This paper discusses some of the key issues in data warehousing practices and opportunities in the healthcare industry. Beginning with an overview of the topic, the paper discusses briefly the current uses of industry data, basic terminology, the myriad users of data and information, and implications for health services delivery and policy.

**Overview and Foreshadow**

Healthcare systems face strong competitive pressures (Curtright, Stolp-Smith, and Edell 2000, Griffith 2000, Rivenson, Wheeler, Smith and Reiter 2000, Langabeer 1998, Zuckerman 2000) and have evolved into a highly competitive industry with healthcare providers having little competitive experience compared to other industries (Zuckerman 2000). Langabeer (1998) has indicated that for the first time in recent history, teaching hospitals are now exposed to a competitive market where smaller hospitals have been competitive for many years. Many organizations have found it necessary to clearly understand and communicate their capabilities (Zuckerman 2000) when evaluating planning responses. Such competitive pressures have led organizations to consider a performance management approach to organizational planning efforts (Curtright et al. 2000), and to consider the applicability of data warehousing for making informed decisions (Awalt and Lawton 2000, Hammer 2002, Orr 2000). Orr (2000) suggested that the single key to business survival in the 1990s and beyond is the ability to analyze, plan and react to business conditions more rapidly than in the past. Information technology development has allowed organizations across many industries to reap benefits in how data are managed and converted to useful information (Birlasoft 2000, Hammer 2002, “Reaping the Rewards” 2002, Love and Burwen 1999, Meditech 2002, Moore and Wells 2000, Orr 2000, Sederholm 2002, Whiting 2001). The healthcare industry has embraced this technology passionately as organizations small and large begin to reap the rewards of data warehousing, data mining, and related information management techniques.
**Terminology**

The focus of this paper is on data warehousing as a model for managing the information of an organization. Awalt and Lawton (2000) described a data warehouse as “the cohesive data model that defines the central data repository for an organization” (p.1.) (see also Figure 1). This complete integrated data model differs somewhat from a data mart, which is a repository of specific information among homogenous users or subsets of an organization. Data warehousing, as defined by Awalt and Lawton (2000), is the process of managing the data warehouse and data marts (see also data warehousing terminology in Appendix 1). Users access data warehouses in a variety of means, most though for some form of structured queries that result in the specific data sought. These data may be represented digitally as images, as is the case with the picture archiving and communications (PACS) systems commonly used in radiology, through graphical visual modes (e.g., histograms, charts, or tables) as in decision support systems, or in other forms of utility to the user. These data may be comprised of clinical, demographic, financial, operating, supply, workforce and other elements of the organization or its environment.

**Current Issues**

Data are used increasingly by a wide variety of employees, patients and other stakeholders in the healthcare industry (Figure 2). Data access and management have become some of the prime areas of importance in healthcare management and consumers are increasingly accessing organizational information (Evans, 2000). As healthcare delivery has become firmly intertwined with its financing mechanisms, payers and providers rely fully on financial and operational data for making appropriate decisions. This array of users has created a nearly endless set of circumstances and applications for managing the information desires of both the organization and its constituents.

**Users**

Many entities are potential users of such data including hospitals, physician practices, ambulatory centers, home health agencies, rehabilitation and subacute facilities, skilled nursing centers, mental health centers and insurers. Data queries and analyses have been facilitated greatly with the development of sophisticated computer languages and networking capabilities (Angus 1997, Bohn 1997, Bontempo and Saracco 1997, Morgan and Lauer 2002, Sederholm 2002). Love and Burwen (1999) indicated that financial and sales/marketing applications are the most popular uses for
business intelligence and data warehousing in the general marketplace, while Meditech (2002) and others see the healthcare industry moving rapidly (Whiting 2001). E-business applications, the ability to reduce transaction and supply chain costs may well be the most promising aspect of the practice (Evans 2000, Hammer 2002), and these hopes continue to expand the user base. In the healthcare environment its users may be best described through some of the most prevalent applications of these techniques in the field.

Applications
Hundreds of sources of information on data warehousing, data mining and information management speak to the possibilities for these techniques (see also www.dwinfocenter.org). Figure 2 lists some of the current practices among healthcare organizations that involve data warehousing. Organizations of greater size, complexity, and resources are more likely to engage data warehousing than smaller organizations, primarily due to the time and costs involved in comparison to the competing interests of other program needs for scarce resources. It is likely however that even the smallest of organizations can benefit from a coordinated approach to data management (Adelman and Moss 2000). Such investigation begins with an assessment or cataloging of needs, followed by an analysis of current system components and architecture as the basis for the internal needs assessment (Dyché 1998, Moore and Wells 2000) and budget development (Inmon 1997) before progressing to design and implementation.

Data conversion and implementation (see Figure 3) is most often fraught with difficulties in execution. Causes identified in the literature range from poor planning (Adelman and Moss 2000, Angus 1999, Awalt and Lawton 2000, Baldwin 2002, Bohn 1997, Dyché 1998, Lerner 2001, Love and Burwen 1999, Moore and Wells 2000, Orr 2000), poor organizational vision (Awalt and Lawton 2000, Hammer 2002), and poor process management (Orr 2000). Several authors have specifically cited the data conversion and implementation as the primary factor in the success or failure of the focused IT strategy (Adelman and Moss 2000, Angus 1999, Bohn 1997, Dyché 1998, Gupta 2002, Moore and Wells 2000, Williams 1997). The individual data mart design and implementation (Figure 4) is often the starting point for planning utility and the ending point for analyzing platform problems (Dyché 1998). Healthcare organization needs, indicated in Figure 2, tend to be discrete data mart issues rather than cohesive data warehousing.
Implications for Health Services Delivery and Policy

It remains clear that healthcare systems face strong competitive pressures that will likely continue for the foreseeable future. The rapid developments in IT that have occurred over the last thirty years may not continue over the next thirty years, but many healthcare organizations have only begun to experience the marvel and facility of the information age. The management of these organizations will remain challenged by internal and external stakeholders (Evans 2000), state and federal regulatory bodies (“Reaping the Rewards” 2002, Whiting 2001) and others (Hammer 2002, Love and Burwen 1999, Orr 2000). Competitive pressures will continue to drive poorer performing organizations toward closure, restructuring or positions of financial uncertainty (Curtright, Stolp-Smith, and Edell 2000, Griffith 2000, Rivenson, Wheeler, Smith and Reiter 2000, Langabeer 1998, Zuckerman 2000). The Health Insurance Portability and Accountability Act (HIPAA) of 1996 has required unprecedented changes in operational behavior and data management to comply with its privacy provisions. All of the issues presented in this paper have profound impact on the delivery of and policies developed for health services. Perhaps the impact that is most unclear is the impact of IT capabilities on business and the subsequent growth of health and political policy questions.

The cumulative effect of competitive pressures and technological development (necessity as the Mother of Invention) will continue to create new business landscapes for healthcare organizations, which will be followed by a construct of beliefs, then policies, over what is right, desired, or permissible. Longest (1998) described health policy as “the collection of authoritative decisions made by government that pertain to health and the pursuit of health” (p.xxi). In the United States these decisions are made through a complex structure of regulation and law making in a theoretically democratic manner. The influence of interest groups and the historical and prevailing trade practices affect the policies that are developed. The ever-turning tide of consumer sentiment and the balancing of political will also play their roles in the development of health policies. These factors will indeed make it difficult to estimate the impact of IT developments in healthcare delivery and policy but they do suggest that issues will continue to challenge healthcare providers, organizations, regulators, and consumers.
References


Figure 1  Components of Data Warehousing (Orr 2000)
Figure 2  Current Practices in Healthcare that Involve Data Warehousing

<table>
<thead>
<tr>
<th>Current Practices in Healthcare that Involve Data Warehousing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar coding</td>
</tr>
<tr>
<td>Business transactions</td>
</tr>
<tr>
<td>Care management</td>
</tr>
<tr>
<td>Clinical data validation</td>
</tr>
<tr>
<td>Clinical practice analysis and guideline development</td>
</tr>
<tr>
<td>Coding schema implementation and architecture design (information engineering)</td>
</tr>
<tr>
<td>Cost management</td>
</tr>
<tr>
<td>Customer service management</td>
</tr>
<tr>
<td>Data auditing</td>
</tr>
<tr>
<td>Data marts</td>
</tr>
<tr>
<td>Data mining</td>
</tr>
<tr>
<td>Database performance and scalability</td>
</tr>
<tr>
<td>Decision support</td>
</tr>
<tr>
<td>e-business</td>
</tr>
<tr>
<td>Enterprise application integration</td>
</tr>
<tr>
<td>Executive information systems</td>
</tr>
<tr>
<td>Graphical visualizing techniques</td>
</tr>
<tr>
<td>Health employer information data set (HEDIS) of the National Committee for Quality Assurance (NCQA)</td>
</tr>
<tr>
<td>Healthcare information and management systems society (HIMSS)</td>
</tr>
<tr>
<td>HIPAA solutions</td>
</tr>
<tr>
<td>Infrastructure management</td>
</tr>
<tr>
<td>Marketing</td>
</tr>
<tr>
<td>Process and workflow re-engineering</td>
</tr>
<tr>
<td>Supply chain</td>
</tr>
</tbody>
</table>
Figure 3  Data Conversion Planning and Implementation (Bohn 1997)
Figure 4  Data Mart Design and Implementation Time Frame  (Dyché 1998)
Appendix 1  Data Warehouse Terminology (Creative Data 2002)

**Bitmapped Indexing:** A family of advanced indexing algorithms that optimize RDBMS query performance by maximizing the search capability of the index per unit of memory and per CPU instruction. Properly implemented, bitmapped indices eliminate all table scans in query and join processing.

**Business Model:** An object-oriented model that captures the kinds of things in a business or a business area and the relationships associated with those things (and sometimes associated business rules, too). Note that a business model exists independently of any data or database. A data warehouse should be designed to match the underlying business models or else no tools will fully unlock the data in the warehouse.

**Corporate Data:** All the databases of the company. This includes legacy systems, old and new transaction systems, general business systems, client/server databases, data warehouses and data marts.

**Data Dictionary:** A collection of Metadata. Many kinds of products in the data warehouse arena use a data dictionary, including database management systems, modeling tools, middleware, and query tools.

**Data Mart:** A subset of a data warehouse that focuses on one or more specific subject areas. The data usually is extracted from the data warehouse and further denormalized and indexed to support intense usage by targeted customers.

**Data Mining:** Techniques for finding patterns and trends in large data sets. See also Data Visualization.

**Data Model:** The road map to the data in a database. This includes the source of tables and columns, the meanings of the keys, and the relationships between the tables.
**Data Visualization:** Techniques for turning data into information by using the high capacity of the human brain to visually recognize patterns and trends. There are many specialized techniques designed to make particular kinds of visualization easy.

**Data Warehouse:** A database built to support information access. Typically a data warehouse is fed from one or more transaction databases. The data needs to be cleaned and restructured to support queries, summaries, and analyses.

**Decision Support:** Data access targeted to provide the information needed by business decision makers. Examples include pricing, purchasing, human resources, management, manufacturing, etc.

**Decision Support System (DSS):** Database(s), warehouse(s), and/or mart(s) in conjunction with reporting and analysis software optimized to support timely business decision making.

**Joint Application Development (JAD):** JAD is a process originally developed for designing a computer-based system. It brings together business area people (users) and IT (Information Technology) professionals in a highly focused workshop. The advantages of JAD include a dramatic shortening of the time it takes to complete a project. It also improves the quality of the final product by focusing on the up-front portion of the development lifecycle, thus reducing the likelihood of errors that are expensive to correct later on.

**Metadata:** Literally, “data about data” More usefully, descriptions of what kind of information is stored where, how it is encoded, how it is related to other information, where it comes from, and how it is related to your business. A hot topic right now is standardizing metadata across products from different vendors.

**Methodology:** The steps followed to guarantee repeatability of success. A good methodology is built on top of real world experience.

**Middleware:** Hardware and software used to connect clients and servers, to move and structure data, and/or to pre-summarize data for use by queries and reports.
Multidimensional Database (MDD): A DBMS optimized to support multidimensional data. The best systems support standard RDBMS functionality and add high-bandwidth support for multidimensional data and queries. Users that need a lot of slices and dices might appreciate a multidimensional database.

Object Oriented Analysis (OOA): A process of abstracting a problem by identifying the kinds of entities in the problem domain, the ‘is-a’ relationships between the kinds (kinds are known as classes, is-a relationships as subtype/supertype, subclass/superclass, or less commonly, specialization/generalization), and the ‘has-a’ relationships between the classes. Also identified for each class are its attributes (e.g. class Person has attribute Hair Color) and its conventional relationships to other classes (e.g. class Order has a relationship Customer to class Customer.)

Object Oriented Design (OOD): A design methodology that uses Object Oriented Analysis to promote object reusability and interface clarity.

OLAP: An acronym for On Line Analytical Processing.

On Line Analytical Processing (OLAP): A common use of a data warehouse that involves real time access and analysis of multidimensional data such as order information.

Performance: Data, summaries, and analyses need to be delivered in a timely fashion. Performance is often a key issue with data warehouses: the right answer isn't worth much if it shows up after the decisions have been made.

Query: A specific atomic request for information from a database.

Rapid Application Development (RAD): Part of a methodology that specifies incremental development with constant feedback from the customers. The point is to keep projects focused on delivering value and to keep clear and open lines of communication. English is not adequate for specification of computer systems, even small ones. RAD overcomes the limitations of language by minimizing the time between concept and implementation.
**Relational On-Line Analytic Processing (ROLAP):** OLAP based on conventional relational databases rather than specialized multidimensional databases.

**Replication:** A standard technique in data warehousing. For performance and reliability several independent copies are often created of each data warehouse. Even data marts can require replication on multiple servers to meet performance and reliability standards.

**Replicator:** Any of a class of product that supports replication. Often these tools use special load and unload database APIs and have scripting languages that support automation.

**Report:** A repeatable, formatted, non-atomic request for information from a database. Usually a report formats and combines several related queries.

**Reporting Strategy:** A top down collection of methodology, products, plans, and teams that ensure business people can get information reliably, accurately, and understandably. It includes choosing tools matched to the organization's particular needs and existing infrastructure, capturing the business models used by the business people, finding source data, integrating all the above into a data warehouse and/or data marts as needed.

**Security:** The right data for the right person. Note that a business analyst may need access to summaries of data s/he should not see. Security systems need to make this easy to implement while making sure outsiders or rogue employees do not see data they should not see.

**Snowflake Schema:** A layering of Star Schema that scales that technique to handle an entire warehouse.

**Star Schema:** A standard technique for designing the summary tables of a data warehouse. "Fact" tables each join to a larger number of independent "dimension" tables. The tables may be partially de-normalized for performance, but most queries will still need to join in one or more of the star tables.