INTEGRATING ENTERPRISE DATA FOR DECISION SUPPORT IN CONSTRUCTION ORGANISATIONS

Tanko Ishaya, James Chadband
The University of Hull, Scarborough Campus, Filey Road, Scarborough, YO11 3AZ, United Kingdom

Lucy Grierson
C Spencer ltd, Mill Lane, Barrow on Humber, North Lincolnshire, DN19 7BD, United Kingdom

Keywords: Decision Support, Data Warehouse, Data Warehousing, Knowledge Discovery, Decision-Making, Data Mining, and Information Integration.

Abstract: Information integration is one of the main problems to be addressed when designing a data warehouse for decision-making support. Possible inconsistencies and redundancies between data residing at the operational data sources needs to be resolved before migrating to a data warehouse, so that the data warehouse is able to provide an integrated and reconciled view of data within the organisation. This paper presents a performance-oriented data warehouse as an integrated data for decision-making support within a construction organisation. The process is based on a conceptual representation of the enterprise, which has been exploited both in the data integration phase of the warehouse information sources and during the decision-making activity from the information stored in the data warehouse. The application of the process has been supported by prototype software tools for data integration, reconciliation and analysis that provides some decision-making support.

1 INTRODUCTION

Construction organisations generate a great amount of operational data about various aspects of their business, such as about customers, contracts, products, operations and employees. These data are normally distributed across various functional systems that support every day operations and decision-making. The data also play an important role in ensuring that those contracts are adequately prized, completed on time within budget and meeting design specifications as highlighted by Ahmed (2000).

Organisational information systems are generally classified into two broad categories: systems that support everyday operations and systems that support decision-making, coined as Operational Support Systems (OSS) and Decision Support Systems (DSS) respectively (Gonzalez, 2005). DSS provides and uses historical information for analysing the business in order to improve the quality of decision-making process. OSS provides and uses data from day-to-day organisational transactions and serves the operational purpose of an organisation. OSS data is physically different from that serving informational and analytic need. The user community, supporting technology and processing characteristics for the operational environment are fundamentally different from informational environment (Jones, 1998). Thus, any successful DSS must have informational data and information integration is one of the main problems to be addressed when designing informational data. These data are mostly modelled in a data warehouse. As popularised by Inmon (2002), a data warehouse is a "subject-oriented, integrated, time-variant, and non-volatile collection of data in support of management's decision-making process. A data warehouse forms the core of a DSS as justified by Adhikari (1996) and Ahmed (2000). It provides support for all levels of management decision-making process through the extraction, transformation and interpretation of internal and external data (March and Hevner, 2005).

Ishaya T., Chadband J. and Grierson L. (2007).
INTEGRATING ENTERPRISE DATA FOR DECISION SUPPORT IN CONSTRUCTION ORGANISATIONS.
In *Proceedings of the Ninth International Conference on Enterprise Information Systems - DISI*, pages 534-539
DOI: 10.5220/0002376705340539

Data warehousing is a process for integrating data from a variety of data sources into a data warehouse that could be used for information analysis and knowledge discovery required to support effective and efficient decision-making process of an organisation. Information obtained from data warehouses enables organisations to examine how successful the business was in the past and its current position. The sum of the past and present information provides decision makers with a better knowledge about the business for strategic decision-making (Gonzalez, 2005).

Despite the popularity of data warehouses in manufacturing and other business sectors, studies and implementations still seem to be very limited in the construction industry – especially within small-to-medium scale organisations. Other areas that have utilized data warehouses have been from the operational perspective. In order to take full advantage of the technology, more research is needed on how to collect and store company wide construction data and how to develop a data warehouse for assisting construction business in the process of decision-making.

The objective of this paper is to report our experience in the development of an enterprise data warehouse for a construction organisation. This work is part of a UK Government funded project in collaboration with a construction company to research, design and develop a central repository (data warehouse) for the company data and to develop analytical tools in order to improve the company's management decision-making process. The paper mainly presents the development process of the data warehouse currently being used for information analysis and a basis for data mining as highlighted in the further work section of the paper. Our approach has been business led and a performance-oriented data warehouse architecture was developed based on organisational requirements. The developed data warehouse is currently being used to support certain aspects of decision-making within the construction organisation. This approach can equally apply to similar construction organisations.

The reminder of the paper is organised as follows: In Section 2, we describe the enterprise information system in construction organisations including a specific analysis of the different data sources and their inter-relationships within the company. In Section 3, we present the development process using a developed performance-oriented data warehouse architecture and the information integration process, and how the Data Warehouse is now being used for information analysis is briefly discussed in Section 4. Finally, we draw some intermediate conclusions with a description of further work in Section 6.

2 ENTERPRISE INFORMATION SYSTEMS IN CONSTRUCTION ORGANISATIONS

Organisations within the construction industry have always tended to be based on an "*every man for himself*" (Macomber, 2004) philosophy of running a business. This type of philosophy may not always in the best for the company, and generally means that the data within the company is always different from organisation to organisation and also organisations within the construction industry are often slower to accept change, especially in the field of Information Technology.

The current industry practice shows that many local databases are maintained by different offices in a construction company to support its management functions. The specific construction organisation referred to in this paper is a Civil Engineering & General Construction company. They operate at a national level with a particular emphasis on the development, refurbishment and maintenance of contracted locations. As well as working nationally, they operate a multi office and multi site business, which incorporates the activities and requirements of a diverse range of staff and partners including contractors, consultants and customers. Like many medium to large organisations, the company operates with a variety of data sources. The interactions between these data sources are presented in Figure 1.

Following is a summary of the different types of data sources existing within the organisation.

2.1 Relational Data Sources

The various relational data sources with the company are stored in a variety of different locations using different relational management systems. These data sources include accidents, accounts, customer enquiries, construction and drawing documents, insurance databases, etc. Some databases have lots of repeated information and are not normalized, others have no repeated information and are normalized up to the third normal form. Obviously access to these databases depends on user access rights and privileges.

After a thorough review and analysis of the various databases from a centralisation point of view, it was discovered that some databases needed to be maintained. Access databases are very detailed with lots of information in them. Other databases are quite sparse and contain information that would be required to address future business needs.

2.2 Internet Data Sources

The company also have an Intranet site with a large repository of documents stored online, these file are then available to anyone who has access to the intranet. The company also uses e-mail as a form of communication, along with more traditional forms of communication. Regulations are in place for the use of e-mail for business purposes.

The intranet contains a lot of data including project information, meeting minutes, cost code information, subcontractor performance reviews, design process reports, subcontractor approval statuses, drawings and customer feedback. Most of these information are either files being uploaded or reports that have been produced using data from the relevant databases and can be used by users with appropriate access privileges.

The intranet is a very useful resource and contains a wide range of documents and information that could be used to enhance the decision making process for the organisation.

2.3 Non-digital Data Sources

There are various non-electronic data that are kept in a paper based format. These are filed and are kept in various offices on the site. The types of data that is kept ranges from contract documents to office overhead documents.

Each Contract within the company has one or more files, each contract file has a pre-defined number of sections and each section has a particular purpose, for example section 1 contains all the correspondence from a client and section 8 contains all the correspondence from the subcontractors. These files can be very large and can therefore be spread across multiple folders.

There are also human sources of data, information and knowledge, which is obviously very difficult to tap into at the moment.

2.4 Interactions between the Main Data Sources

The interactions and relationships between and within the various data sources of the company were analysed to produce figure1 below.

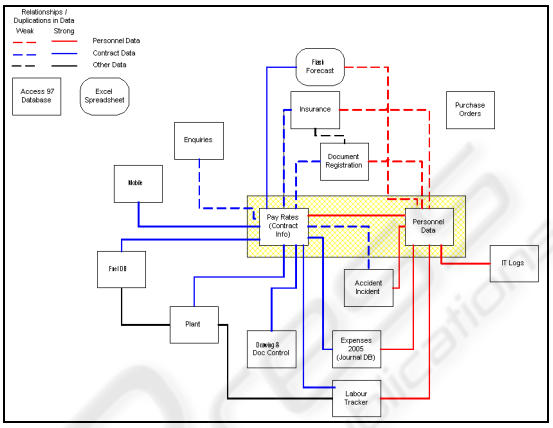

Figure 1: Interactions between the main data sources.

The analysis of each of the data sources shown in Figure 1 shows a clear cross-functional nature of decision support requirement, which demands significant integration efforts as well as the technical challenges it highlights.

These operational systems and data sources are segregated along the physical boundaries between management offices, which cause a series of problems for the construction business including, but not limited to the following:

- difficulty in sharing of data between functions
- multiple entries for the same data are generated for various operational systems
- too much time and resources taken to analyse data for a particular purpose
- slows down the decision making process, since data obtained from different sources may conflict with one another
- difficulty in maintaining data consistency and integrity
- maintenance across organisational units
- too much time and resources taken to analyse data for a particular purpose

The wide-range of critical information requirements for business decision-making and their diverse sources (both internal and external) to the organisation, presents a clear challenge to the development of a data warehouse and decision support tools that utilises them. Next section presents our approach in addressing some of these challenges.

3 DATA WAREHOUSE DEVELOPMENT PROCESS

Data Integration is a key part of many organisational activities. This can be a very time consuming and an expensive endeavour, so it is critical that it is well analysed and designed. The development of the data warehouse was business lead, which is performance-oriented that is inline with the company's requirement for monitoring team, project and financial performance. A data warehouse provides an architectural model for the flow of data from operational systems already in place for decision-support (Decker, et al., 1997; Levi and Arsanjani, 2002). The data warehouse therefore organises data in context of the associated requirements and processes (Matthes, et al., 2005), based on which we developed a performance-oriented architecture (as shown in figure 2) consisting of four main components: data sources, data staging, information storage, and information access.

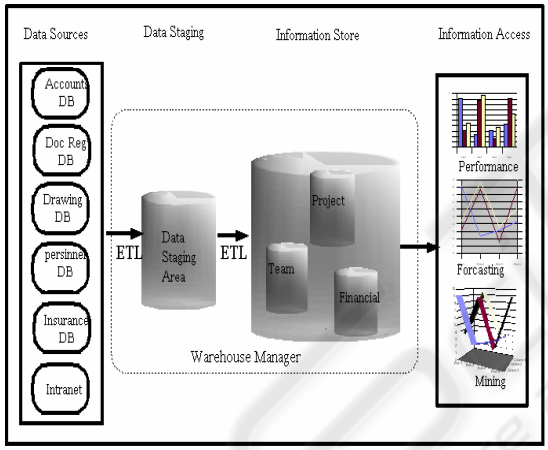

Figure 2: Performance-oriented data warehouse architecture.

The data source components include sources from not only the company's operation databases, but also the relevant data on its intranet – i.e. all the available digital data sources. One of the first challenges here has been access and determining what data to upload to the warehouse. Two approaches were employed in order to address these challenges. These are availability-based and need-based approaches. The availability-based approach examined the data that was currently available in operational systems and the available data was selected to the data warehouse. The need-based approach examined the data that would be needed for decision-making support based on the business requirement.

Recognising the cross-functional and cross-organisational nature of decision support requirements, various disciplined approaches including the General System Theory (Checkland, 1999), Joint Application Design (JAD) were used in order to understand both the operational and strategic vision of the organisation. The business-driven approach focuses at the initial stage of the analysis of the software system within which a goal-oriented model of a business is created and lead to the development of the above architecture. The following requirements were identified during the requirement capture stage:

- The system shall contain features that allows the executives to monitor the general performance of the company
- Monitoring of the performance of the teams.
- Monitor the performance of individual members of a team.
- Monitor financial performance
- Forecast trends in order to explore new business opportunities

However, while, some data loaded into the data warehouse may not have immediate use in relation to the need-based business requirements, we felt that this data may be useful in the future, since it may be easier and cheaper to store the data in to the information store than work to collect the data at a later stage. Section 3.1 and Section 3.2 presents a detailed description of the data warehouse modelling and implementation process.

3.1 Data Warehouse Modelling

A data warehouse is simply a relational database using multidimensional (star schema) modelling techniques that organises data from a view of multi-dimensions as against the traditional two-dimension data base modelling process. Dimensions are the perspectives or entities with respect to the purpose of keeping the records. These can exist in the form of Start schema, snowflake schema or fact constellation schema (Kimball and Merz, 2000). Core to multidimensional modelling is the identification of fact table(s) that determines the numerical measurements of the business. It contains the facts as well as keys to each of the related dimension tables (Calvanese, et al., 2006; Jarke, et al., 1999). As shown in Figure 2, different business grains were identified and lead to the development of three meta data stores. Figure 3 presents one of the three models – team performance dimensional model.

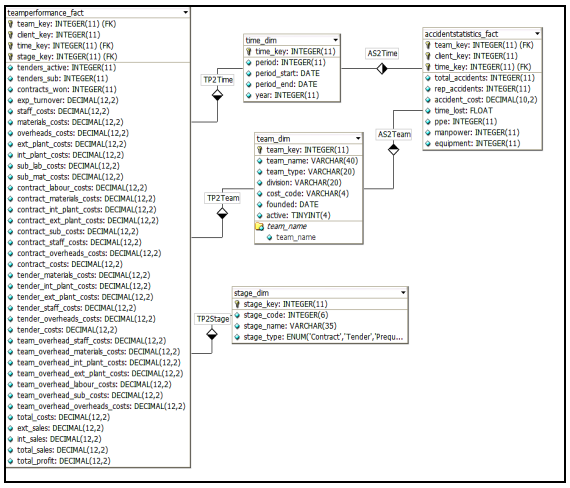

Figure 3: Team-performance dimensional model.

3.2 The Data Warehouse Manager

The warehouse manager consist of tools for extraction, transformation, loading and the warehouse itself. The warehouse was implemented using the MySQL database server. The MySQL server was chosen for a number of reasons firstly it was necessary that a database server had the appropriate technical features. These features would ensure that the data warehouse would operate with adequate performance. One of the most important features was the need to create and execute stored procedures which allow a number of SQL statements to be executed in a single command.

MySQL was also chosen as it fits well in the IT culture of the construction company. The IT culture within the organisation is such that it uses wherever possible software that runs on Linux based operating systems and wherever possible is created on an open-source license (typically LGPL, ASF or BSD style licenses). Further to this the organisation has some experience using the MySQL server.

The warehouse manager contains, along with the data warehouse a number of tools for extracting the data from the operational data sources such as those shown in Figure 2 and transforms them ready to be inserted into the data warehouse. The ETL components have been written in the Java programming language, again there were a number of reasons for the choice but the main reasons were; Java is a platform independent language that can be written and executed on a variety of platforms, Java's database connectivity allows pure Java (platform-independent) access to a wide variety of database servers. Java also has support for XML based information, such as web pages (written in HTML) and text documents.

The first step of the ETL application is to connect to the operational data sources and extract certain columns and tuples as required by the data warehouse model and place them in the data staging area, which may be a temporary database, memory or some other location. Next, certain aspects of the data may be cleansed, transformed or translated into another format or style, examples of which include the conversion of contract numbers between data sources, the filtering of client information and translation of team and personnel information.

Once the data has undergone all necessary transformations, it is then integrated using the following process; firstly the dimensional tables are populated with the relevant data, such as time, team and contract data, next the various fact tables are populated. The fact tables contain the key business information such as the amount that has been spent on a project over a period of time by a particular team. Section 4 presents a description of how the data warehouse has been used for information and knowledge discovery and its application for decision-making support.

4 INFORMATION ACCESS AND DECISION SUPPORT

The ability to access information from the data warehouse is the primary determinant of success from user perspective. Information access tools that are now being used for a certain level of decision-making support. Sections 4.1 and 4.2 presents a description of the tools and how the decision-making support that they provide.

4.1 Information Access Tools

Constructing a centralised data warehouse has made a large number of decision support applications available. The data warehouse has provided the facility for management of the organisation to investigate the amount that has been spent on a project over its whole life or just at specific stages. It has allowed the management to look at varying time frames and also specific teams and divisions within the organisation. The data warehouse system has also provided access to personnel and health and safety records and using all of these data sources in a single system can be of greater benefit than each of the separate database sources. With the data

warehouse system, even a simple query using SQL will allow the users to obtain information.

The system has also produced dynamically generated charts and figures that have been accessed via a web browser, and although currently has only been used internally on the company's intranet, there is no reason that this could not be accessed externally in the future.

4.2 Decision Support

The implementation of this system has provided a number of decisions made within the organisation to be improved, primarily the senior management team of the organisation can customise the level of detail in which the information is returned to them, as a result they can look at a company's wider view, then break it down where necessary and look at particular areas. This may result in the senior management noticing some discrepancies at a particular time and they can drill down into the data to discover the cause. This is very useful if there is a particularly high cost for a project, high number of accidents or trends in the numbers of staff leaving. The management can investigate any of these problems and will be able to determine the cause, if necessary change company procedures or retrain staff, etc.

Certain safety statistics have also been of great benefit in the decision making process, it is now possible to link any injuries to costs in a much easier way than before and of course it is even possible if it is requested to link staff to incidents on site. Making use of the charts mentioned in 4.1 it is much easier for a member of management to see trends and patterns in data, this is particularly useful for when a quick analysis is required as it ensures the user does not have to spend time number-crunching before the information is obtained and the decision is made.

5 CONCLUSIONS AND FURTHER WORK

Integrating operational data in historical form can assist managers in answering questions about the business – its performance, business trends, and what could be done to improve general performance and stay at a competitive advantage. This paper has presented how performance-oriented approach to data warehouse development provides a well integrated data and how it is currently being used to provide support for decision-making within a construction organisation. Decision support systems tend to focus more on detailed information and are targeted towards mid-level managers. The integrated data has provided the basis for data mining as further work in order to provide a higher level of consolidation and a multi-dimensional view of the data, as high level executives require the ability to slice and dice the same data and also to drill down to review details of specific data and information.

REFERENCES

Adhikari, R. (1996). Migrating legacy data, *Software Magazine* 16(1), 1996, pp75-80

Ahmed, I. (2000). Data warehousing in construction organizations, Construction Congress, VI 2000 In *Proceeding, ASCE, 2000*, pp.194-203.

Calvanese, D., Dragone, L., Nardi, D., Rosati, R., and Trisolini, S, M. (2006). Enterprise Modelling and Data Warehouseing in TELECOM ITALIA, In *Journal of Information Systems* 31(2006), pp 1-32

Checkland, P. (1999). *Systems Thinking Practice*, John Wiley & Sons, New York, 1999

Decker, K., Oaks, A., and Salinas, M. (1997). Building a Cost Engineering Data Warehouse, In *AACE International Transactions*, Morgantown, 1997

Gonzalez, C. (2005). Decision support for real-time, dynamic decision-making tasks, In *Journal of Organizational Behaviour and Human Decision Process*, 96(2005), pp142-154

Inmon, W,H .(2002). *Building the Data Warehouse*, 3rd ed, Wiley, New York, 2002

Jarke, M., Lenzerini, M., Vassiliou,Y., Vassiliadis, P (Eds.), *Fundamentals of Data warehouses*, Springer, Berlin, 1999

Jones, K.(1998). An Introduction to Data Warehoursing: What Are the Implications for the Network, In *International Journal of Network Management*, 8(1998), pp42-56

Kimball, R and Merz, R. (2000).*The Data Webhouse Toolkit*. Wiley, USA, 2000

Levi, S., and Arsanjani, A. (2002). Developing and Integrating Enterprise Component and Service, in *Enterprise Component and Service*. Vol. 45. No.10

Macomber, J, D., 2004. IT Strategy for Construction Companies: A Pragmatist's Vision, Retrieved October 2006, from http://ocw.mit.edu/NR/rdonlyres/Civil-and-Environmental-Engineering/1-464Spring2004/A0885389-423D-4B68-91BE-5E89450572FD/0/jmacpragvision.pdf

March, S, T., and Hevner, A, R., 2005. Integrated decision support systems: A data warehousing perspective, In *Journal of Decision Support Systems* (2005) 13 pages, retrieved September 2006, from www.Sciencedirect.com

Matthes, F. et al. A Process-Oriented Approach to Software Component Definition, retrieved September 2005, from http://citeseer.nj.nec.com/cache/papers/cs/22844/