Third International Conference on Construction in the 21st Century (CITC-III)
"Advancing Engineering, Management and Technology"
15-17, September 2005, Athens

Implementing Data Warehousing in the Construction Industry: Opportunities and Challenges

Irtishad Ahmad
Professor and Chair, Department of Construction Management
Florida International University, Miami, Florida, USA

Salman Azhar
Ph.D. Candidate, Department of Civil & Environmental Engineering
Florida International University, Miami, Florida, USA

Abstract
Construction organizations deal with large volumes of project data containing valuable information. These data are normally stored in operational and application databases that are designed to support day-to-day business transactions. While making critical decisions, decision makers would like to quickly analyze existing data to discover trends so that predictions and forecasts can be made with reasonable accuracy. Existing construction databases do not offer such functionality. The concept of data warehousing provides a powerful mechanism to solve the problem of access to appropriate information by decision makers. A data warehouse is a dedicated database created by combining data from multiple databases for purposes of analysis. It supports reorganization, integration, and subsequent utilization of data that enable users to access information quickly and accurately. This paper examines the opportunities and challenges in implementing the data warehousing technique in the construction organizations. The paper first briefly reviews the in-place database management systems in the construction industry. Next, the concept of data warehousing is presented with examples related to the construction industry. At the end, the challenges construction organizations may confront while implementing the data warehousing technique are discussed.

Keywords
Data Warehouse, Information Management, Decision Support Systems, Information Systems, Database Management System

1. Introduction

In construction organizations project data are typically stored in operational and application databases to support construction operations and decisions. The data are often non-validated, non-integrated and stored in a format that makes it harder for decision makers to make quick decisions. This is a typical problem in many business organizations and is caused by an inadequate information technology (IT) infrastructure (Ang and Teo, 2000). Further, for construction firms, these operational databases are likely to be geographically and/or functionally dispersed, thereby making them difficult and inconvenient to access in the relatively short time that is available for decision making. A decision maker may have to wait for days or weeks for responses from IT or Management Information Systems (MIS) personnel that handle requested database queries in order to tap into the data. Such long waiting periods can cause irreversible damage to project performance and reduce the value of the information. Moreover, oftentimes data are not organized properly to support decision making.

The majority of database management systems used in construction organizations are based on the concept of O*n-line Transaction Processing (OLTP)*. OLTP data are updated continually on a regular basis and are suitable to support day-to-day business operations. Such transactional databases are designed to answer *who* and *what* type questions. They are not very good at answering *what-if*, *why*, and *what-next* type questions (Ahmad and Azhar, 2002). The reason is that data in OLTP databases are not necessarily organized to support analytical processing.

For example, payroll data can provide information on wages of regular and overtime labor. Additionally, a scheduling database can provide information on project progress. However, if an estimator or scheduler is interested in knowing "what is the effect of employing overtime labor on project productivity?" OLTP databases cannot provide a direct answer. The executive making this query must rely on subordinates to extract related data from accounting and scheduling databases. These data need to be interpreted, and then computations are needed, in order to get the answers. The whole process could take days to weeks depending on the size and complexity of the project (Yang and Yau, 1996). The issue is to be able to quickly analyze existing data to discover trends so that predictions and forecasts can be made with reasonable accuracy and in time to aid in the decision making process.

The situation described above could have been different if the organization had implemented the concept of data warehousing, which has emerged as one of the most powerful tools for solving the problem of access to appropriate information by decision makers. Data warehousing is based on the *Online Analytical Processing (OLAP)* concept as opposed to OLTP. Basically, this technique supports reorganization, integration, and subsequent utilization of data (from different databases and data-sources) that enable users to access information quickly and accurately. With the appropriate user-friendly query tools, users can experiment with different views of the data, thus gaining a better understanding of the situation in order to make better decisions.

The data warehousing technique has been successfully implemented in many industries. For example, in manufacturing for order shipment and customer support, in retail for user profiling and inventory management, in financial services for claims analysis, risk analysis, credit card analysis, and fraud detection. It is also used in transportation for fleet management, in telecommunications for call analysis and fraud detection, in utilities for power usage analysis, and in healthcare for outcomes analysis (Adrians and Zantinge, 1996).

The purpose of this paper is to examine the value of implementing data warehousing techniques in the construction industry. If properly implemented, data warehousing could facilitate coordination and promote greater interaction and responsiveness in the construction process by generating trend analysis, discovering patterns, and aiding executive-level decision-making. This paper will briefly discuss all these aspects.

2. Database Management Systems in Construction

The use of database management systems (DBMS) in civil engineering and construction is not new. Like organizations in the service and manufacturing industries, the construction industry benefited from the advancements that took place in the DBMS technology. Earlier systems were developed for stand-alone applications such as the CAD systems, estimating, scheduling and inventory control (Chen et. al., 2004). Efforts were made in the early 1980's to integrate various construction-related databases. Much of the early integration efforts revolved around the idea of integrating 2D drafting with 3D modeling, integration of graphical and non-graphical (or computational) design information and integration of two or more applications, such as analysis and design (Amor and Anumba, 1999). In the 1990's, emphasis was placed on process integration (i.e. design and construction) and multidisciplinary integration (such as integrating various construction - related departments). During that time, different interfaces were developed to integrate CAD systems with project planning, scheduling, cost estimating and cost control tools and methods. Examples of such efforts, available in the literature, include COMBINE, CIMSTEEL, ATLAS, and ICON among others (Bjork, 1999). Another important development was the formulation of standards for data exchange such as STEP (Standard for the Exchange of Product Data), and IFCs (Industry Foundation Classes).

Recently, with advances in computer software and hardware technology and due to the exponential growth of Internet-based communication technology, the trend has shifted towards the development of web based project management systems and Enterprise Resource Planning systems (ERP). In the following paragraphs we briefly describe some of the recent research efforts on these topics around the world.

The AIC (Automation and Integration in Construction) Research Group at the University of Salford, UK, has developed a web based integrated environment prototype called SPACE (**S**imultaneous **P**rototyping for **A**n integrated **C**onstruction **E**nvironment). SPACE provides users with a multi-disciplinary computer environment where project information can be shared between the various construction professionals using integrated databases (SPACE, 2004). CIMS (**C**omputer **I**ntegrated **M**anagement **S**ystem) developed at the Hong Kong Polytechnic University provides an integrated information control, material control, cost control, progress control, human resources management and tools for quality assurance (CIMS, 2004).

Enterprise Resource Planning (ERP) is another area that is now being hailed as a foundation for the integration of organization-wide information systems. A recent study by ML Payton Consultants (ML Payton Consultants, 2004) indicates that more than half of the large US and European construction companies have some sort of ERP systems (customized/in-house or external/ commercial) in place. The extent of implementation, however, varies and most are not very effectively implemented or utilized. At least half of these companies complain that commercially available ERP systems are hard to customize for an individual company's operations and this is one of the main reasons behind their slow adoption or ineffective utilization. The top ERP systems for construction include One World by JD Edwards, SAP and BAAN.

The commercial software vendors are also making prominent progress towards the development of integrated project management databases. One such example is Prolog Manager™ developed by the Meridian Project Systems Inc. Prolog Manager automates everyday management functions and provides features to track the project data from design to close out. Built on the Microsoft ® SQL Server ™ 7.0 database platform, Prolog Manager enables multi-project control of procurement, cost control, document management, collaboration, and field management within a single application (Prolog Manager, 2004).

Various applications and efforts in database management systems mentioned above have the objective of providing accurate information about an ongoing or completed project activity. However, most of them are not designed to, or are limited in their ability to, generate trend analysis, discover patterns, and aid executive-level decision-making (Chau et. al., 2003). Hence executives would still not be able to make critical decisions based on the success and failure of previous projects unless some specially prepared data are is provided to them. OLTP based IS (information systems) departments would typically spend excessive amounts of time, money and effort to provide the crucial data needed to support such decision-making activities. For effective decision-making, there is a need to develop integrated database management systems that can organize data from different projects over a period of time and can provide answers to analytical queries. The recently developed data warehousing technique is a response to the limitations of OLTP based database management systems.

3. The Data Warehousing Concept

The traditional view of database management is based on the need for data to support transaction processing. Conversely, data warehousing evolved as an answer to the need to support analytical processing for informed decision-making. The primary purpose of a data warehouse is to provide easy access to specially arranged data that can be used with decision support applications, such as management reporting, queries, data mining and executive information systems.

In general terms, a data warehouse is a dedicated database created by combining data from multiple databases for purposes of analysis. It is generally populated with data from two sources. The most frequent source is the periodic migration of data from Online Transaction Processing (OLTP) systems such as financial and accounting, scheduling, cost, human resources and so on. The second source is externally purchased databases (such as lists of incomes and demographic information, and in the context of construction, material pricing data) that can be linked to internal data. A data warehouse collects all data into one system, organizes the data for consistency and easy interpretation, keeps "old" data for historical analysis, and makes access to, and use of data a simple task so that users can do it themselves without great technical proficiency in data handling (Gray and Watson, 1998).

Typically, data warehouse architecture has three components or tiers, as shown in Figure 1:

1. Data acquisition tools (back end) that extract data from transactional databases (i.e. OLTP systems) and external sources, consolidate and summarize the data, and load it into the data warehouse.
2. Associated software for managing the data. (Maintained with the data in the warehouse itself)
3. The client (front end) software that enables users to access and analyze data in the warehouse.

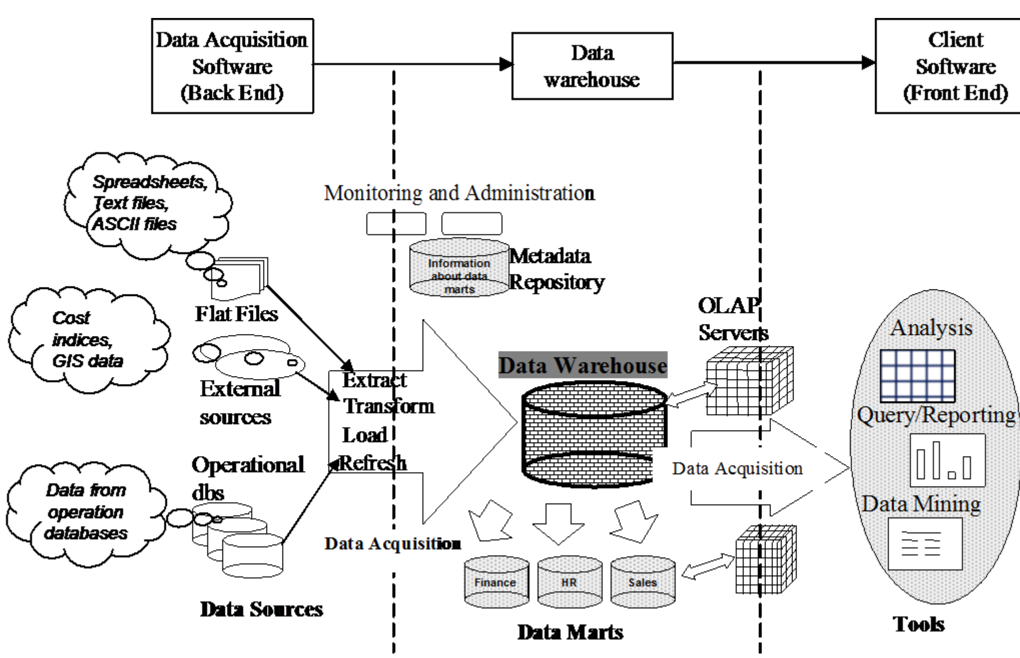

Figure 1: A Generic 3-Tier Data Warehouse Architecture
(Adapted from Chaudhuri and Dayal, 1997)

4. Data Warehousing in the Construction Industry

There are three recent instances of the application of data warehousing in the construction industry. In one a Chicago based construction company with offices located in nine states developed a data warehouse of its accounting data. The purpose was to support the decision makers for making critical price forecasting decisions and analyzing the company's profits and losses in different activities over a period of time. The system provides quick answers to executives for all types of forecasting and trend-type queries and generates ad hoc reports whenever they are needed. Using a web interface, the company personnel have presentation-quality access to information at all hours. The system has resulted in a substantial savings in preparing cost estimates for future projects and enhanced the productivity of the decision makers (nSpin, 2004).

In the second instance, a Construction Management Decision Support System (CMDSS) using data warehousing was developed at the Hong Kong Polytechnic University for materials inventory control (Chau et. al., 2003). The system was designed for both novice and experienced users and was implemented in the Hong Kong Polytechnic University Student Dormitory construction project. The system could provide answers to queries such as "Determine the total amount of ceramic materials stored in the warehouses from suppliers in Beijing?" Using such queries, the construction managers could easily determine the inventory trend of the materials and the amount of each material type. It helped the managers in formulating an appropriate inventory decision or a warehouse storage strategy.

Very recently, Zhiliang et al. (2005) developed a system named EXPLYZER which can extract useful information from the accumulated documents through the use of data warehousing technique and then analyze it. The system adopted a data standard based on the XML (extensible markup language) format to extract information from documents generated by a Web-based project management system. The system can search across different kinds of documents to extract relevant information and perform various types of analysis such as the Earned value analysis, Productivity analysis etc based on the users' requirements.

The following sections include three scenarios that illustrate how the concept of data warehousing can benefit the decision makers in construction organizations, particularly the owners.

4.1. Scenario I: Productivity Information

Reliable productivity information is essential for the effective planning and management of construction projects. Most often information remains hidden within the past project data. Depending on the type of contracts, owners would have these data in one of the two forms. (1) In lump-sum contracts, productivity data will be contained in periodic progress payment requests and schedules submitted with them. These documents typically contain information about how much physical work has been done and how much time has been spent. (2) In cost-plus contracts, owners will typically have more detailed information about productivity in the form of payroll data derived from daily/weekly time-cards. These data can be reconstructed (integrated) from existing transactional data sources, de-normalized (a process of integration from normalized data, typically used for transaction processing) and then can be loaded in the data warehouse for subsequent use by the decision makers. This concept is demonstrated in Figure 2.

4.2. Scenario II: Conceptual Estimates

Another area where data warehousing can improve effectiveness is the estimating process, especially the screening or conceptual estimates developed at the planning stage of the project for various alternatives. A data warehouse based on previous project data can be used to develop parametric estimates by selecting the most appropriate parameters. Typical examples of the parameters are physical measures of size and capacity of the facilities. Thus a data warehouse can be very useful in providing information to develop semi-detailed and, with more effort even, detailed estimates particularly at the planning stage of a the project. In addition, during the process of developing estimates, queries can be run and sensitivity analyses can be performed to see how various parameters affect project cost, schedule, and productivity.

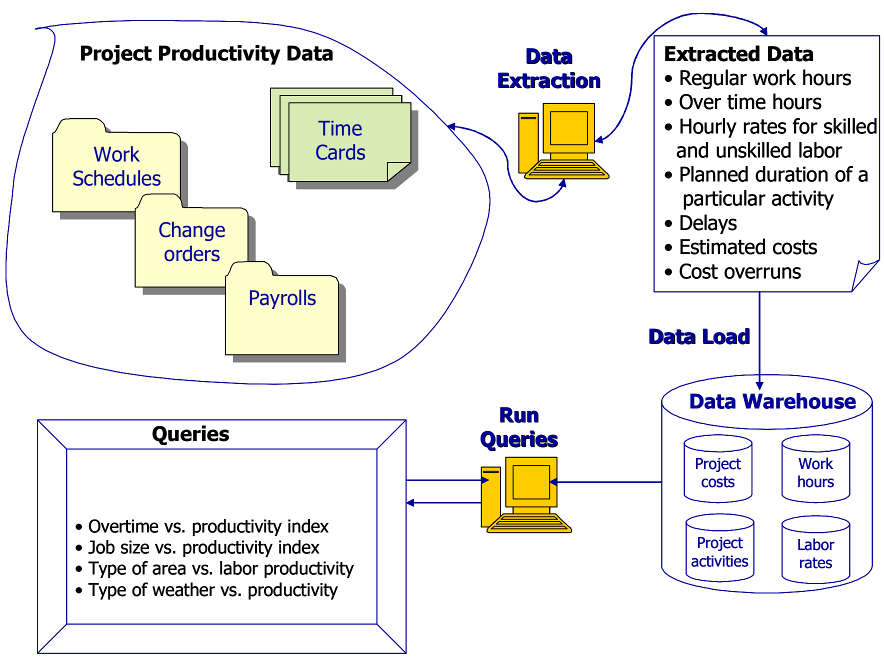

Figure 2: Conceptual Framework of a Data Warehousing System to Provide Information about Productivity Analysis

4.3. Scenario III: Site Selection for Residential and Commercial Development

The Site selection process depends on a number of spatial and business-related factors making it a complex decision-making task. It is common for the decision makers to use their subjective judgment and gut feelings based on their experience in selecting the most appropriate site for development. The reason is that data for site selection originate from varied sources and are not organized in a format that decision makers can readily use to derive any meaningful information. One possible solution of this problem is to develop a decision support system (DSS) using data warehousing technique to help retrieve data from different databases and information sources and analyze them in order to provide useful and explicit information. Based on this concept, the authors developed a DSS to assist builders/developers in site selection for residential housing development. The following paragraph will illustrate the conceptual model of this system. The system development details can be found in an article by Ahmad et. al. (2004).

A conceptual model of the developed DSS for site selection is shown in Figure 3. The DSS utilizes GIS application and employs the data warehousing technique in order to narrow a vast list of available sites for sale down to a manageable short list of a few technically feasible sites. Subsequently, an analytical processing technique is used to rank order the candidate sites for selecting the most appropriate one. The DSS can be used for both short-term and long-term decision making. Short-term decision making is referred to reviewing and analyzing available information for a particular project. While long-term decision making involves analyzing trends and discovering patterns of occurrences of certain events over several years. For example, generating lands sale trends in different types of urban and rural areas. These trends can aid the executives in making effective decisions regarding site selection.

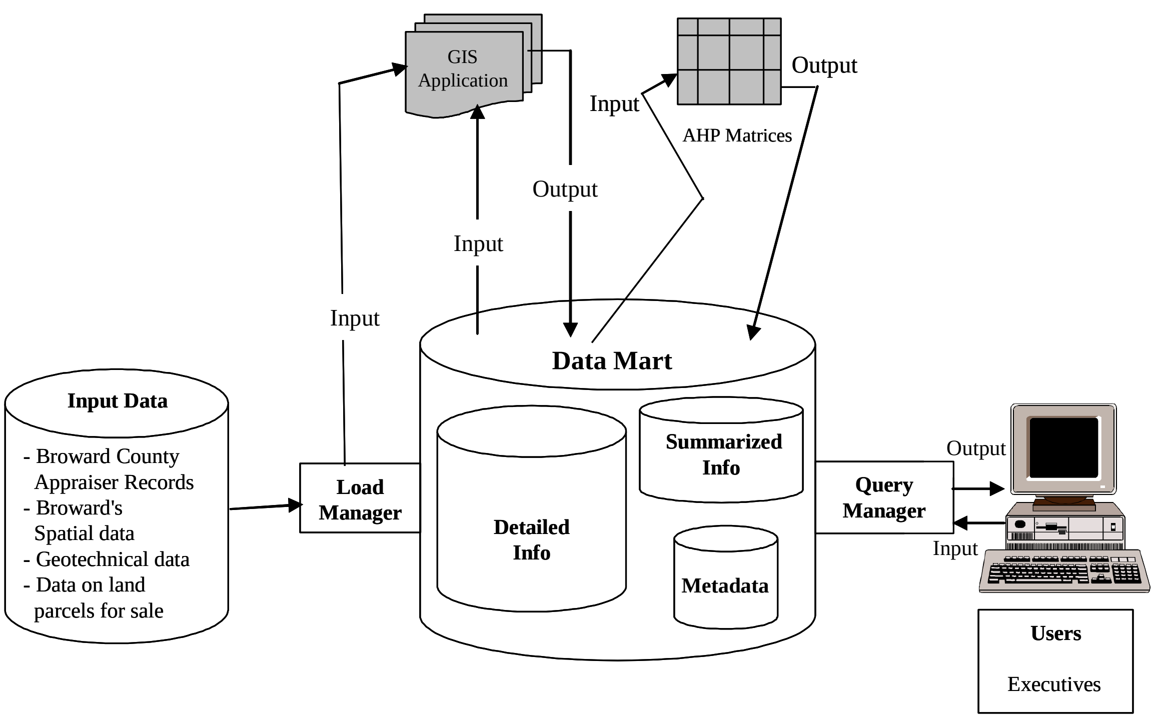

Figure 3: Conceptual Model of the DSS for Site Selection (Ahmad et. al., 2004)

5. Challenges in Implementing Data Warehousing in Construction

Construction is an industry bound by traditions, not necessarily by choice, but because of the ways organizations are set up, and have worked, over the years and because of their dependence on age-old norms and rules. Implementation of concepts like data warehousing is challenging in construction, as it would have an impact on these set-ups, norms and rules. In construction organizations, executive decisions are mostly based on experience and gut feelings. By implementing data warehousing, the critical decisions could be made using available information (such as trend analysis, summarized and detailed reports, etc). It could help junior managers to make decisions based on the facts rather waiting for their superiors to make decisions based on their experience. This change could help to convert the construction industry from a rigid vertical structure to a flexible horizontal one that provides better coordination more control.

From a practical point of view, data warehousing implementation in a construction organization poses huge problems. These problems are, organizational, financial (a substantial investment is required), and technical (IT experts are needed to build or maintain data warehouse). The cost of a typical data warehouse ranges from $500,000 to $2 million and its implementation time is from 1-3 years. This investment in time and money requires strong commitment from top level management. The data warehouse can be created in-house, outsourced in part, or out-sourced completely. The favored approach is to keep the warehouse project in house, with the IT department responsible for construction; however outside consultants may be hired to provide expertise in specialized areas (Gray and Watson, 1998).

Although data warehousing could be implemented for any group of construction organizations (such as owners, consultants, contractors etc.), the authors suggest initial implementation in the owner's organizations for two reasons: (1) Owners have a wider perspective as they are involved with the project from inception to completion, and in most cases, are the ultimate users of the constructed facilities. (2) Owners are in control of project funds. The extent of this control, however, depends on the type of contractual arrangements, but owners are the ultimate stakeholders, as far as the overall investment in a project is concerned. Obviously, we are not considering one-time construction clients as owners. We are focusing only on the type of owners that are continuously involved in construction projects. This type includes public agencies, such as the state department, county/city and state governments, school boards, etc.

Moreover, the fragmented nature of the construction industry coupled with the sequential nature of project tasks makes it almost impossible for any organization other than the owner's to be the primary manager of the data warehouse. Given the adversarial relationships that exist in most traditional project delivery systems, the owner will be in the best position to effectively and efficiently manage and control the data query, retrieval and analysis for enhanced and informed decision making by all parties concerned. Also, since it is the owner who initiates and provides the capital for most construction projects, it will only make sense if that entity can claim the ultimate ownership of the data warehouse.

Recently, the authors conducted a research study to conceptually develop a data warehousing system for a public owner organization in the State of Florida. The estimated cost of the system is around $1 million with an implementation time of 45-55 weeks. It is expected that the initial development costs will be recovered in a period of 3-4 years. Besides improvements in data management and decision-making practices, the system is expected to result in up to 20% savings in direct salary costs and another 20-30% savings in operational costs. The details about this research project will appear in a separate paper.

6. Overall Benefits of Implementing Data Warehousing in Construction

Successful implementation of data warehousing in construction organizations will help decision makers to make more effective decisions, which will improve the productivity of the industry in general. Decision makers will be less dependent on information technologists or database managers to make schema changes in order to extract information. Perhaps more importantly, data warehouse capabilities would enable managers to model problems that would be impossible through less flexible systems with lengthy and inconsistent response times.

7. Conclusions

A data warehouse is an informational database, specifically designed to support strategic planning and decision making. Data warehousing technology can enable construction companies to consolidate information from diverse operational systems into one source for consistent and reliable information. This will give construction project managers the opportunity to incorporate wisdom and insight into their decision making process.

With successful implementation of the data warehousing technique in construction organizations, managers will be less dependent on IT professionals, and will be using data more effectively. However, data warehousing developmental efforts must be supported by an organization's top executives. The decision to invest in data warehousing must be made carefully - it is not always the most cost-effective option, it may be advisable first to build a summarized reporting structure using operational data before investing in a full-blown data warehouse. These structures can eventually be ported to the data warehouse of the future. Most of the benefits are long-term and far less tangible.

8. References

Adriaans, P. and Zantinge, D. (1996). *Data Mining*, Addison Wesley Longman Limited, Reading, Massachusetts.

Ahmad, I., and Azhar, S. (2002). "Data Warehousing in Construction: From Conception to Application". *Proceedings of the First International Conference on Construction in the 21st Century: Challenges and Opportunities in Management and Technology"*, April 24-26, 2002, Miami, Florida, USA, pp. 739-747.

Ahmad, I., Azhar, S. and Lukauskis, P. (2004). "Development of a Decision Support System using Data Warehousing to Assist Builders/Developers in Site Selection". *Automation in Construction,* Vol. 13, No. 4, pp 525-542.

Amor, R., and Anumba, C.J. (1999). "A Survey and Analysis of Integrated Project Databases". *Proceedings of the International Conference on Concurrent Engineering in Construction: Challenges for the New Millenium*, CIB Publication 236, Espoo, Finland, August 25-27, pp. 217-227.

Ang. J., and Teo, T.S.H. (2000). "Management Issues in Data Warehousing: Insight from the Housing and Development Board". *Decision Support Systems*, Vol. 29, pp. 11-20.

Bjork, B.C. (1999). "Information Technology in Construction: Domain Definition and Research Issues", *International Journal of Computer-Integrated Design and Construction,* Vol. 1, No. 1, pp. 3-16.

Chau, K.W., Cao, Y., Anson, M., and Zhang, J. (2003). Application of Data Warehouse and Decision Support System in Construction Management. *Automation in Construction* Vol. 12, No. 2, pp. 213-224.

Chaudhuri, S., and Dayal, U. (1997). "OLAP Technology and Data Warehousing" *ACM SIGMOD RECORDS*, Heward Packard Labs.

Chen, C., Udo-Inyang, P., and Schmitt, F. (2004). "Integration of a Database Management System and a Knowledge-Based Expert System in Construction: A Review", *On-line* http://asceditor.unl.edu/archives/1994/chen94.htm, Accessed on June 13, 2004.

CIMS. (2004). "Computer Integrated Management System". Online at http://www.bre.polyu.edu.hk. (06/12/2004).

Gray, P.G., and Watson, H.J. (1998). *Decision Support in the Data Warehouse*, Prentice Hall, NJ.

ML Payton Consultants. (2004). "Use of Enterprise Resource Planning in the Construction Industry: Summary of Findings," *http://www.mlpayton.com/pages/summaries.html* (06/12/2004).

nSpin Case Study. (2004). "Flexible Access to Enterprise Data and Decision Support". Online at http://www.nspin.com/industries/cases.html, (06/21/2004)

Prolog Manager, "Prolog Manager: Power Application for Total Project Control". Online at *http://www.mps.com/products/PM/index.asp*, (06/12/2004).

SPACE. (2004). "Simultaneous Prototyping for An integrated Construction Environment". Online at *http://www.surveying.salford.ac.uk/aic/space.htm*. (06/12/2004).

Yang, J.B., and Yau, N.J. (1996). "Application of Cased-based Reasoning in Construction Engineering and Management". *Proceedings of the Third Congress held in Conjunction with A/E/C Systems,* Computing in Civil Engineering, American Society of Civil Engineers, New York, pp. 663-669.

Zhiliang, M., Wong, K.D., Heng, L., Jun, J. (2005). "Utilizing Exchanged Documents in Construction Projects for Decision Support Based on Data Warehousing Technique". *Automation in Construction*, Vol. 14, No. 3, 405-412.