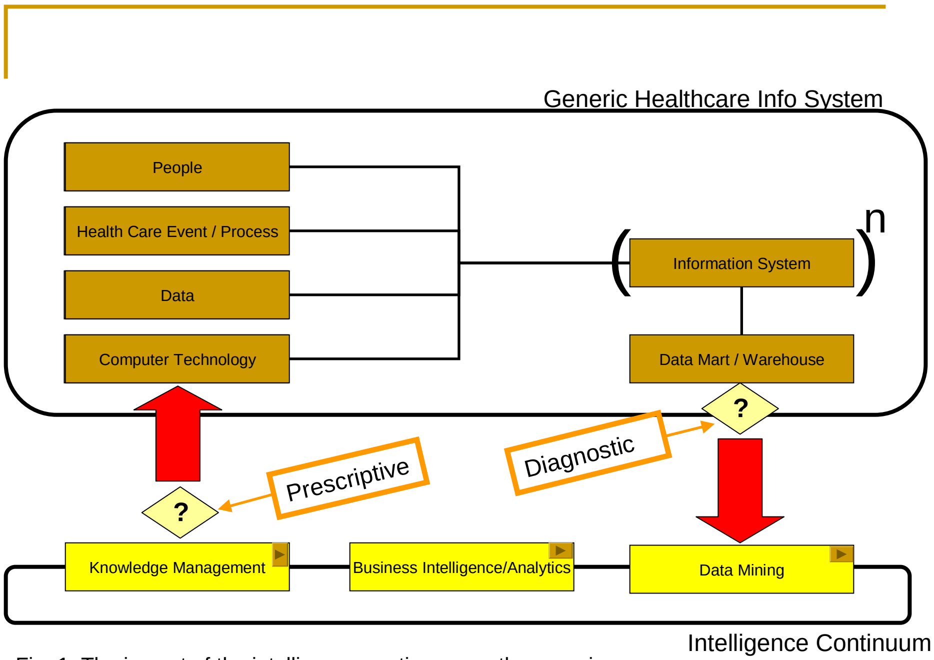

Fig. 1: The impact of the intelligence continuum on the generic healthcare info system [18]

The Intelligence Continuum is a collection of key tools, techniques and processes of today's knowledge economy; i.e. including but not limited to data mining, business intelligence/analytics and knowledge management. Taken together they represent a very powerful system for refining the data raw material stored in data marts and/or data warehouses and thereby maximizing the value and utility of these data assets for any organization. The first component is a generic information system which generates data that is then captured in a data repository. In order to maximize the value of the data and use it to improve processes, the techniques and tools of data mining, business intelligence and analytics and knowledge management must be applied to the data warehouse. Once applied, the results become part of the data set that are reintroduced into the system and combined with the other inputs of people, processes, and technology to develop an improvement continuum. Thus, the intelligence continuum includes the generation of data, the analysis of these data to provide a "diagnosis" and the reintroduction into the cycle as a "prescriptive" solution (Fig. 1). Each of the components of the intelligence continuum is now explored in more detail.

Data Mining

Due to the immense size of the data sets, computerized techniques are essential to help physicians as well as administrators understand relationships and associations between data elements. Data mining is closely associated with databases and shares some common ground with statistics since both strive toward discovering structure in data. However, while statistical analysis starts with some kind of hypothesis about the data, data mining does not. Furthermore, data mining is much more suited to deal with heterogeneous databases, data sets and data fields, which are typical of data in medical databases that contain numerous types of text and graphical data sets. Data mining also draws heavily from many other disciplines, most notably machine learning, artificial intelligence, and database technology.

Data mining then, is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns from data [4,5]. Clinicians accomplish these tasks daily in their care of patients using their own "personal CPU;" however, the enormous amounts and divergent sources of information coupled with time constraints limit any clinician's ability to fully examine all issues. Data mining algorithms are used on databases for model building, or for finding patterns in data. When these patterns are new, useful, and understandable, we say that this is knowledge discovery [5]. How to manage such discovered knowledge and other organizational knowledge is the realm of knowledge management [4].

Business Intelligence/Analytics

Another technology-driven technique, like data mining connected to knowledge creation is the area of business intelligence and the now newer term of business analytics. The business intelligence (BI) term has become synonymous with an umbrella description for a wide range of decision-support tools, some of which target specific user audiences [6,7]. At the bottom of the BI hierarchy are extraction and formatting tools which are also known as data-extraction tools. These tools collect data from existing databases for inclusion in data warehouses and data marts. Thus the next level of the BI hierarchy is known as warehouses and marts. Because the data come from so many different, often incompatible systems in various file formats, the next step in the BI hierarchy is formatting tools, these tools and techniques are used to "cleanse" the data and convert it to formats that can easily be understood in the data warehouse or data mart. Next, tools are needed to support the reporting and analytical techniques. These are known as enterprise reporting and analytical tools. Online analytic process (OLAP) engines and analytical application-development tools are for professionals who analyze data and perform business forecasting, modeling and trend analysis. Human intelligence tools form the next level in the hierarchy and involve human expertise, opinions and observations to be recorded to create a knowledge repository. These tools are at the very top of the BI hierarchy and serve to amalgamate analytical and BI capabilities along with human expertise. Business analytics (BA) is a newer term that tends to be viewed as a sub-set of the broader business intelligence umbrella and specifically focuses on the analytic aspects within BI [7].

Conclusions

The main goal of all healthcare organizations is to provide the best patient care possible. Given the demographic, financial and technologic challenges facing these healthcare organizations, an environment must be provided in which the appropriate management of operating costs can be attained while maintaining or exceeding an acceptable level of patient treatment. Improved levels of patient treatment involve the use of methodologies such as clinical pathways, computerized physician order entry and clinical decision support systems. Direct areas of cost management range from personnel, equipment, supply and pharmaceutical costs. Since the data for these care and expense categories often reside in different data systems, the integration of data in a data warehouse that is available for complex analysis is critical to the successful evaluation of entire patient care processes. Housing the clinical data in electronic medical record repositories provides further analytical capability that will lead to potential improvements. With trends moving toward the capture of all this data in digital form, researchers, clinicians and administrators are provided with a more efficient and effective means of accessing data to form hypotheses about disease initiation and progression, search for patterns in certain populations, conduct surveillance studies of new drugs, identify adverse events, improve prescribing practices and, perhaps most importantly, identify potential study candidates for clinical research purposes.

The intelligence continuum emphasizes the need for the continuous analysis of the current state of data and other socio-technical inputs by applying the tools and techniques of the knowledge economy to develop prescriptions for the attainment of a superior future state and enhanced extant knowledge base. While this is a global description, it can be used to explain the development and installation of a reporting system and the mechanisms required to harness the data in intelligent format to enable its use for clinical and administrative functions. In building an intelligence reporting capability, organizations are advised to consider two major alternatives:

- The design and implementation of a Data Warehouse; or
- The design and implementation of a series of data marts (the virtual data warehouse).

These two alternatives are representative of “schools of thought” espoused by individuals considered to be leading authorities in this specific IT discipline. There is, obviously, much more that is involved in implementing an organizational reporting system and structure. Other considerations beyond the scope of this paper include decisions about whether to purchase a system off the shelf or trying to build one internally and the selection of an extraction tool (software) that will enable the actual extraction of data from one system or another into the Dimensional Model and or the warehouse. Additionally, consideration should be given to the export capabilities of the system as other internal customers will want to export the data for their own reporting purposes. In all situations, the intelligence continuum will provide the capability to effectively extract data for learning from yesterday’s events to provide superior healthcare tomorrow.

References

1. Wickramasinghe, N. 2006. “Knowledge Creation: A Meta-Framework.” *Intl. J. Innovation and Learning* (IJIL) Vol. 3, Issue 5, ppg. 558-573.
2. Drucker, P. 1993. *Post-Capitalist Society*. Harper Collins, New York. ppg. 6-352.
3. Wickramasinghe, N.; Schaffer, J. “Creating Knowledge Driven Healthcare Processes With the Intelligence Continuum.” *Intl. J. Electronic Healthcare* (IJEH) Vol. 2, Issue 2, ppg.164-174. 2006.
4. Cios, Krzysztof J. (Ed.). “Medical data mining and knowledge discovery;” Physica-Verlag, 2001.
5. Fayyad, Piatetsky-Shapiro, Smyth. "From Data Mining to Knowledge Discovery: An Overview", in Fayyad, Piatetsky-Shapiro, Smyth, Uthurusamy, *Advances in Knowledge Discovery and Data Mining*, AAAI Press/ The MIT Press, Menlo Park, Calif., 1996.
6. Kudyba, S. and R. Hoptroff. *Data Mining and Business Intelligence: A Guide to Productivity*. Idea Group Publishing, Hershey, Penn. 2001.
7. Dhar, V.; Stein, R. *Seven Methods for Transforming Corporate Data into Business Intelligence*. Prentice Hall, Upper Saddle River, N.J. 1997.

