Use of Open Source Business Intelligence Software in Hospital Management

Michal Miklas

A Dissertation submitted to the University of Dublin,
in partial fulfilment of the requirements for the degree
of Master of Science in Health Informatics

2010

Declaration

I declare that the work described in this dissertation is, except where otherwise stated, entirely my own work, and has not been submitted as an exercise for a degree at this or any other university.

Signed: _____

Michal Miklas

Date: **3rd September 2010**

Permission to lend and / or copy

I agree that the Trinity College Library may lend or copy this dissertation upon request

Signed: _____

Michal Miklas

Date: **3rd September 2010**

Acknowledgements

The Author acknowledges and wishes to thank the following, without whom this dissertation would not have been possible:

To my supervisor, Professor Jane Grimson, for your guidance and support throughout the completion of this work.

To the course coordinator, Dr Lucy Hederman, for your help and support throughout the Health Informatics programme.

Summary

The purpose of this research is to find out if the hospital mangers can benefit from the implementation of Business Intelligence (BI) software. It is focused on the On-line Analytic Processing (OLAP) approach for querying hospital data and the use of the pivot tables and charts for interactive data analysis and presentation. The OLAP tools and techniques are used for querying of large data sets usually available in a data warehouse or a central data repository, in which data from various hospital systems are integrated. These tools allow non-technical users to extract the information using predefined analytical data models in the form of self-service information extraction. On-line Analytic Processing, in combination with the front-end tools included in the BI software, provide the decision support tools required by the hospital management.

The information required by hospital managers for the making of decisions can include for example the average weekly number of admissions from Accident & Emergency (A&E) in the last four weeks. This information is helpful in planning of bed occupancy for the week. The employment of BI software with OLAP techniques is expected to ease the access to this type of information by reducing the time needed to collect and process the data and by presenting the information in a suitable way. Analysis of data provided by the BI software with the OLAP approach should allow the hospital managers to easily reveal the information related to performance of individual hospital employees or departments. This assists in the management of the performance and quality of the hospital care by triggering further investigation of individual cases identified by using of BI software.

Research Questions

- Can hospital managers benefit from the utilisation of the techniques and tools available in Business Intelligence software such as the On-line Analytic Processing (OLAP) approach for querying hospital data with the use of the pivot tables and charts for the interactive data analysis and presentation?
- Can the availability of the information for the management of the hospital be improved by the use of these analytic tools?
- Can this be achieved by implementing Open Source (OS) software without increasing hospital annual costs for the software licenses?

Findings

The hospital managers were pleased with the simplicity of the information analysis provided by the Open Source Business Intelligence software and the availability of data without the need for manual integration and processing. For some of the interviewed hospital managers, the integration and automation of data processing seemed to be more relevant than the analysis views. Nevertheless, the analysis of the data using the OLAP interface to access the processed data was seen as extremely helpful when compared to the current processes used to achieve the same results.

The managers agreed that the availability and accessibility of the information presented through the implemented BI system would be an improvement compared to current ways of obtaining of the information used in the decision making.

The zero licence cost of the OS software cannot be on its own considered as an advantage of the OS software as the Total Cost of Ownership of OS system would need to be assessed.

Table of Contents

Table of Figures

Table of Tables

Chapter 1: Introduction

This section of the dissertation begins with an overview of the research field. The background information provides a brief description of the areas and topics related to the research subject and are relevant for the formulation of the research objectives and questions. The next section of the introduction focuses on the main themes leading towards the definition of the research questions. The research questions and objectives will be presented in a standalone section. This is accompanied by a section where the research methods and techniques used in this study are explained. Finally the introductory chapter also deals with the motivational factors for the research and looks at the overall structure of the dissertation.

1.1 Overview

If an individual makes a decision the accessibility and availability of information is the influencing factor. The relevant information presented in a suitable way therefore plays important role in all human activities which involve the making of decisions.
In the management of a company there are many decisions made at various levels of management and affecting various periods of time. All information about the company's processes, activities, results and performance available at the time when the decision is made can contribute to making of the correct decisions. This affects many aspects of a company's management including planning, resource allocation or monitoring. The information which influences the decision making in a company can come from various areas, for example finance (revenue, cost, cash-flow), procurement (warehouse/stores stock balances, purchase orders, warranty claims), marketing (customers, complaints, sales) or work processes management (work orders, repairs, support calls).

The computer based Information Systems (IS) have been used in business for decades now. The volume of data stored in company systems has been growing as the computers improved and became more affordable. Furthermore the functionality of these systems increased to enable the IS to cover more business functions and processes. All data available in the company systems can be transformed into valuable information which can be used in the process of decision making.

The role of data in the company systems extended over the time. Originally, the data recorded in the computer systems was mainly used to describe and capture certain elements of the company's environment, process or activity including their attributes. The data could be viewed and amended within certain processes during interaction with the computer IS. The potential usage of the data, for

the discovery of more information, grew with the volume of the data stored in these systems over time. The structure of data in these systems was optimal for transactional operations: fast insertion of new records, fast updating of data and quick searching. For the analysis of the data, a different approach to the data structure design had to be used and the transformation of the data into a more suitable form was required. Also the data discovery and information extraction needed to be done on a copy of the data since these activities could not affect the performance of the transaction system. This led to the emergence of the data warehouse or central data repositories. These can contain the data in the form which is more suitable for the extraction of the information used to support the decisions. The systems which utilise the data collected over the time to produce the information used in the decision support are called data-driven decision support systems. (Power, 2007)

The situation with hospital information systems and the data included in these systems is similar. The information needed to support the decisions in hospitals comes from various areas and supplements the information relevant for business organisations. The relevant information can be related to patient administration (admissions, discharges, accident and emergency visits), patient treatment (diagnoses, treatments and prescriptions), laboratory (orders, tests, results) or radiology (orders, scans, reports). The same applies that the system for the administration are designed to be able to record, query and search information quickly but the ability of the system to provide insights into the data and help discover more information is limited. The availability of data-driven decision support systems based on the data warehouse or central data repository, integrating data from more systems, or modules of a single system present in the hospital could enhance the decision making in the processes of hospital management.

The following sections of the introduction will provide the background of the areas which influence the use of the data-driven decision support systems and the overall situation with the availability of the information used in hospital management.

1.2 Background

The information which can be provided by a decision support system (DSS) has to be based on the existing knowledge, data or their combination. The understanding of the domain encoded in a form of rules or models is used by the knowledge- or model-based decision support system (Power, 2007). These types of DSS are sometimes called Expert systems (Giarratano and Riley, 2005). The data-driven decision support systems do not contain the domain knowledge. These systems provide access to the information based on the data recorded in the information systems. The data-driven DSS are set to transform the data into a form understandable by the domain expert, the user of the DSS, to support a decision. The data-driven DSS can contain rules to automate the processing of data and highlight any important information.
An example of hospital data is the list of patient admissions with attributes related to patient or admission. The information produced by the data-driven DSS, based on this data, can be the number of admissions in the previous month or the department with highest number of admissions in the previous year. There is knowledge required to apply this information to make decisions which impact on the future. This knowledge is not part of data-driven DSS and is possessed by hospital managers. An example use of knowledge is the change in staff planning for the future months or years based on the workload in hospital departments.

The history of the use of information technology in healthcare and hospitals varies between countries, but in general the adoption and acceptance of Information Technology (IT) in healthcare was slower than in other areas, for example in business or banking. However, this represents an advantage for the adoption of the data-driven decision support systems in hospitals as hospitals can benefit from the experience of these systems being used in non-hospital or non-healthcare environments. If the technologies and techniques used to build hospital systems are similar, and this is true mainly in the case of the hospital administration systems, the adoption of the tools which have successfully been used in other sectors can be straightforward.

Data-driven decision support systems that are in the business sphere are currently referred to as business intelligence systems. Business Intelligence *"describes a set of concepts and methods to improve business decision making by using fact-based support systems" (Power, 2007).* The "intelligence" refers to a smart way of combining information from various company systems and additional data sources for providing summary information based on the facts related to the business recorded over the time and stored in the company systems. The summary information in data-driven DSS is available at various levels of aggregation and allows the decision maker to move between

these levels easily using a user friendly visual and interactive interface. Business intelligence systems can also be represented by groups of data presentation and data analysis tools. The business intelligence software can include also a data warehouse. Alternatively the data warehouse can function independently from the reporting and data analysis tools.

In healthcare data-driven decision support platforms are referred to as business or clinical intelligence systems. The “business intelligence” term is used when the DSS is focused on the administrative and management of the hospital. The “clinical intelligence” term is used when working with the clinical data. The DSS principles, features and function however remain the same regardless of the content or the nature of the data.

Without the use of data-driven DSS, the transformation of the data into information can be awkward and laborious. This process would involve manual data extraction, integration and summarization with the help of proprietary reporting modules of various systems and spreadsheet application. This process is often not automated and takes too long time.

While the history of business intelligence systems goes back to the 1990s (Power, 2007) and maturity of these systems can be observed in the past decade, the healthcare and hospital sectors only started discovering these data-driven systems recently. The application of business intelligence systems principles is more suitable for the administrative decision makers. These would benefit most from the data-driven DSS when utilising the system’s reporting and data analysis functionality with a focus on performance, care quality improvement, planning and monitoring.
The hospital environment can also benefit from business intelligence system capability to connect to various data sources including legacy systems. This is done either by connecting to their interfaces directly or by importing data extracted from these systems. The ability to utilise the interfaces of the system to which the DSS can connect depends on the data storage design and structures and the standard of the data access.

1.3 Study Research Question and Objectives

This study focuses on the management and administrative processes which involve decision making at various levels of the hospital management including mainly the planning and monitoring related to the quality of the provided care.

1.3.1 Research Objectives

The purpose of this research is to find out if the hospital managers can directly or indirectly benefit from the implementation of Open Source Business Intelligence (BI) software in hospitals. It is focused on the On-line Analytic Processing (OLAP) approach for querying hospital data and the use of the pivot tables and charts for interactive data analysis and presentation. The OLAP tools and techniques are used for querying large data sets usually available in the form of a data warehouse or central data repository which integrates data from various hospital systems. These tools allow non-technical users to extract the information using the predefined analytical data models in the form of self-service information extraction. The OLAP in combination with front-end tools included in the BI software provides the decision support tools used in the processes of the hospital management.

The research will attempt to prove that the availability of the information for the management of hospitals can be improved by these analytic tools and the following can be achieved:

- make the information available in the most appropriate form in a timely manner and make it easily accessible
- reduce the time required to obtain the information by combining the data from various systems
- automate the process of the data integration and summarisation so that the end user can focus more on the analysis of the data than the data manipulation and integration
- make the reporting and analysing data visually attractive and interactive.

The implementation of Business Intelligence software is expected to provide an advantage, when compared to not using of any software dedicated to data analysis and reporting, in terms of the availability of the information which can be provided by querying the data recorded in various modules of the Hospital Information System (HIS).

The aim of the research is also to demonstrate that this can be achieved by implementing Open Source (OS) software without increasing hospital annual costs for the software licenses. Another major advantage of the use of the OS software is the freedom in the selection of the implementation, support and maintenance providers. The decision to focus on the OS software in this research was made also in order to demonstrate that the OS BI software has matured in terms of simplicity of use and user friendliness as well as the seamless process of implementation.

1.3.2 Research Question

The research question is the following: “Can hospital managers benefit from the utilisation of the techniques and tools available in Open Source Business Intelligence software such as the On-line Analytic Processing (OLAP) approach for querying of the hospital data and the use of the pivot tables and charts for the interactive data analysis and presentation?”

The research secondary questions are:

- Can the availability of the information in the management of the hospital be improved by the use of these analytic tools?
- Can this be achieved by implementing Open Source software without increasing hospital annual costs for the software licenses?

1.4 Research Plan, Methodologies and Techniques

The research is carried out as a case study and is based on a pilot implementation of selected Open Source Business Intelligence software in a private hospital. The implementation project contains the assessment of a current set of tools used for the extraction of information from the hospital systems and its transformation and presentation. The assessment of the implemented Open Source Business Intelligence analytic tools is conducted as part of research when the software is deployed, demonstrated and used by the hospital managers. While assessing the implemented software the emphasis is put on the impact on the information availability.

1.4.1 Research Methods and Expected Project Duration

The implementation project runs in three phases and is completed within three months since its commencement. Hospital managers participating in the research are interviewed in two rounds during the software implementation project. The information is acquired in a form of semi-structured interviews carried out with each person individually. The first round of interviews concerns the current ways and forms of obtaining the information needed in the management of the hospital.

The following topics are discussed during the semi-structured interviews:

- Identification of information relevant for decision making by the hospital management
- Reasons why the information is needed
- How the information is accessed and what is the form of information
- When and how often is the information required

- How long does it take to obtain the information if needed outside predefined periodicity
- What is the origin of the information
- Can the information be produced from the data recorded in the hospital systems

The requirements for the design of the analytic data models are also discussed during the first round of interviews. The system implementation is carried out over four weeks after the initial interviews. The implementation also includes the deployment of sample data models. The second set of interviews is conducted after the hospital managers have used the system for a period of four weeks. These interviews are focused on the users experience with the analytic tools provided in the software. The duration of a single interview is 30-60 minutes. There are notes taken during the interviews and also after the interview which are analysed after the interviews are completed. The full content of interviews is not recorded in any way. This increases the willingness of those interviewed to share their opinions more openly.

An interview is also conducted with the IT director regarding the experience and future plans of the hospital with the OS software and the hospital attitude towards the use of OS software.

The questionnaire was designed and used only to support and formalise the feedback collected during the structured interviews. The questionnaire was sent to hospital managers prior to interviews aimed at feedback from using of the software. The questionnaire is described in detail in Chapter 4.

1.4.2 Literature Review

The literature review on the use of the Open Source Business Intelligence software or on studies of use of this software is carried out to provide background information. The literature review focuses on the use of information in the decision making process of management in general and also specifically in the management of the hospital. The literature is explored in order to find evidence of use of BI software and use of OS software separately.

The search for the information in this research consists of two stages. Initially the most recent information about the commercial and OS BI software is explored using internet search engines, looking for information about the features of BI software and implementations of this type of software in the hospital environments. This includes the collection of information about the use of OLAP and pivot tables for the analysis and presentation of hospital data. The internet web searches will provide the most relevant and up-to-date information from the selected areas when focusing on the latest technologies. In addition, studies connected to the subject of this research will accompany the internet searches and the information from the software vendor pages.

The search will include the combination of following expressions:

- hospital
- management
- executive
- data analysis
- administration
- business intelligence
- open source
- healthcare
- performance management
- service quality management
- online analytic processing (OLAP)
- data warehouse
- data repository
- data transformation
- information needs
- performance indicators
- decision support system

1.4.3 Data Privacy and Identity Confidentiality

The hospital information used in this dissertation does not contain any clinical data and does not contain any information by which it would be possible to identify a patient. There are some detailed data required in the process of building the analytic data models but these are not used in the dissertation. The example of such data is the date of admission, the name of department and the name of clinician. None of this information is mentioned in the dissertation and is only available in the data models presented to the hospital managers in order to assess the software features.
During the system implementation it is also required to work with some patient related information, like the time of the patient's visit, diagnosis group or the home city of the patient. This information is however used only inside the hospital and only used for the purpose of the software demonstration. None of the detailed information is included in any way in the dissertation.
Individuals' opinions and comments are included in the dissertation but this is done anonymously, so the name of interviewed person is not mentioned in the dissertation.

1.5 Motivation

The initial review of existing studies from this area showed that while there are studies dedicated to the elements of data warehousing, OLAP and visual interactive data analysis tools and studies done on the use of the Open Source Software (both groups of studies also from the area of healthcare), there was no study found on the use of OS BI software in healthcare or in the hospital environment.

1.6 Dissertation Structure

The information acquired by the literature review is presented in Chapter 2 and Chapter 3.
Chapter 2 of dissertation is focused on the role of information in managing a hospital. This chapter contains descriptions of role of information in management in general. The activities performed by the hospital managers, who are dependent on the availability of information are also discussed in this chapter. The benefits of data-driven decision support systems and their use are described in the last part of Chapter 2. Chapter 3 focuses on the characteristics of business intelligence systems and the functions provided by these systems. The data analysis and presentation functions and tools which provide these functions are discussed. The focus is also on the various BI systems both OS and commercial and the vendors of commercial software. The use of OS software in the hospital environment is described later in the chapter.
Chapter 4 elaborates on the methodology of the research and the techniques used. This chapter also includes the study limitations.
Chapter 5 contains the descriptions of the process of implementation of selected BI OS software in the hospital. This chapter contains the outcome of all interviews carried out with the hospital managers members including the analysis of the data collected during the interviews.
Chapters 6, 7 and 8 contain the research findings, summary and conclusion.

Chapter 2: Role of information in managing of hospital

This chapter begins with an overview of the functions of hospital management and the activities conducted during the management of a hospital. The management information systems will be discussed further including the position of the decision support functions and features within these systems. The theory of the decision making process will be explained based on available literature. Finally the types of decision support systems will be described with the focus on data driven DSS.

2.1 Hospital Management

The hospital management or hospital administration functions are similar to businesses to a certain extent. This applies especially in the case of a private hospital which carries many characteristics of a business organization. A private hospital provides services, deals with customers – patients, suppliers, forms a hospital business strategy and strives to maximize the profit through the reduction of costs, increase of resource utilisation effectiveness and continuous improvement of provided care (Butler, Leong and Everett, 1996). The tasks carried out by hospital managers consist of planning, scheduling and monitoring of resources, the quality of services and performance in general. The hospital operation management according to Butler, Leong and Everett (1996) includes “cost containment, capacity planning and personal scheduling”. The article also suggests that the operations cover “much broader range of decisions”. It defines the operation strategy as “the procurement and allocation of resources for the development of operations capabilities such as low costs; superior quality; prompt, dependable and innovative service delivery; and flexibility” in order to “gain competitive advantage in the marketplace.” The activities and tasks performed by the hospital employees involved in hospital management at the operational level, can be derived from this strategy. The formulation of the strategy itself can be considered as one of the most important tasks of hospital managers. However, the decisions regarding the allocation of resources for the achievement of hospitals goals derived from the strategy are made on regular basis.
The information required by the hospital managers for making decisions can include for example the average weekly number of admissions from Accident & Emergency (A&E) in last four weeks. This information is helpful in the planning bed occupancy for the week. The analysis of data using data-drive DSS should allow hospital managers to easily reveal the information related to performance of individual hospital employees or departments. This assists in the management of performance and quality of hospital care by triggering further investigation of individual cases identified by the use of BI software.

2.1.1 Organisational Performance

The operational decisions and the possible corrections to the operational strategy are dependant on the organisation's performance. The information about the performance is part of the information system of the hospital which includes the hospital employees, procedures, processes and their interactions. The information and information flow is to a great degree stored and contained in the IT systems.

The performance of the organisation can be measured using performance indicators. The hospital performance indicators can be from the following categories: productivity, cost efficiency, clinical quality, patient satisfaction, employee attitudes and behaviour, adaptability and survival, and financial outcomes (Fottler, 1987 cited in Butler, Leong and Everett, 1996). The performance indicators can be defined in many ways and cover various areas but the qualitative information provided by the indicator should always be considered in context with the circumstances. "While performance indicators will always be imperfect, management researchers have an obligation to refine measurements and relate them to controllable structure and process factors" (Fottler, 1987) cited in (Butler, Leong and Everett, 1996).

2.1.2 Performance Management

Performance management is an approach to leading and developing people which contributes to effective management of individuals and teams in order to increase the performance of the organisation (Armstrong and Baron, 2005). In order to evaluate the level of performance there must be a mechanism in place which allows the measurement of performance. This is the basic point of performance management. Without the ability to measure performance it is not possible to prove the improvement of an organisation's performance. As often said: "if you can't measure it you cannot manage it" (Armstrong and Baron, 2005).

Performance indicators are used to measure the performance. There are two kinds of indicators: quantitative and qualitative. The performance indicators can be classified in the following way (Armstrong and Baron, 2005):

- Finance – income, cost, revenue, profit, return on investment
- Output – units produces, customers served, sales, contracts closed
- Time –response time, delivery duration, turnaround time

2.1.3 Organisation Capabilities and Measuring of Performance

The performance of operations can be measured in the following areas based on the classification by Schroeder, Anderson and Cleveland (1986) cited in Butler, Leong and Everett (1996):

- Cost
- Quality
- Delivery

Although the classification was defined for manufacturing operations, the article and book search referring to these items from the healthcare field revealed that cost, quality and delivery are the key areas from the point of view of healthcare operations performance (Butler, Leong and Everett, 1996). The relation of these operational capabilities to the previously stated hospital performance measures and performance indicator types can be described by Table 2.1.

Operations capability	**Hospital performance measures**	**Performance indicator class**	**Indicator Type**
Cost	Cost efficiency & Financial Outcomes	Finance	Quantitative
Delivery	Productivity & Patient Satisfaction	Output & Time	Quantitative
Quality	Clinical quality & Patient Satisfaction	Output	Qualitative

Table 2.1: Relation between Indicators Categories

Each of the identified hospital operation capabilities involves a subset of management actions that includes hospital activities. These can be measured to monitor the planned performance and to investigate underperformance in order to achieve overall improvement. The management actions are influenced by the availability of accurate information on the hospital operations performance.

Cost

The pressure from the organisation's environment to reduce costs in order to increase profit margins or to offer lower prices than competitor results in the focus on the cost containment (Butler, Leong and Everett, 1996). Cost related information about hospital activities is normally available from the hospital financial system. Cost containment however does not impact financial planning directly but

seeks opportunities to increase the efficiency of all cost related hospital activities and processes. The management actions related to hospital cost containment include common management operation issues as described in Smith (1981) cited in Butler, Leong and Everett (1996), such as “internal information/control systems, work coordination and communication systems, manpower planning and scheduling, scheduling of patients and patient services, facility and services”. The relation between the cost containment, quality and delivery of services however is not covered in this work. Inventory and material management also belong to the important cost containment activities (Butler, Leong and Everett, 1996).

An example of relevant finance information can be the list of material types with highest value in the hospital warehouse, material in stock with the oldest purchase date (e.g. medications approaching expiry dates) or material types with the most frequent in-and out- warehouse transactions with highest value. Other groups of hospital measures that can be utilised for cost containment are workload, bed occupancy rate or service diversity.

Delivery

The delivery performance is based on the speed of the production and delivery in a manufacturing organisation, the availability of the service and the duration of the service provided. The delivery performance in health care concerns the availability of patient rooms, nurses and clinicians at the time the care is required/requested, the patient waiting times and the treatment duration (Butler, Leong and Everett, 1996). Hospital management needs to find a balance between the high service availability and promptness of service and cost involved in the sustainability of these (Butler, Leong and Everett, 1996). The successful implementation and use of performance measures regarding the service delivery require strict recording of each step of the patient pathway over the full course of their interaction with the hospital. An example of a delivery performance indicator can be the length of patient stay in hospital, the turn-around-time of a specific hospital workflow or department, e.g, radiology or the A&E average waiting time.

Quality

The improvement of quality in the health service and reduction of costs are equally important. The higher the quality of provided care according to Butler, Leong and Everett (1996) the higher the cost, so it is important to keep the relationship between these two in balance. Preventive medicine however, as shown by many studies, reduces health care cost, improves care quality and results in increased patient health (Johnson, 2008). Unlike cost and delivery , it is difficult to find a good way in which quality can be measured. The quality of patient care and treatment is defined in the literature as the best possible care which can be provided with the knowledge available at the time of

treatment (Lohr et al, 1990; Donabedian, 2003 cited in Greaney, 2009). This is however difficult to measure. One of the measures used in the past to indicate the quality of care provided by hospital was the mortality rate which was considered very controversial given other aspects which contribute to quality such as patient comfort, responsiveness and hospital technical equipment (Butler, Leong and Everett, 1996). The mortality rate also has to be considered individually in each hospital setting, hence is not suitable to be used as generic indicator of quality. We can look at hospital quality according to Wyszewianski (1988) cited in Butler, Leong and Everett (1996) using the following three dimensions:

- Interpersonal – how responsive and attentive the health care professionals are
- Amenities – how appealing, comfortable and private the facilities are where care is provided
- Technical – how medical science and knowledge are applied to diagnosis and treatment

Based on these dimensions and the fact that quality is perceived through the speed and the duration of care, the delivery measures mentioned in previous section can be used as indicators of hospital quality level, e.g. average waiting time or length of stay.
Some of the attributes of the provided care, e.g. attentiveness or comfort, are based on the subjective perception of care receivers. The performance indicators derived from these attributes can be only based on the data collected in the form of surveys which provide feedback about patient's satisfaction and is used as the indicator of health care quality.

2.2 Management Information Systems

Management Information Systems (MIS) are systems whose function is to assist the managers to do their jobs. The information and functions provided by such systems should be in a form which suits the activities and tasks of people working at various levels of an organisation's management. The purpose of the MIS is to serve the managers during the process of high level managerial decisions. It concerns the definition of the organisation strategy, designing services or products and long term allocation of resources. The management and decision making contained in an organisations operational activities involve does not utilise MIS (O'Brien, 1999). The MIS primary function – the support of decision making – can be however found in various forms and extents in the form of a system module in other types of information systems used in the organisation such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM) and Supply Chain Planning (SCP). The advantage of having a standalone MIS would be in focusing on the decision making and

the ability to integrate data and knowledge from various sources of information and from a number of different systems within the organisation.
"Information from the MIS is relevant in all three phases of a typical strategic management cycle, strategy formulation, strategy implementation and performance feedback" (Naranjo-Gil and Hartmann, 2007):

- strategy formulation – exploration and evaluation of strategic alternatives
- strategy implementation – financial analysis, results monitoring and resource deployment
- performance feedback – drivers of success and the causes of failures

The Management Information System or Management Support System describes a system which automates or supports the process of decision making and consists of one or combination of the following system types (O'Brien, 1999):

- Decision Support Systems (DSS)
- Expert systems (ES)
- Executive information systems (EIS)

While all three system types as stated above can be utilised in the process of decision making there are differences in what each system provides, how it is done and what is contained within the system.

2.2.1 Decision Support Systems

Decision support systems are those which provide support or automation in the process of decision making. There are several classes of DSS (Power, 2002) which are discussed in more detail below.
A particular decision support system is in general focused on a specific problem and can be successfully used only in areas or domains where the knowledge or the exact procedure of making decisions can be fully described. This makes it impossible to use a broad-range DSS which would be making or suggesting decisions in a field in which the process of getting to a decision cannot be fully described and justified. The most practical use of a generic type of DSS could be seen, not in the suggesting of the best of the available options, but simply providing assistance. The assistance is provided in the form of offering as much information as possible in a suitable form, possibly automating evaluation of the available options, but leaving the final selection on the user of the DSS.

2.2.2 Expert Systems

"An expert system is software that attempts to provide an answer to a problem, or clarify uncertainties where normally one or more human experts would need to be consulted" (Wikipedia, 2010c). The term "expert systems" covers both systems based on general knowledge and systems which contain expertise from certain domain; therefore sometimes this term can also be used when talking about knowledge-based systems (Giarratano and Riley, 2005). Expert systems can also be considered as a type of DSS or at least can be used to support the process of decision making. The difference between an ES and a DSS is that while the DSS uses selection of solutions based on identification of available options, their evaluation and the actual action of choice, the ES uses more complex reasoning in support of the decision making. It tries to analyse and explain the offered decision or solution. The expert system can also contain various elements of thinking or decision making simulations and modelling as well as providing forecasts and predictions based on provided historical data.

2.2.3 Executive Information Systems

The EIS serves the managers by providing summary information in an interactive and user-friendly way, allowing the user of the system to combine information from various sources and at various level of aggregation. These systems are usually used for the analysis of trends and monitoring of performance and can easily point out problems or unusual results. EIS do not contain any knowledge used for the decision making apart from the data load and processing rules which can define how the information is extracted and integrated across data from various sources. The EIS belongs to groups of DSS known as data-driven DSS (Power, 2002). At the same time the EIS can serve as a front end to specialised DSS which focus on a specific decision making process (Wheeler, 1994).
The term EIS has over time been replaced by Business Intelligence (BI) systems. These systems still cover the same set of elements and tools in areas of reporting, analytics and dashboards.
This study will deal with BI software particularly with the tools available which provide reporting and analytic functions. The Business Intelligence software tools will be further discussed in Chapter 3.

2.2.4 Hospital Management Information Systems

The concept of hospital MIS is not different from the MIS used in other types of organisations. The difference is in the data contained in the systems. Due to the nature of patient related data more complex transformations might be required when translating the data from the source systems to the repositories utilised by MIS. This depends on the structure of the data used in the system and

data storage system, which can be different from standard Relational Database Management Systems (RDMS). “Typical information elements contained in a hospital MIS relate to financial data on operational budgets, cost information per patient, per activity or treatment, and non-financial data on such diverse issues as number of treatments, bed occupancy, staff absenteeism and discharge rate” (Tan, 2001 cited in Naranjo-Gil and Hartmann, 2007).
Apart from the content of the MIS – the information which is contained and available for the users of the system – when defining the functions of the Hospital Management Information Systems, the Hospital MIS usually provides only the functions related to the management of the hospital and traditionally does not include clinical management (Tan, 2001). Another difference can be seen in the use of different MIS categories in health care in general, not only in the hospital environment. The systems with functions described earlier as characteristic of Executive Information Systems or Business Intelligence systems are mainly used in health care for organisation management and finance related activities whereas the Decision Support Systems and Expert Systems are used to support the clinical decisions concerning diagnostics and treatment.

The application of the Hospital MIS providing support to hospital managers can have influence on the following areas:

- patient administration
- human resources allocation and scheduling
- asset management
- workflow efficiency
- procurement
- finance.

The Hospital MIS functions from the system architecture point of view fall according to Tan (2001) into three processing phases where the main functions from each of these phases are data acquisition, data storage and data presentation. The function relevant to the end-user of the Hospital MIS – a hospital manager for instance – is the form of the data or the way it is displayed. This is also the most important function because the way the data is presented influences the delivery of the information. Summary tables and statistical reports are usually sufficient for the user of the system, but graphical presentation of the data is however recommended as it can provide “better intuitive feel of data trend” (Tan, 2001).

2.3 Decision Making

When the role of computers and IT in the process of decision making should be reviewed, it is first necessary to uncover what the decision process consists of. If the decision is as an activity which can be decomposed into stages or steps then the various information systems can be examined in order to see how they can support these steps. Use of any system (computer or non-computer) to support the process of decision making can be seen as use of a "decision support system". For the support of decisions made by hospital management, the focus will be on the computer based information systems and management information systems which provide functions which can be used by the managers and which affect their decisions.

The decision making consists of the following three steps based on the "Three Phase Decision-making Process" defined by Simon (1960) cited in Forgionne and Kohli (1995a) which is known to be the most popular classification describing human decisions:

- Intelligence – definition of problem or opportunity and gathering of facts and information
- Design – definition of the decision criteria formulation of the options
- Choice – application of the criteria and making of selection.

The description of these steps as stated above is simplified for the purpose of finding how information systems can support of the manager at any level during the process of decision making. In the *Intelligence* phase the decision maker identifies the problem and acquires the information regarding the problem from various sources. The step of decision making process can be supported by the use of EIS. Such a system is suitable for providing relevant, aggregated and integrated information from various sources. The decision maker can benefit from the composition of the information provided in easily understandable and readable form.

"Executive information systems (EIS) can facilitate this process by:

- filtering, compressing, and tracking critical data
- providing prompt access to detailed information behind text, numbers, and graphics (drill-down capability) in a very user-friendly manner that meets the user's decision style" (Gray, 1994) cited in (Forgionne and Kohli, 1995b).

The use of EIS system can help the manager to identify a problem for example by highlighting a drop in trend of a monitored performance indicator or point out an opportunity by providing finance related data identifying areas of potential cost reduction. Expert Systems are also considered to be systems for the support of the intelligence phase. Their focus is however more on the identification

of the problem and possibly its explanation. ES does not provide the functionality for gathering information used in the next phase of the decision making process.
The second step in decision making is the *design* of selection criteria. Expert Systems are more suitable in this phase. Using the ES, a decision model can be designed, which would incorporate the criteria used in the final phase. The DSS can be also used for the modelling and simulation. The use of these systems is however more common for clinically-related decisions rather than by hospital managers decision making.
The same applies to the last step of the decision making process where ES or DSS can be used where the explanation of the *choice* is required. These systems can explain the selection usually only for a narrow problem area where the knowledge can be embedded in the system. “Expert systems (ES) can capture and deliver the problem-solving expertise to decision makers” (Zahedi, 1993) cited in (Forgionne and Kohli, 1995b).

2.4 DSS History and Classification

The history and classification of DSS described in this paragraph is based on the information available in the article “A Brief History of Decision Support Systems” (Power, 2007). The studies on the usage of decision systems started in 1960s and the first experiments were done with a model-driven management decision support system. Also the study carried out by Scott Morton in 1966 was a model based system focused on the management decisions involving recurring business planning. Management information systems were defined by Gordon Davis in 1974 as systems which provide information for support of operations, management and decision-making functions within an organisation.
Research conducted by Steven Alter in 1980 identified seven classes of decision support system based on the dominance of data-driven or model-driven characteristics. The “File drawer” system was defined as the most data-driven system and the “suggestion models system” as the most model-driven system. In year 1979 an article was published in which John Rockart described the use of an Information System (IS) for display of critical success metrics for manager. This led to development of Executive information systems.
DSS are divided into the following categories (Power, 2002):

- Model-driven
- Data-driven
- Communications-driven
- Document-driven

- Knowledge-driven
- Web-based

Model-driven DSS were, from the beginning, focused on scheduling and financial simulation. These DSS do not require a large set of data for their operation and can work with only small volumes of data inputted into these systems as parameters.

Data-driven DSS are concerned with access to the large volume data generated over the time related to the activities of the company. On-line Analytical Processing functionality included in the data-driven DSS provides a high level view on the business historical data. Executive Information Systems are an example of DSS. The data-driven systems, initially referred as data-oriented, are developed from model-driven DSS used together with relational databases. The first EIS used predefined screens with business measures designed for managers by an analyst. These screens, called dashboards today, were static and did not provide any interaction. Also later developments and the use of EIS was related to and made possible thanks to commercial relational database software. Data warehousing and the use of On-line analytical processing as opposed to on-line transactional processing broadened the category of data-driven DSS. "Nylund (1999) traces the developments associated with Business Intelligence (BI) to Procter & Gamble's efforts in 1985 to build a DSS that linked sales information and retail scanner data" (Power, 2007). Business Intelligence as a type of software was described as set of methods and concepts used for the support of the business decision making based on the facts by Howard Dresner of the Gartner Group in 1989 (Power, 2007). Data-driven DSS were from the beginning connected with the use of relational databases which allowed efficient recording and storing of a company's historical data.

Communication DSS are systems used for the collaboration, communication and utilisation of technology for the support of decision making. The tools used within these systems included originally groupware, video-conferencing and boards. Currently there are many portal-based solutions available which covers all types of communication and collaboration including the latest communication trends utilising social networking.

Document driven DSS are systems for decision support which allow effective management, storage and analysis of document content for the purpose of decision making. These systems handle all types of document such as text, video, audio or images. With growth in data volume contained in the documents produced by people, the key function became efficient search mechanisms and display of relevant search results. The search function is embedded in portal systems focused on document

management as well as directly in operational systems allowing effective search in personal or group documents.

Knowledge-driven systems specialise in problem solving expertise. This consists of knowledge of a particular narrow subject or a problem from a certain area and the ability to solve such a problem. Knowledge-based systems are often based on rules. An example of such rule-based system is MYCIN, a system used to help the physician to make diagnosis based on the patient symptoms. Many knowledge-based systems utilise improving artificial intelligence (Power, 2007).

The last group of the systems are web-based systems which are actually a combination of the previously described DSS. These are mainly data-, communication- and document- driven systems. Web-based systems benefit from the availability of the web services and web browsers which make the DSS systems easily accessible. Web based systems also allow providing DSS in form of “software as service” (Power, 2007).

The next chapter will focus on the aspects of BI systems as a type of data-driven DSS and will explore mainly the functionality related to OLAP techniques and the presentation of the data acquired by the usage of OLAP: the cross-tabular style of the pivot tables and the pivot charts. The emphasis will be mainly on the use of these tools and techniques as part of Open Source Business Intelligence software.

Chapter 3: Hospital Data - Analysis and Reporting

This chapter initially focuses on the structure of data available in hospital systems. The concept of a data warehouse is explained here as well as the transformation of data into a form suitable for reporting and analysis. The chapter further describes the purpose of BI software and its functions and features. Various BI software vendors and OS BI software are also covered and the chapter also captures potential areas for the use of BI systems in the field of healthcare. This chapter also includes a review of available studies dedicated to the use of BI and OLAP too and their use in the healthcare environment.

3.1 Hospital Systems Data

There are many disparate computer systems used within a hospital. Most of these systems are focused on the patients and patient related events and procedures. The patient data contained in hospital systems include demographic data, account information, personal identification and treatment data (Tan, 2010). Yet another category of patient data can be represented by the data recorded against each patient and related to the care provided. This "patient administration" data could be considered as a subgroup of treatment data or an intersection of all the above data types. The patient administration data although related to a specific patient, when aggregated, reflects the quality of care provided. The data can contain, for example, the duration of the hospital stay or the list of laboratory tests carried out prior to patient surgery. This data can be referred to as secondary data since it does not have to be stored in hospital system database but can be easily derived from data which is recorded elsewhere. At the level of an individual patient, this secondary data has information value which cannot be provided for example by the date of patient admission alone.

3.1.1 Data Warehouse

In order to be able to transform the operational data stored in hospital systems there needs to be a mechanism in place which includes the following:

- The data used for information discovery (analysis and reporting) should be a copy of the original data so the performance of original systems is not affected
- The data should be in a form which is more suitable for information discovery
- The data transformation should incorporate derived data using predefined calculations/formulae

- The data can be redundant if this improves the processing of data in process of information discovery
- The data should be integrated from various hospitals systems in order to enrich the information discovery process.

These are the features of Data Warehouse (DWH) concepts used in business organisations since the 1990s. The term "business data warehouse" was introduced in article "An architecture for a business and information systems" by Barry Devlin and Paul Murphy published in IBM Systems Journal (1988). The data warehouse concept as storage of organisational electronic data for the purpose of data reporting and analysis was described in detail for first time in the book "Building the Data Warehouse" by William Inmon published in 1991. "The data warehousing concept is intended to provide an architectural model for the flow of data from operational systems to decision support environments" (Wikipedia, 2010b). The data, in its transformed form contained in the data warehouse, are the basis of data-driven decision support systems.

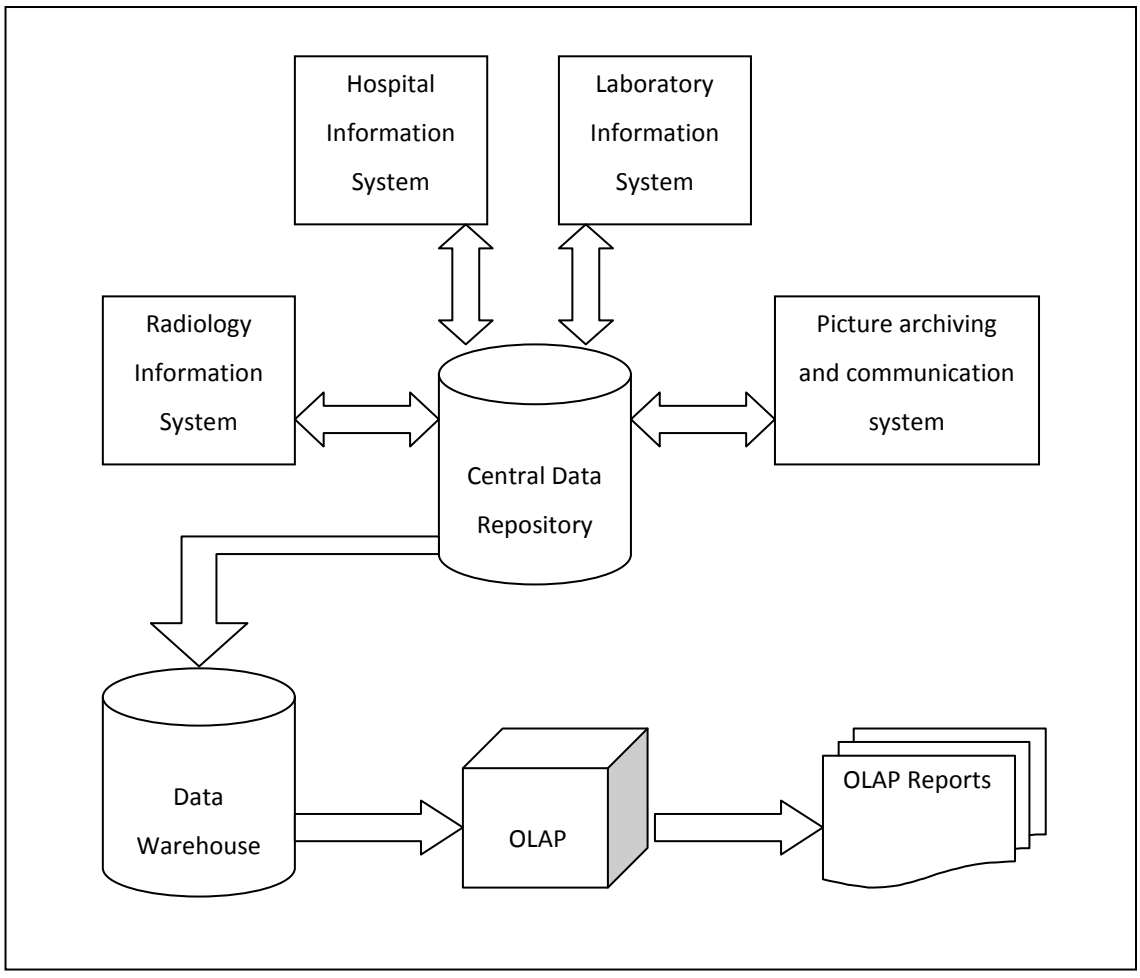


Figure 3.1: Hospital Systems, Data Repository and Data Warehouse data flows

The integration of various source systems is one of the most important features of the data warehouse approach. The integration of data coming from various data sources should bring additional value compared to analysis and reporting based on separate data sets. In order to integrate the data, there must be either a unique identifier present in the data from different sources on the level of entity; for example a unique patient identifier, or there need to be common entity attributes coming from two separate systems which allows the data analysis based on these common attributes. The most usual attribute normally available in most data sets is the date of the event. In the case of businesses, the data are most commonly from areas like sales, marketing, finance, customers, asset management and procurement. In a hospital environment with multiple systems the integration would consist of linking of data from the Hospital Information System. The integration of data is also important due to the existence of many legacy systems. The integration of data becomes with the evolution of hospital systems less important. Various functions provided before by standalone systems, are now all available within single hospital-wide system. "The Enterprise Resource Planning (ERP) philosophy is an attempt to integrate all departmental and functional processes throughout the enterprise into a single, integrated Hospital Management Information System (HMIS), enabling enterprise wide information management and decision making on all organisational operations (Tan, 2010)." The structure of the data within such systems would already provide sufficient integration of the data within the source system and such an ERP system might even provide data analysis and reporting functions. The data warehousing approach will still be required for further data analysis and also if the data from additional systems needs to be included in reporting and analysis (Figure 3.1).

The transformation of the data into a form suitable for reporting and analysis in a data warehouse is another important aspect of the data warehouse concept. The structure of the data in source systems (given that there is an interface which provides access to source system data) is designed to fit the transactional nature of the systems. If there is a relational database used to store the system's data, the structure of tables and relations is suitable for system transactions, such as creating new records, altering records or record search. As a part of the transfer of data from source system to data warehouse, the data is transformed into a form more suitable for reporting and analysis. The transaction system database tables normally contains the descriptive information in the form of internal database identifiers and use these together with lookup tables, when the information is retrieved and displayed in source application. These links are either completely translated into a flat structure, where the tables used for reporting contain all quantitative and qualitative entity attributes or are transformed into a simplified relational schema which is more suitable for data analysis.

3.1.2 Data Mart

A data mart represents a subset of data warehouse data which contains data related to a particular function or to a particular department. Data marts are usual in large organisations where the set of data marts actually forms the data warehouse. The structure of individual data marts within the warehouse is the same. The data in each data mart only relates to a particular category. An example of hospital data marts can be data marts based on hospital department, patient's county or insurance company. The data mart approach is useful if the groups of data are likely to be analyzed on their own and can contribute to better performance of data warehouse and can simplify the configuration of user access rights.

3.1.3 Clinical Data Repository

A clinical data repository (CDR) is a repository of patient related data integrating the data from various systems containing patient data. "CDR offers a comprehensive source for storing and retrieving relevant, reliable and accurate clinical information" (Tan, 2010). This repository of data is not meant to be used to build summary information neither is it supposed to provide general diagnosis or treatment information.

The CDR is used to retrieve comprehensive information related to a single patient at a time. Ideally the CDR is built first with the main hospital system and other applications or systems then utilise the same data repository to read and write patient related detailed data. A clinical data repository can also be implemented as a real time copy of data available in all hospital systems. This copy can be used to read and write patient data as well as the original systems and CDR is always kept synchronised with all source systems' data. The data repository is not suitable for data reporting and analysis because it contains the data in a form suitable for transactional operations. Therefore even if there is a CDR present in the hospital, a data warehouse is still required for data reporting and analysis. The presence of a CDR however simplifies the data integration in the process of building of the date warehouse solution. Some additional integration is still required in case the data warehouse needs to include non-clinical or non-patient related data.

An innovative approach which combines concepts of data repositories and data warehouses is represented by the software product Amalga from Microsoft (Holland, 2009; Holland M., 2009) which defines itself as 'Unified Intelligence'. The Amalga platform is in essence a data repository constantly fed by data from all hospital systems in real time. The noticeable feature of this solution is that there is no need to define the data structures and load procedures for the data coming from other systems. The Amalga software always contains a real-time copy of all transactional data in its original form and at the same time creates a simplified copy of the data which is more suitable for

data querying and analysis. According to a Microsoft study (Holland, 2009) this transformation is done automatically which enables non-technical users to have instant user friendly access to up-to-date clinical, financial, operational and administrative information. The studies and presentations criticise the standard DWH and BI approaches which rely on predefined data structures and analytic data models and the updating of data in DWH in batches. However, one of the studies adds that "Amalga complements, rather than competes with traditional business and clinical intelligence solutions" (Holland, 2009). This appears to be due to the fact that the business intelligence functionality related to the data reporting and visual analysis would still be needed since Amalga is primarily focused on the optimal data structures, storage mechanisms and seamless data transformation and integration.

3.1.4 Data Structure and Storage

The data in hospital systems is easily accessible by third party data extraction tools if the data storage engine used is a standard Relational Database Management System (RDBMS). This type of data structure and storage is most suitable for transactional systems. It is appropriate for the systems which work with the patient related administrative data and all patient's diagnostic and treatment related events recorded during the patient hospital stay. There are however still many legacy systems used across hospitals utilising proprietary data storage engines which make reading the data more difficult or impossible in the case where the system does not incorporate some type of standard data interface.

The issues with the data extraction and integration for the purpose of loading the data into DWH are the same as the issues with the system interoperability and integration from the data exchange point of view. "A large number of legacy clinical systems, which do not provide the support for interoperating and sharing information either logically and/or physically, have been independently created or administered in the National Health Service. In addition, most of the healthcare systems in England and Wales are built with different system platforms and Database Management System (DBMS)" (Tan, 2010). In order to extract data from systems without a standard data read or exchange interface, there has to be at least a reporting module or data export function available within these systems which would allow exporting of data into flat text files.

The ability to analyze the data extracted from hospital systems assumes the availability of the relational-like data format which can be transformed into entity based data stored in the data warehouse. While the relational structure of data as designed for the purpose of the hospital system might be not suitable for reporting or analysis, the relational character of the data structure simplifies transformation of data into a more suitable form in the data warehouse.

3.1.5 Data Quality and Coding

The ability to analyse data is dependent on the quality of the data, the data coding and classification system. The form of the data in the hospital systems is based on the application design. Modern Information Systems should include data validation of every single field which is entered by the user of the system. In general the occurrence of free form based fields in the application should be minimized since it makes data processing, aggregation and analysis more difficult. The fields for the data input should only accept the entry of values from a predefined set. An example of such attribute can be the patient county. If this field could accept any entered value, the analysis of the number of patients by county would become very awkward. An example of this would be values 'Dublin', 'Co Dublin', 'Co. Dublin' which all refer to County Dublin but will be displayed as three different counties when analysing the patient groups. The classification and coding of all attributes of all entities (unless a detailed description is required) should be required and mandatory.

The cross-system classification of attributes is also important. Also when the attributes are coded, unless the same coding is used in all systems or there is mapping defined between the systems, the integration of data from various systems would not be possible for the purpose of analysis. "The adoption of consistent and shared coding standards across all organisational units is one approach to reducing and eventually eliminating the possible proliferation of data redundancies and data update anomalies throughout the different databases and systems" (Tan, 2010). While this mainly emphasises the elimination of redundancies and anomalies in the systems, it also eliminates designing of additional clean-up and transformation procedures required for the transfer of the data to the data warehouse.

3.2 Business Intelligence software

As explained in previous sections, Business intelligence software evolved from Executive Information Systems and as such the primary function of the BI software is to provide summary information related to an organisation's performance and processes to managers. The Business Intelligence term covers many software tools with various functions, which can be used in various areas not only business or performance related. Business Intelligence usually relies on the data available in DWH but the data warehouse and DBMS on which it is based does not have to be part of BI system. The software tools which are used for data extraction, transformation and load (ETL) data into DWH are often included in BI systems. TechEncyclopedia defines the BI software as: "the software that enables users to obtain enterprise-wide information more easily. Such products are considered a step up from the typical decision support tools because they more tightly integrate querying, reporting, online analytical processing, data mining and data warehousing functions" (techWeb, 2010).
The BI software normally includes BI tools with the following features:

- Access to content via Web Portal
- Report Repository
- Data Model Repository
- Detailed Reporting
- Summary Reporting (cross tabulation)
- On-line Analytic Processing
- Data Analysis (Pivot tables and charts)
- Report Designer
- Data Model Designer
- Data Mining
- Dashboards
- Trend Analysis and Forecasts

The architecture of a BI system is usually based on the client- server model. Most BI systems provide these functions via web portals which can be accessed from web browsers. Sometimes the design features which include the definition of reports and data models are provided by standalone client applications. The web portal and client applications connect to an application server to access the definition of reports and analytical data models. The data analysis tools were in the past provided by standalone applications connecting to the BI server, lately however this functionality is also provided through the web portal. The functions like data modelling, DWH design and ETL procedures design,

execution and management are usually provided by standalone applications which connect to the application server to work with the content of DWH and analytical data model and report repositories. The BI system server repositories are used to store the definitions of the reports, analytical data models and ETL procedures. The definition of DWH structure is normally separated from these repositories.

3.2.1 Online Analytical Processing

The data analysis in BI tools is provided by Online Analytic Processing (OLAP). It is an approach to answer the analytical queries quickly. The analytical query is a request for information which can be provided by the data stores in the DWH. The analytical queries used in OLAP work with the concept of dimensions and measures. Measures can be defined as simple counts of items but can be also based on the quantitative attribute of an entity. The dimensions are based on qualitative entity attributes and can be considered categories, which are used as groups when retrieving summary information. More than one dimension can be used in an analytic query so the analytic queries are often called multidimensional queries. Queries are executed using the multidimensional analytic data models. Each data model consists of a predefined set of measures and dimensions. This approach allows the retrieval of summary level information from the data stored in the DWH in a way which allows the translation of a multidimensional query into the data query language used to extract and sum the relational data.

The translation of requests formulated in natural language to the multidimensional analytic query is not difficult with the use of visual composition of the query. The query execution speed is enhanced by caching included in OLAP server and can be improved by designing analytical data models aggregations which contain pre-calculated summary measures. The pre-calculated data can be optimized using the statistical data collected over time on the most frequently executed queries. The DWH structure of the data designated for the multidimensional analysis is specific to the OLAP approach. For each analysed entity there is a table containing all quantitative information related to the entity called a “fact table”.

The dimensions are defined in dimension tables and the relation to fact table is defined in the relational database schema. The dimension tables can be simple lookup tables but can also contain a dimension hierarchy. The hierarchy allows the grouping of measures at various levels. An example of a report using OLAP and the diagram showing how the data are transformed is demonstrated on Figure 3.2. This figure demonstrates a transformation of admission record attributes into dimensions. The OLAP report displayed in this figure can be used to analyze the “Average Length Of Stay” by “Insurer“ and “Admission Date” dimensions defined in the underlying OLAP data model.

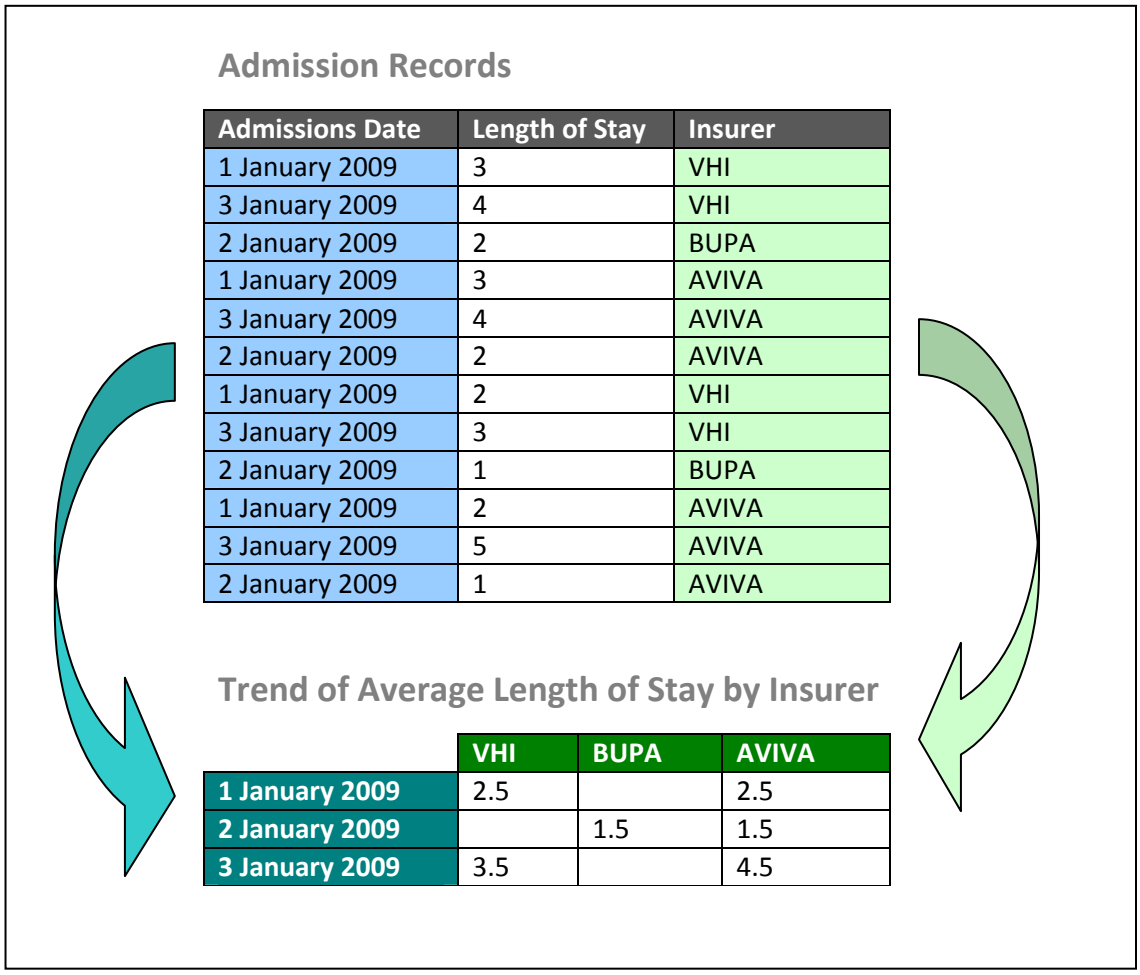


Figure 3.2: Example of reporting based on OLAP

The OLAP within the BI software is provided as a service or interface to which the BI reporting and analytic tools connect to run the analytic queries. There are several commercial software vendors who have OLAP server software in their portfolios, the biggest (based on the market share in 2006) are (Pendse, 2008b):

- Microsoft: SQL Server Analysis Service
- SAP: Business Warehouse
- IBM: COGNOS TM1
- Oracle: ESSBASE, OLAP Option
- Microstrategy: OLAP Services

There is also an Open Source OLAP server available called Mondrian (Weins and Cope, 2008; Scamuzzo, 2009). This OS project is managed by Pentaho and although this server is included in the commercially available Pentaho BI Suite Enterprise Edition, it is also available under open source licence as a part of the Community Edition of Pentaho software. The Open Source software licensing

models and Open Source Business Intelligence software including Pentaho will be discussed in section 3.3.1.

The Multidimensional Expressions (MDX) Introduced by Microsoft in 1997 (Pasumansky, Zare and Whitehorn, 2005) is a standard analytical query language. The OLAP interface follows standards too by using Extensible Markup Language for Analysis (XMLA). This allows querying of any standalone OLAP server from various reporting tools. The OLAP server is however usually included in BI software.

3.2.2 Pivot Table and Charts

The result of an analytic query is summary information presented in the form of cross tabulation. This table normally contains grouping which is displayed in the form of groups in column and row headers and measures displayed in the body of the table. There can be multiple dimensions used on row or column headers, or two or more levels of the same dimensions.

The static results returned by multidimensional queries are used for high level summary reporting. With the reports designed to show static contingency tables the query is predefined by the data analyst and the user does not have the option to amend the report content.

The pivot table is a tool included in BI software which allows the analytical queries to be changed interactively and quickly get results for changes done in the query definition. The pivot table presents the data in form of cross tabulation with option to:

- change the query filter using the dimensions not included in table layout
- rearrange, add or remove dimensions displayed in table column and row headers
- move up or down in the dimension hierarchy to focus on a greater details or move to higher level of data aggregation
- swap column and row headers in the table
- order dimensions items by the measures displayed in table body
- display detail records which constitute the summary information.

Pivot charts complement the pivot table and the cross tabulation view on the data. These charts dynamically change the content depending on the data displayed in the pivot table. Sometimes the chart elements can be defined as actionable allowing modification of the pivot table and chart query. In this way the chart itself can be used to change the aggregation level by clicking on chart elements. A bar in the bar chart representing certain groups can, for instance, change the query and display summary of the sub-groups of the group on which the user clicked. Alternatively the pivot chart can

be configured to display detailed records which are 'behind' summary figures when the user clicks on a chart element.

3.2.3 Use of OLAP for Decision Support and in Healthcare

When looking for information on the use of OLAP as DSS or its use within the DSS system, two of interesting studies were discovered including studies done on the impact of the use of OLAP in the healthcare environment. The search was done primarily on ScienceDirect web site and using Google Search.

One of the studies (Palaniappan and Ling, 2008) focuses on clinical decision support and combines OLAP and data mining with decision trees to improve decision making. It uses patient administration data together with clinical patient data to show how data mining decision trees can be integrated into OLAP as hierarchies to support the diagnosis of diabetes. The study enhances the OLAP analytical data model which on its own does not include any knowledge by adding decision tree related to diabetes and results from laboratory tests. It creates an artificial hierarchy by transforming the measured plasma glucose results into hierarchy items based on the definition of plasma glucose ranges from the decision tree. In this way the analytical data models can provide support for diagnosis or just give additional attributes by which the patient data can be analysed. The clinician can then easily identify which groups of patients have the highest probability of having diabetes.

A great benefit from using OLAP is presented in the study "Doing more with more information: Changing healthcare planning with OLAP tools" (Tremblay et al, 2007). This study was carried out in a regional health care agency in Florida where the focus was on the impact of OLAP on decision making. The study in its introduction emphasises the qualities of the OLAP concept, mainly the ability of users – decision makers – to work with large volumes of data and to decide on how the data is presented. It is compared to technologies preceding OLAP which did not allow the extent of flexibility, interactivity and intuitiveness which is provided by tools based on OLAP and provided only views on data predefined by data analysts. The study further pointed out that in the past in general there have been many implementations of DWH which were not successful because they were not used together with OLAP or not designed to be used with OLAP. In these unsuccessful implementations, the DWH did not have the expected effect since the reports were static and did not provide any interaction. The process of getting additional information or a different view on data always involved an IT person with technical knowledge of the DWH data structures. The study analyses the impact of using of OLAP on the performance of employees and examines the nature of tasks carried out by them in the processes of health planning.

The main function of this health planning agency is to regulate healthcare providers in order to efficiently meet healthcare needs. In this process the following needs to be assed regularly: utilisation, costs, population, resources and quality of care. Such assessment is a data intensive process and therefore suitable for the application of DWH and OLAP techniques. A data warehouse was built in 1998 but although the data contained in the DWH was very detailed, there were only three major static reports produced from the data with the focus on the number of uninsured and their utilisation and costs. The static reports naturally caused the need for further analysis of the reported figures. Due to this a decision was made to provide the planners direct access to DWH data through OLAP and the use of analytic data models.

This research was done mainly through utilising interviews and direct observation of the employees work before the OLAP based solution was implemented, during the implementation and when the project was completed. The focus was on the work experience, work activities and tools used when carrying out the work tasks. 'Task-technology Fit' theory was used to assess the impact of the OLAP implementation (Tremblay et al, 2007). In the pre-implementation environment the tasks were categorised as: knowledge acquisition and knowledge packaging. The knowledge packaging consisted of the collection and composition of information from various sources and with various level of aggregation in a visually attractive form for the requestor of the information. A major portion of the activities were focused on the data collections, merging and transformation into a presentable form. These were done manually utilising the data extracted from the DWH and combining them with data from additional data sources. When there was a change in a request for information the process had to be largely repeated.

The introduction of OLAP allowed employees to reduce the amount of 'manual' work and focus on 'working with information'. "The added flexibility provided by the OLAP tool allowed the health planners to leverage their individual skills while allowing them to exercise some new skills, in particular: interpretation and judgment. Interpretation became a large part of their work." (Tremblay et al, 2007). The study concluded that the use of proper technology such as OLAP has the potential to improve performance in the field of healthcare planning. One of the main outcomes was the change in the type of tasks on which the employees worked. "Rather than data collectors, they began performing more as consultants." (Tremblay et al, 2007)

3.2.4 Commercial Business Intelligence Software Vendors

There has been major consolidation of the vendors in the BI software market in the past decade (Pendse, 2008a). Large IT companies acquired smaller BI vendors in order to include their BI tools into their software portfolios or to enhance the range of products and services offered. The Table 3.1 contains the top software vendors and their BI software and indicates major acquisitions in the BI field. It can be also seen from Table 3.1 that the companies which bought the BI software vendors were the biggest IT companies and are leaders in the corresponding market with their software products like Database, Enterprise Resource Planning, Finance and Operating System. The acquisitions could be put into three categories (Pendse, 2008a):

1) acquisition of competitor (Business Objects – Crystal, Hyperion – Brio)
2) acquisition of tools which completed BI range of products (Microsoft – ProClarity)
3) acquisition of full BI suite (IBM – Cognos)

Company	Acquired Company	Acquired software
SAP	Business Objects	Business Objects, Crystal, Cartesis
IBM	Cognos	Cognos 8, Planning, Controller, TM1, Executive Viewer
Oracle	Hyperion, Siebel	Essbase, Hyperion Planning, HFM, former Brio, Siebel Analytics
Microsoft	ProClarity	ProClarity

Table 3.1: BI software market acquisitions

All of these companies can now offer a full range of BI software tools, some of them including DB and OS platform, consisting of tools providing all the functions which are expected from BI tools as described earlier in this chapter.

3.2.5 Potential of Business Intelligence in Healthcare

This section contains a summary of articles discovered during the internet search which contain fields suitable for the application of business intelligence and give an indication of the status of BI in healthcare. Some of the articles are based on or are related to healthcare in USA. The healthcare system in the USA differs from the Irish healthcare environment. However, the performance and efficiency which are most commonly addressed by the use of BI systems in hospitals are applicable to ways the private hospitals and clinics operate in Ireland.

Research carried out by Business Week in 2008 (BusinessWeek, 2009) which focused on the benefits of the BI platform in large healthcare organisations concluded that the implementation of BI is a natural step towards the improved performance. This improvement results in better patient care and more efficient utilisation of resources. During the research, managers of six healthcare organisations were interviewed in order to gain their opinions on the subject. The research indentified three areas of performance management on which the BI should have a positive impact: clinical, financial and operational. Although the research was sponsored by Microsoft it was conducted independently and offered the following valuable conclusions:

- the volume of data generated in healthcare is huge and growing vastly; it is necessary to have tools in place to transform this data into useful information
- BI can represent the central basis for the support of decision making
- a change in the company culture is required towards fact-based decision making
- the main drivers of changes in healthcare are the reduction of costs and the improvement of quality of provided care
- an essential aspect of successful BI implementation requires involvement and acceptance of clinical staff from the very beginning of the project.

The article ”Hard Times in Healthcare: Business Intelligence Provides Support” (Ament and Hildreth, 2009) describes the difficult situation in healthcare in the current economic crisis. It points out the problem of rising costs, problems with reimbursements from insurers, lowering employer contributions and suggests that striving for better efficiency and increasing performance can be accelerated by using Business Intelligence. It sees the capability of business intelligence to support and justify decisions by evidence based on facts. The areas of administration and procurement, reimbursement and clinical care are identified as those which can gain most from the use of BI tools. The article suggests that many of the healthcare organisations already have some form of data warehouse solution in place and should implement BI to unlock the information. The outsourcing and “software as a service” approach is mentioned as an option to reduce the costs related to the use of BI software. The BI projects can start as small implementations focusing on the areas with highest Return on Investment before greater investment is made in a BI system.
Another article published in 2009 suggests areas best suited for the application of BI in healthcare aimed to healthcare providers (Wanless, 2009). It starts with the analysis of demand. The healthcare organisations like hospitals and clinics operate in the system of private or semi-private healthcare system as business entities. If the healthcare sector was seen as a market with constant demand in the past, the present climate of economic turndown shows that the demand for health care services

can fluctuate and be unpredictable. The article explains that when an analysis of demand and revenue is required, based on various groups of patients, services and methods, the BI software provides a suitable solution. Except for areas focused on financial and organisational issues, the article identifies the patient registry as the best target for application of clinical intelligence.

The same author published another article in 2007. While in most cases BI provides a great opportunity to start utilising data recorded thanks to the implementation of Electronic Health Record (EHR) in the past, this article looks at the BI as one of the reasons for the introduction of EHR. The benefits from the implementation of the EHR and its use together with BI will be seen only after some time. With the understanding of the features provided by BI software and the need to electronically store patient information, the BI itself can be triggering the process of EHR adoption in organisations which have not yet commenced this process (Wanless, 2007).

In the article "Build or Buy – Considerations for Healthcare Organizations" by Madsen (2009), the author discusses the options the organisations have with the access to BI software. It lists the pros and cons of purchasing a ready-to-deploy BI solution and developing the BI internally within the organisation. This comparison does not include the actual development of the software but covers all services required for successful BI implementation including software installation, configuration, the design and development of the DWH structures, data loading procedures, analytical data models, analytical views and reports. It emphasises that in the case of buying a professional service the organisation needs to understand the system, the data structures and the flows used within the BI implementation fully. There must be a backup or recovery plan in place in case the contract is cancelled or the vendor stops supporting the software or the content. If there are employees with sufficient skills available, the BI software can be implemented and populated with data and reports by the team created internally within the organisation. This requires understanding of BI as a tool 'allowing a lot' if used properly. Such projects need to have sufficient human resources dedicated to the project fully for period project duration in order to be completed successfully.

The use of the BI software in hospitals can be beneficial also in the following areas:

- Communication with the public and marketing where the patients are considered as customers
- Communication with competitors (other hospitals) and investors
- Presentation of achieved and provided quality of service
- Demonstration of growth and efficient use of resources.

3.3 Open Source Business Intelligence Software

This part of the Chapter 3 describes the principles of Open Source software and explains its attributes. It suggests which characteristics of open source software are relevant for its use in healthcare and reviews articles capturing the situation of OS in this area. Available open source BI software is presented next and its position in the BI market is explained together with various implementation and licensing options and business models. Finally the known OS BI projects and implementations in healthcare are mentioned.

3.3.1 Open Source

Despite its qualities and growing popularity in recent years, open source software can still be perceived by many as software put together by groups of amateurs in their spare time as their hobby and hence cannot be accepted as an alternative to commercial (closed source) software. One of the principles of open source software is indeed the collaborative development of software products where developers who may be geographically distant contribute to the development of software without direct financial reward. The collaborative development brings to the world of software a higher level of innovation compared to classic development since the ideas can be shared among many people at various stages of development. Other key characteristics of open source software are the access to source code and the freedom in the use and distribution of the software.
Open source software has been defined by Open Source initiative mainly by the following (OpenSourceInitiative, 2010):

- software is freely distributed without any fees related to use
- distributed software includes the source code
- amended software including all modifications in the software source code must be made available under the original license.

The open source system is also seen by many as something that is free. Although the use of the OS software does not have to involve license costs, the cost of use of any software – not just OS – should be always expressed by sum of all cost related to software implementation, configuration, maintenance and support.
The development in many OS projects is in reality not done by enthusiasm alone only and many commercial software vendors have released their software products as OS meaning that all people have access to the source code and the use of the software does not require purchasing a licence. Many successful software projects which started as open source from the beginning are managed by commercial organisations too and the contribution to the software development is restricted to

company employees or registered developers only. These companies generate their income from the OS solution implementations which include the installation of the software, its configuration and customization and from additional services mainly including support, training, documentation and application administration. Many companies offer the option to purchase licensed versions of the open source product, developed and managed by them, which often satisfies requirements of organisations for assurance of support, maintenance, rapid fixing of identified software issues and priority updates and upgrades. The licences are also instrumental in guaranteeing legal protection. The services mentioned above might be part of a support package rather than included in the licence itself.

There are more ways for commercial companies managing OS projects to generate their income (Wikipedia, 2010a):

- dual-license model which allows the company to release the software under open source and also commercial license where the software released under commercial licence may contain additional features, documentation, testing, and protection from legal liability
- the additional functions are available in the form of a plug-in or module which interacts with the open-source software; such additional software can be closed source
- the software is provided in a form of a service where the provider hosts the application on its own hardware; the customer pays for service instead of paying for using of software
- additional services are provided based on a contract normally including support, trainings, manuals, consultations.

Besides the zero licence cost, open source software has the following qualities: reliability, customisability, freedom of choice, support from big companies and scalability (Waring and Maddocks, 2005; Garand, 2009). The reliability of the software is based on the availability of the software source code. With many people having access to the source code, discovering software issues and fixing of problems is much speedier and more frequent. The customisability of the OS software represents a major advantage over proprietary software. The openness gives the opportunity to develop additional software functionality or adjust the software features to completely fit the user needs. This does not necessarily mean that all organisations which employ OS at any level in their IT infrastructure will want to modify the source code internally to adjust the software functionality. However, there is always the possibility for them to have a system integrator or a software company carry out such modifications for them. This also illustrates the next OS feature which is the freedom of choice, since modifications can be ordered from any company and

there is no restriction in the selection of suppliers of a solution. Many OS projects have been adopted and are supported by large IT corporations which elevate OS software credibility and increase trust in OS software. Scalability is also one of the OS features which support the growth of the OS based solution from small departmental pilots to large scale organisation-wide implementations.
There are a few negative attributes of OS software including version proliferation, complexity of licences, implementation issues, lack of skilled workforce and migration related to high short-term costs (Waring and Maddocks, 2005).
A study looking at the use of OS Software in the UK public sector (Waring and Maddocks, 2005) concluded in 2005 that the OS software is not only an issue for IT and programmers. It is an alternative to the world of commercial closed source software which should be taken seriously due to its maturity and ability to reduce financial costs and dependability on software vendors.
The perception of OS by Small and Medium Enterprises was examined in a study conducted in 2009 in France (Garand, 2009) which pointed out problems with licence fee increases when the company using a software product which is key to its operation does not have a choice and has to accept it. "Oracle is the dominant solution for data base software and in 2008 it announced a 25% increase in the License costs. Users do not want to be held hostage by any one vendor and Open Source Software is an excellent alternative of 'freedom' software" (Garand, 2009). The study concluded that there are certain groups of OS applications which are not well known among the companies which can represent an interesting alternative to commercial solutions. These were the data management, CRM and business intelligence applications. The companies were not reluctant to consider use of OS software but there were couple of areas identified with the opportunity for improvement such as consolidation and communication of the strong points of OS software and support of OS software.

3.3.2 Open Source in Healthcare

There are many studies supporting the ongoing increasing acceptance of OS in healthcare. A study focused on the status of the healthcare related OS projects identified and examined 174 projects available in SourceForge webpage, a source code repository for OS projects (Janamanchi et al, 2009). The study revealed that most of the projects were related to "bio-informatics, data formats, database and medical science applications". It was also concluded that many of the successful projects were sponsored, indicating that sponsorship plays an important role. The number of active healthcare projects showed growing interest in the development of OS software in this field.

Many interesting ideas are presented in the article "Open Source software in medical informatics: why, how and what" (McDonald et al, 2003). The article lists open source applications available at the time of its publication and describes the development of medical record and research databases which were planned to be released as OS software. It explains how OS can help standardisation: "Basing more of the product on shared, open-source tools frees up resources to improve and add functionality to the market focus of their proprietary software, and a large base of Open Source underpinning would have the side effect of more standardized and more connectable software packages" (McDonald et al, 2003). It clarifies how the sharing of knowledge and collaboration improves efficiency in the utilisation of resources when building applications for healthcare: "A widespread use of Open Source parts would also benefit commercial medical system developers by reducing their development risks, and allowing them to focus capital and energy on their area of application expertise, and the implementation, support and integration services that only they can provide. In such a new world, development will be less costly, and innovation will be faster (McDonald et al, 2003)." The communication and marketing related to available OS software are seen as problematic. It suggests there should be some kind of notification mechanism which would inform the medical community about OS software. This is important for healthcare software but applicable to world of OS in general. The key to the successful spreading of OS software is the need to be informed about OS projects so the resources are not wasted on 'reinventing the wheel', but rather invested into improvements and innovations.

Beaumont Hospital implemented OS software in two phases with the start in year 2004 (Fitzgerald and Kenny, 2004). The licensing cost and their prediction together with the calculation of Total Cost of Ownership (TOC) for a 5-year period estimated a saving of 20 Million Euro. This transformation towards the use of OS system was triggered by reducing budgets and there were OS alternatives available for many of the originally proprietary software. The experience from the migration emphasised the need for the involvement of all hospital staff members not only the IT department. The change of software tools which are used on a daily basis by hospital employees must be supported and accepted by all the system users. It was also necessary to spend some time on the identification of suitable solutions and their evaluation since there was lack of "slick marketing campaigns for OS".

The article "Open Source Software in healthcare" (Carnall, 2000) describes the quite common situation which can occur with the use of closed-source proprietary software. The vendor of the HIS implemented in Walton NHS Trust ceased operation which meant the immediate cessation of support and with no access to source code of the system. There was an arrangement made and the source code was obtained and ported from UNIX to Linux which also reduced overall cost of running of the system since it had lower hardware requirements compared to the original system. This article

also noted that many OS projects are based UNIX systems hence it is not difficult to migrate them: "as many as 50% of legacy healthcare systems were written in some form of Unix, and the Posix-compliance of Linux generally means that porting them is reasonably straightforward."

3.3.3 Open Source Business Intelligence

As stated in the previous section BI started appearing among OS software only in recent years but the improving quality and general increase in popularity of OS solutions is giving rise to a growth in the number of BI OS implementations and use, in both commercially supported enterprise edition and community version installations. The potential of OS BI tools was already seen in 2005 given the success of wide spread OS platforms such as Linux operating system, MySQL database, Apache web server and PHP programming language (Gruman, 2005). The status and successful history of these OS software packages indicated that software of other types, like Business Intelligence, may become successful following OS principles and either started as an OS project or transformed into one by opening of the software source code. These OS applications were expected to be suitable for small sized companies or as departmental projects although in terms of quality and available features not quite comparable with closed source software in 2005 but "good enough" (Gruman, 2005). Based on the success of open source database software it was natural to anticipate that it will be the OS BI software, if available, that will be most likely implemented on top of the existing OS databases to provide data analysis and reporting functionality. "A popular reporting tool like BusinessObjects' Crystal Reports is not expected to support open-source databases because of the vendor's relationships with proprietary database developers such as IBM, Microsoft and Oracle" (Gruman, 2005).

The OS BI solutions appear to be successful in mid-size company markets or as new projects within large organisations where they are implemented not as a replacement for traditional BI software but to complement it, so they coexist with the original BI software (Kelly, 2009; Madsen, 2009b).

The idea of selecting OS software as a BI solution can be perceived as not suitable due to the sharing of software source code which might over the time incorporate company's business know-how. Modifications or additional features developed in companies which implement OS BI software however do not have to be provided back to community if the modified software is not distributed further (Madsen, 2009b). There also does not have to be any concerns in regard to the logic implemented within the BI system since any content developed within the BI system, such as data models, reports or analytic view, does not become part of the OS software which can be a common misconception. "Reporting tools are examples of platforms that are often marketed as applications.

The difference is that platforms typically do not encapsulate specific business processes or logic" (Gruman, 2005).

The Gartner analysis from 2009 confirms the expected increased usage of OS and predicts its further growth between 2010-2012 expecting the number system deployed to increase five-fold by the end of this period (Bitterer, 2009). There are five OS BI software projects stated in this report "that have made a name for themselves":

- Actuate BIRT
- JasperSoft
- Jedox/Palo
- Pentaho
- SpagobBI

"The provided statistics from the vendors and also Gartner's inquiry request statistic about open-source BI clearly indicate that deployment are on a sharp rise" although the total number of actively used installations is difficult to estimate because many implementations are carried out by technology partners, resellers and software companies which integrate OS BI into their product and the estimation of the increase in number of OS BI software implementations is derived from several types of measurements (Bitterer, 2009):

- number of paying customers which purchased licence or service subscription
- number of registered users (not every registration indicates active use of the system)
- number of downloads (least significant)

The Gartner's report notes that while the common areas for traditional BI software is the finance and telecommunications sectors, OS BI software is more often implemented in mid-size organisations mainly from government, public sector, manufacturing and healthcare (Bitterer, 2009).

The other important conclusions of Gartner's analysis were mainly the following (Bitterer, 2009):

- OS BI adoption doubles every year
- OS BI features are sufficient although still behind commercial BI capabilities
- lack of locally available developers with appropriate skill set may be crucial.

Another research sponsored by Pentaho, one of the OS commercial vendors, was carried out in July-August 2009 in order to map the situation with the OS BI software market (Madsen, 2009b). The research was sponsored by Pentaho and it appears to be done independently. The research findings criticised the difference in functionality between community and enterprise editions of Pentaho software. The research confirms Penatho as the leader in this market although this was not the primary aim of the research. The research focused on the use of OS BI and DWH software in organisations and the analysis of benefits and challenges. It was based on the response of 1000

people surveyed. The main reasons for the adoption of OS BI were costs, vendor independence and ease of integration. Many organisations also use OS BI to show the alternative to the vendor of the traditional BI software which they use in order to re-negotiate the licence fees and contract terms. Interestingly this helps the competition between the OS and traditional BI software and can contribute to the improvement of the traditional BI software too. The results of the survey contained the following conclusions:

- most of the OS BI projects were from the commercial OS vendors and one-third of OS implementations uses paid services and support from OS vendors
- the longer experience the organisation has with OS software the more likely will start using more OS software tools from various categories
- some vendors may lose potential customers due to different functions provided by community and enterprise versions of the OS software

3.3.4 Open Source Business Intelligence Software

The Gartner's study from the previous section (Bitterer, 2009) was used as an indicator in order to identify and describe the most significant OS BI projects together with the overview of the OS BI tools from the presentation from OpenLogic, company providing support for OS software, from year 2008 (Weins and Cope, 2008) and presentation from OW2, consortium dedicated to OS development, from year 2009 (Scamuzzo, 2009).

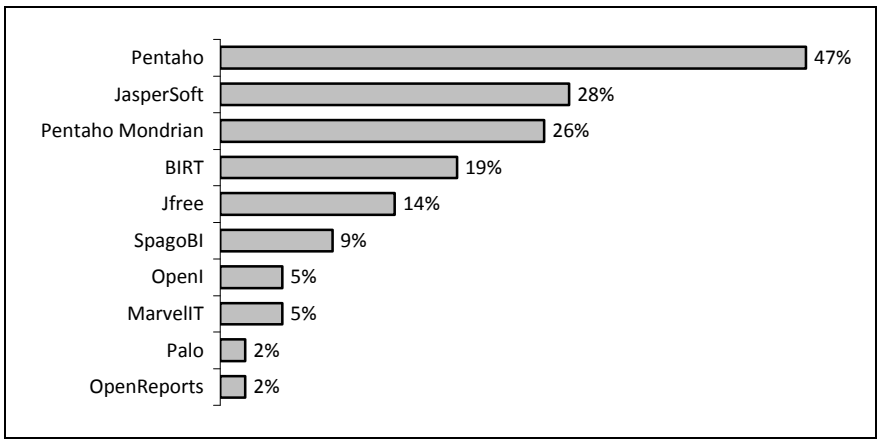


Figure 3.3: Open Source BI Tools in use

The result from the document "Open Source Solutions: Managing, Analyzing and Delivering Business Information" by Madsen (2009) displayed in Figure 3.3 was also considered. The chart displays Mondrian as a separate tool/project but this is now considered a Pentaho component, freely available and used by other BI software tools too.

Based on these documents the following OS BI tools will be further described:

- Actuate BIRT
- JasperSoft
- Pentaho
- Jedox/Palo
- SpagobBI

Actuate BIRT is a visual report designer based on Eclipse - open source Integrated Development Environment. It requires knowledge of Structured Query Language (SQL) and in general is not aimed at the business audience with regard to report development and self-service ad-hoc report design and execution. The deployment of reports designed in BIRT requires proprietary server product, but the BIRT designer is fully Open Source which allows its integration with other software tools and BI systems. It contains libraries allowing the export of generated reports in many various formats. The reports designed in BIRT can contain a combination of data, charts and images. The reports designed in BIRT are detailed reports which are suitable for listing of items showing the item attributes. BIRT is the only single software tool described here, the following BI software packages consist of sets of tools for data integration, reporting, analysis, dashboards and web portal for user access.

JasperSoft BI software provides a BI platform consisting of tools for detailed and analytic report design, report execution and Extraction Transformation and Loading (ETL) tools. The JasperSoft BI solution is based on a client-server architecture with a thin client where the users access the content over a web interface. The Jasper Server is used for administration and for scheduling of events and ad-hoc queries. The Jasper Report (based on iReport) is the detailed report designer and viewer tool. Jasper Analysis is based on the Mondrian OLAP server and JPivot in the community edition. The enterprise (commercial) version provides advanced analytical tools providing drag-and-drop functionality and interactive charts. Both editions, commercial and community, contain Jasper ETL - the integration tool. The enterprise edition of the Jasper software includes additional installation, configuration, management and monitoring tools. The comparison of JasperSoft and Pentaho BI platforms contained in the OpenLogic presentation (Weins and Cope, 2008) from August 2008 favoured JasperSoft over Pentaho, both BI systems have however been changing dynamically over the past two years. The user interface is a web based application but there is no portal like solution included in JasperSoft.

Jedox Palo Suite is another comprehensive OS BI platform providing all common BI tools. It is distinctive in that it has its own in memory OLAP server, which is used by the Palo BI Suite as well as Palo plug-in for Excel. The Palo Suite is available as a free version and a premium edition with additional functions and inclusion of software maintenance and assurance. Alternatively paid support can be purchased for the free edition too. The web version of the OLAP client in contrast to the community edition of JasperSoft BI and Pentaho BI is not based on JPivot component. The Palo web OLAP client can connect to both Mondrian and Palo servers thanks to the use of XMLA standard. There is also a desktop application Palo OLAP client available with added charts; the web version does not contain charting. The Palo OLAP web client is an interesting alternative to JPivot, since its drag-and-drop user interface is more user-friendly but it misses a couple of useful features, such as drill-replace (style of changing of grouping detail/level in which only the items of the grouping hierarchy from lower level are displayed when the user click on the group with subgroups, replacing all items displayed before), drill-through (displaying of detailed items/transactions when user clicks on an aggregated measure) and mainly the charts.

Pentaho BI is the most comprehensive BI platform in terms of the features and tools available. It includes the web portal with user access to predefined and ad-hoc detailed reports, analysis with predefined analytic views with the option to modify and save them and interactive dashboards. The Pentaho server components include a scheduler, workflow engine and Mondrian analysis server. The design functions are provided by desktop applications connecting to Pentaho repositories. Pentaho Report Designer and Design Studio are used for design of detailed and summary reports based on relational and OLAP data and design of dashboards (in enterprise version only). The Pentaho Workbench schema designer is used for the design of the Mondrian OLAP analytic data models schemas. Pentaho Data Integration – the ETL tools for the data integration formerly known as Kettle – consist of application called Spoon, Pan and Kitchen. These tools are used to design, execute and schedule the data transformation and integration tasks and jobs. Finally the Pentaho BI also includes Pentaho Data Mining. This tool is used for data mining with the orientation on statistics and data dependency analysis. The enterprise edition of the software provides different web OLAP client based on a closed-source component. It also contains additional tools for BI platform installation, monitoring and administration and tool for designing of interactive dashboards.

SpagobBI is the only OS BI software described here which does not come with two different licences. All software features are hence available to everyone without any differences. The software tools include the Reporting, OLAP, KPI, Dashboards, Data Mining and ETL. The OLAP client used on the web

portal is JPivot. The web interface can be configured to run as a standalone application as well as a web portal.
All of the software packages presented here utilise, to a certain extent, OLAP technology for the effective data analysis and reporting and the front-end tools based on the data presentation employing the cross-tabular view and pivot table functions.

3.3.5 Open Source Business Intelligence in Healthcare

Little evidence was found about the successful implementation of the OS BI software in the healthcare environment. The main reason is naturally that with the OS software freely available for download, implementations cannot be recorded and tracked unless the organisation decides to present the use of OS BI software publicly. Another factor already mentioned in previous sections is that OS BI software can be included with other software or implemented as a part of larger software implementation integrating various systems. In such cases it is also not possible to record the number of installations. The only evidence of the number of successful implementations is the information presented by the vendors which provide the commercial versions of OS BI software.

When looking for evidence of implementations and use of OS BI software in healthcare the main source of information were the web pages of the OS BI vendors. There were couple of projects found presented on the vendors websites demonstrating the successful use of the OS BI software in healthcare. The following projects supporting the success of the OS BI use are in healthcare from the Pentaho website (Pentaho, 2010):

- Pentaho software was implemented in Loma Linda University Health Care supporting mainly the analysis of billing related data allowing the analysis by time periods, service and physicians. Pentaho was chosen also because of the ease of its integration.
- The Implementation of Pentaho software in National Health Service Islington was focused on reducing costs by utilising the patient admission data. The decision to use Pentaho software was explained by Pentaho functionality providing the equivalent of proprietary software.
- The implementation in Cardiac Science demonstrates the integration of Pentaho software with proprietary database software. Pentaho software provided reporting and analysis of large volume of sales, marketing and accounting data. The software was selected also due to user friendliness and ease of use.

Chapter 4: Methodology and techniques

This chapter provides more details on the methodology and techniques used in this research. It also includes the study limitations and risks.

4.1 Case Study

The literature review was used as a basic step for this research in order to get insights into the use of Open Source Business Intelligence in healthcare and for the support of decisions by hospital management today. This research further uses the case study methodology in order to answer the research questions. Case study is a methodology used mainly in the social sciences. It uses detailed investigation of a single event with all its attributes (Yin, 2009). The subject of the examination in this research is represented by the implementation of software. This includes the usual phases of software implementation: assessment of the current situation; gathering of the user requirements; collecting of user feedback from the use of the software. The software pilot implementation allows testing and evaluating the software without the need to change any of the existing processes and procedures. Because of this the research is not employing the action research methodology. There were two techniques used for the collection of the qualitative data in this case study: semi-structured interviews and questionnaires.

4.2 Semi-structured Interviews

The semi-structured interview was selected as the main tool for the collection of information. The data collected in the interviews was analysed and used during the project and for the formulations of findings and conclusions. The interviews consisted of several topics which were used for the formulation of open-ended questions. Interviews were not recorded in order to gain greater engagement by the interviewees. The data was collected by making notes during the interview and also by writing a summary after each interview. The interviewees were both hospital managers and IT people involved in the management of the hospital at various levels. The areas discussed during the interview were a mixture of technical and non-technical topics and the agenda for the interviews was adjusted based on the person being interviewed. There was a pilot interview conducted in advance with a person not involved in the research but working in a hospital environment. The interview topics and their order were adjusted based on the outcome of the pilot interview. The

interview topics were made available to the interviewees in advance; the document in Appendix B was sent to them by email. The same document was used for all interviews.
Hospital managers participating in the research are interviewed in two rounds during the software implementation project. The first round of interviews concerns the current ways and forms of obtaining the information needed in the management of the hospital. The second set of interviews is conducted after the hospital managers have used the system for a period of four weeks. These interviews are focused on the users experience with the analytic tools provided in the software.

4.3 Questionnaire

The questionnaire was designed and used only to support and formalise the feedback collected during the structured interviews. The questionnaire is attached in the Appendix A. The questionnaires contained a list of software features and characteristics. The first group of the questions was focused on the overall attributes of the reporting and analysis software. The second group of questions was related to features specifically available in the software which was used in this study. The first page of the questionnaire contained the questions which were supposed to gain the participants' opinions and their perceptions of the software and its features. The questionnaire was intended to collect qualitative data only and it was not intended to process the responses collected on returned questionnaires using statistical methods due to a small number of research participants.
The first version of the questionnaire was tested with a person not involved in the research to make sure the questions were easy to understand and the questionnaire was not too complex overall. Based on the feedback some of the questions were rephrased and simplified and the number of questions was reduced with less important questions removed and others merged. The questionnaire was sent to the participants in the research – these were the same people who were interviewed before the software implementation. The questionnaire was distributed by email by attaching it to the email in PDF format. The questionnaires ware printed and filled out on paper. Some of the questionnaires were filled during the second round of the interviews where the questionnaire questions assisted in the interview process.

4.4 Data Collected and Other Tools

The data collected during the interviews which will be included in this dissertation will contain the opinions of the interviewees on the presented and used software. The collected data is completely anonymous, there are no names, roles or positions of the interviewed people mentioned anywhere in this document except the background of the interviewed people: administration or IT. The people who were interviewed received and signed the consent form which is in Appendix C and also received the information about the research in the form of information leaflet – see Appendix D.
The ethical approval application was submitted to the School of Computer Science and Statistics Ethics Committee by email on 9th February 2010 and approval was received from on 23rd February 2010. An ethical approval application was also submitted to the hospital on 11th February 2010. The research project was presented at the hospital research board meeting on 18th February 2010. The research project was approved at the Medical Advisory Committee on 4th March 2010.

4.5 Study Limitations

This study is a qualitative study and is mainly based on the data collected in the interviews with people. The collected data mainly represents subjective opinions of those interviewed. Also due to the nature of the project, the hospital environment and the time available for this project, it was not possible to involve more hospital managers. The research methodology used in this project - case study – is focused on implementation and demonstration of the software in a single hospital. The conclusions therefore cannot be generalised and will serve only as proof of concept.
The project does not try to compare various open source software for reporting and data analysis. There will also be no comparison of the Open Source and Commercial Business Intelligence software carried out. The software made available in the pilot implementation will be compared with the current situation with regard to obtaining information based on the patient administration data recorded in the hospital systems. There is currently no software used in the hospital dedicated solely to reporting and data analysis and which integrates data from various systems. Hence, the comparison with the system implemented does not create a full picture in the context of other available reporting solutions, both Open Source and Commercial. The validity of the results of the study could be questioned because of this; however the research questions only try to find out if OS BI software would bring benefits over not using any BI solution at all. With an existing reporting system the focus would be on the comparison of the available features and possibly the cost involved in the software licence and overall software management between the existing software and the OS

alternative. Such a comparison is not made due to of the fact that there was no comprehensive BI software package used in the hospital before the pilot implementation of OS BI software.
Only a single OS BI software package is selected for use in this study. This could be considered as another limitation. However, the components used in the particular OS BI software for data analysis are common across other OS BI software packages, so although the feedback from using the particular software package will be based on the overall experience of the solution, the components and technology utilised in the chosen software would be similar if an alternative OS BI software had been used.

Chapter 5: OS BI Software Implementation Project

This chapter will cover all phases of the pilot implementation of the software. The first section describes the pre-project stage including the description of the selection of the software for the implementation and the methodology followed throughout the course of the project.
The interviews with the people from the hospital are described in the second section of this chapter in order to capture the current situation regarding the information used in hospital administration and the channels and procedures utilised for obtaining this information. The source of the information is discussed as well as the transformation and integration of patient related administrative data from the hospital systems. The essential information required in the management of the hospital is also discussed based on the outcome of the interviews. The third section contains details about the design of the sample data models and reports based on the hospital data with an emphasis on the relevant information identified during the interviews. The demonstration of the system followed by the post implementation interviews are presented at the end of this chapter.

5.1 Project Preparation and Software Selection

There are several Open Source Business Intelligence software projects which were already reviewed in Section 3.3.4. In this part of Chapter 5 the software which selected for the purpose of this research will be described together with the reasons leading to this choice. The methodology of the software implementation project is also included in this section.

5.1.1 Software Selection

The review of the available OS BI software introduced five BI software vendors which provide and manage the development of OS BI software. Since the focus of the research is on summary reporting and analysis of data with the emphasis on the indicators of performance and service quality, one of the most important criteria was that the software should contain features offering qualitative data analysis. This is achieved by the utilisation of the On-line Analytic Processing approach and visual components which provide pivot tables and pivot charts functionality. A report defined in Pentaho software which uses this component is called an “analysis view”. Another important criterion was the portal functionality where the reports can be designed, viewed, saved and published in one place with role-based access to various information and accessibility from any computer on the hospital network without the need to install any client software. Finally the selected software package should

contain tools for data integration allowing the connection of a reporting data repository to any source of data available in the hospital systems.

The software should also be platform independent including independence from any particular database software which is used for the report definition and data repository. These criteria were defined based on the experience of the researcher due to the size of the research project and the limited time available.

Figure 5.1: Pentaho login screen

The selection of Pentaho software (Figure 5.1 and Figure 5.2) for the pilot implementation was largely subjective taking into account the above criteria and the assessment of the software packages listed in Section 3.3.4. The Pentaho software seemed to comply with all the basic criteria and the history and evolution of the Pentaho BI software suite and the record of software implementations suggested improving quality and richness of the features the software provides. For the purpose of the project the Community Edition (CE) of the software was chosen, since only this version provides fully Open Source software without any licence restrictions or costs. This version does not include the automated installation process, offers only basic documentation and as expected does not come with any professional support. The community using the CE edition of Pentaho however offers plenty of sources related to the installation instructions and configuration as well as documentation of the software features which can be found on-line.

The user friendly nature of the web portal which supports working with the analysis views was an important attribute affecting the decision although the alternative software packages which provide similar user environments were reviewed briefly. The review of alternative BI OS software was done mainly by reviewing the “demo” installations of the software available on the vendors/projects web sites. The analysis views in Pentaho use the JPivot component which is also used in other BI OS software. The Pentaho portal supports the editing of various attributes of the view including the chart properties, and the modified analysis view can be saved and shared with other portal users. The saved analysis views can also be organised in folders. The access to analysis views can be defined at the level of the folders or analysis views for individual users or users with certain roles assigned.

Figure 5.2: Pentaho main page

The analytic server which provides the OLAP (On-Line Analytic Processing) interface to the data is called Mondrian and is also part of the Pentaho solution. Although it is utilised by other OS BI software, the fact that it is part of the Pentaho software and the development is managed by the same team suggest that a stronger integration is used and communication with other elements of the software is well tuned. The Pentaho software by default uses a lightweight database management system (DBMS) called HyperSQL for the storage of the application data and initially also for the storage of the report data. The software contains database initialisation scripts if another DBMS is used as the application and reporting data repository. It is possible to use any SQL (Structured Query Language) compliant database system with a Java driver available for the connection to the database. Open source database MySQL was used in this project as the repository

of the patient administration data which was the foundation for the reporting and data analysis. All tools included in the Pentaho software, both server components and client tools, are based on Java and because the Java Runtime Environment, which is used for the running of the software, is available on many platforms, the Pentaho software is platform independent too. It can be installed in any UNIX or Linux environment as well as on any version of Microsoft Windows server. The ability to deploy and configure the Pentaho software on Linux was considered as an additional benefit in terms of reducing the overall software cost.

5.1.2 Software testing

The Pentaho 3.5 Community Edition was downloaded from the Pentaho web site for initial testing. The Pentaho BI server software comes in a single archive file and when the files are extracted, it is a matter of execution of a single script file which starts the Pentaho server. The Pentaho web portal is then accessible in the web browser within a couple of seconds on a specific web address, with the predefined “demo” reports and analysis views. The testing of the software began by running the Pentaho server in the Linux environment and the initial steps involved the configuration of the server hostname and the change of the port on which the service was running. When the sample analysis views which came with the software were explored and examined, the additional Pentaho client software tools were downloaded for the design of the data load procedure, the definition of the analytical data models. The MySQL database software was also installed and the configuration of Pentaho amended to use this DBMS.

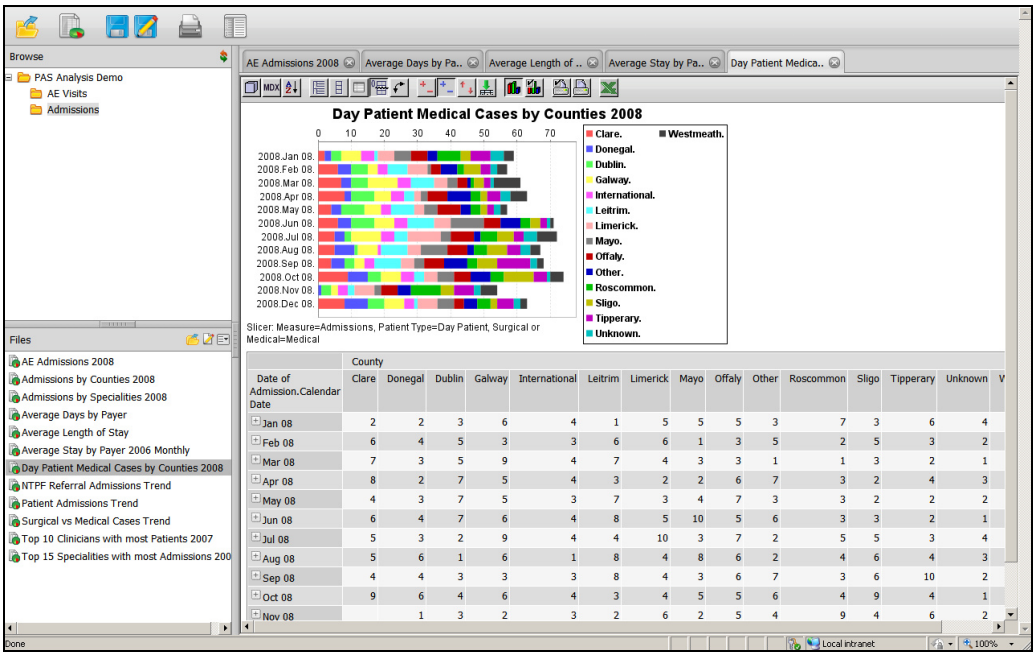


Figure 5.3: Pentaho report repository and sample report

The sample data models and analysis views available in the software were based on sales data, so it was necessary to prepare initial content showing the potential of the software for patient administration data. The nature of the data required for the hospital management was identified in the hospital annual report which is publicly available on the hospital web site. A set of random sample data was generated utilising the attributes found in the annual report with a focus on variability and the option to produce reports and charts similar to those included in the hospital annual report. A simple load procedure was designed to load the data followed by the design of the first analytical data models utilising the patient administration data. A couple of reports based on the new analytical data models were also created in order to demonstrate to hospital management the potential of the software (Figure 5.3).

This test installation of the software was done on a laptop including the server and client tools. This exercise helped to familiarise the author with the software configuration and deployment processes as well as with using the tools for the design of the data models and load procedures. This was all prepared before the initial meetings with the hospital managers and was used in a brief presentation of the software during the meetings. A copy of the configured software, including the data, data models and couple of predefined analysis views was also made available on a server computer through the use of the internet for the purpose of testing the systems capabilities and to provide a brief introduction to demonstrate the aim of the research project. After the initial interviews the participants were encouraged to try to log in to the Pentaho portal available on the internet and run the available reports in order to get better a picture of how the user interface looks and works although this was not necessary at this stage of the research.

5.1.3 Implementation methodology

The implementation of a BI solution in a hospital, and in healthcare in general, should follow a software implementation methodology. The principles and the approach to the implementation of BI software is well covered in the document “Business Intelligence Healthcare Whitepaper” (Steed, 2009). This document includes the techniques and methods which are relevant specifically for the implementation of BI in healthcare. Besides the detailed description of the steps involved in software implementation, requirements gathering and project set up, it also includes an explanation of the benefits of BI software in the healthcare environment. This document describes how a BI solution can be deployed based on Microsoft BI software. However, the principles and methods of BI implementation covered in this paper are applicable to BI healthcare project based on any BI software. Although the pilot implementation of the OS BI software in this research is only a small scale project, the “top-down” approach described in this document could be employed. The

approach is driven by the user requirements, defining what the output of the system should be, and consists of the following elements:

- Requirements Gathering – what information is relevant for the user and in what form will it be delivered
- Metrics – definition of the indicators based on interviews from the requirements gathering session
- Presentation – how the information will be accessed in the system and a definition of the way the information will be presented
- Data Requirements – how the information requested can be satisfied by the data recorded in the hospital systems
- Source Data – identification of the data required to produce the information requested and the method by which the data can be integrated from various systems to satisfy the requirements
- Design of Database – definition of the structure of the data used for reporting and data analysis including the definition of the data load and transformation procedures
- Project Organisation – definition of the project roles and their responsibilities

In a small implementation such as the pilot project presented in this dissertation the number of people involved in the project was kept minimal in order not to disrupt the usual activities and tasks of the hospital employees. The project was lead by the researcher including the requirements and data gathering, definition of the data structure, analytical data models, analysis views and reports and the software deployment and configuration. Only a little assistance was required from the hospital IT staff in terms of the server set up and the extraction of the data from the hospital patient administration system.

5.2 Information and Requirements Gathering

The gathering of the details about the information which are used for decision making and the user requirements with regard to the pilot software implementation were carried out through interviews with the hospital managers. The software features were briefly presented to the interviewees and the user requirements were related to the content of the reports and the information which will be available through the software. The requirements did not include the software functions and features. The interviewees are to various extents involved in working with data and information used for decision making. In the initial round of interviews there was also a member of the hospital IT

department interviewed in order to gain details about the hospital IT infrastructure and the hospital systems. This was done from the point of view of data storage, data flow and also to find out about any BI or reporting software currently used in the hospital and to learn about the hospital's attitude towards OS software.

The mixture of roles of the interviewees gave the research the opportunity to capture a variety of opinions and perceptions of what is needed for decision making. There were 5 people including the members of the hospital IT team interviewed with responsibilities from the following areas:

- Clinical and Corporate Governance
- Finance
- Facilities
- IT
- Nursing.

The topics for the structured interviews were prepared using the information collected during the literature review. The purpose of the interviews was to find out what affects the decisions made in the management of a hospital, what activities are involved in this process and how the people involved in management gain and use information. The focus was also on finding out what is involved in the assembly of information in terms of people, hospital systems and software tools.

5.2.1 Information Required for Decision Making

A number of hospital management activities were identified in the literature review including the following:

- Performance Management
- Resource Planning and Allocation
- Patient Care Quality Monitoring

During the interviews more details were sought regarding these activities and the information used for decisions. The purpose of this was not to capture the details of the activities. The focus was rather on how the information is used. There were a number of metrics identified during these interviews which are used in this hospital and a few of these metrics were later included in the design of the analysis views and the reports when the system was implemented.

The information identified as being required for hospital management comes from various hospital systems and also includes information which is not recorded in any of the hospital systems and is only available in paper format or is delivered verbally during conversations and meetings. The

hospital computer systems include a system for patient administration, a laboratory system, a picture archiving and communication system, a business management system, a document management system, an incident management system, a policy and education system, a theatre system, and a call centre system. Some of these systems are standalone but a number of them are modules of the main hospital system. All of these systems were mentioned as a source of information, relevant to decision making.

There are monthly meetings of hospital managers held which involve the preparation of various reports and compiling of information from these systems. There are also clinical audit meetings held regularly during which cases are reviewed and various metrics are discussed based on the data from the hospital systems related to the monitoring of patient treatments and clinicians' performance. These meetings require the preparation of details related to the provision of care and treatment. It is essential for these meetings to be able to identify those cases which require further explanation or investigation. This is done by a variety of predefined metrics and indicators. The following indicators based on patient administration data were identified as significant by the hospital managers during the interviews.

Admissions within 28 days

- This indicator represents number of re-admissions of discharged patients by the same clinician within 28 days after the discharge
- The indicator is reported as single figure for all admissions and also as the number of re-admissions for each individual department and clinician
- The variation of the indicator includes the re-admission by any clinician and the re-admission of patient by the same clinician who discharged the patient
- The number of re-admissions is reported as a total number and also as a percentage of all admissions
- It is used as an indicator of the quality of care provided and for further investigation and explanation in cases where the departments and clinicians have the highest number of re-admissions
- The indicator is based on the patients' admissions and discharges recorded in the hospital patient administration system.

Length of Stay over 28 days

- This indicator reports the number of all patient admissions for which the length of stay in hospital is over 28 days

- The indicator is presented as the total number of patient admissions as a percentage of all recorded admissions
- It is based on date of admission and discharge recorded in the patient administration system
- It is used for the identification of clinicians with the highest number of patients with stays of over 28 days in order to improve the patients care and to reduce the cost
- The long stay cases identified by this indicator are further investigated with focus on discovery of whether or not the patient had experienced an adverse event which resulted in an increased stay and also with regard to insurance related reimbursements
- The indicator is also reported separately for medical and surgical cases

Pre-operative blood test

- This indicator reports the number of surgeries for which there was no blood test carried out within 24 hours prior to the surgery
- It relies on the times recorded in the laboratory and theatre systems
- It is reported as number of surgeries without the blood test performed within 24 hours as well as the percentage of all surgeries
- It triggers further investigation of cases for which a blood test is missing

Laboratory tests over expected duration

- This indicator shows the number of laboratory tests which took longer than expected based on the type of the test
- The number of test over the expected time is reported as the total number of tests and as percentage of all tests
- This indicator is calculated for all tests and for individual test categories
- The indicator can identify laboratory workers with the highest number of tests with higher than expected duration
- The indicator allows further investigation of tests with long duration

Out of hours laboratory tests

- This indicators shows the number of tests ordered out of hours

Bed occupancy

- This is an indicator showing the percentage of bed occupancy for the week

Discharges after 10 AM

- This indicator shows the number of the discharges after 10 AM and as percentage of all discharges
- This indicator is used to improve the discharge process in order to improve the bed utilisation

Patients admitted from Accident & Emergency (A&E)

- This indicator is calculated as weekly average of admissions from A&E for period of previous 10 weeks
- The indicators assists with weekly bed planning where the weekly average of A&E admissions is used to indicate the number of beds which need to be reserved for A&E admissions

Surgery Turn-Around-Time (TAT)

- The surgery TAT is determined by duration from the time when the patients arrive in theatre until they are discharged back to ward
- The surgery TAT is based on data recorded in the theatre system which records surgery attributes and times of various events related to each surgery

In addition to the indicators, the following areas of information were identified as essential for the management of the hospital:

- Patient related information including medical history, demographics, insurer, length of stay, patient satisfaction
- Nurse related information including qualifications, education, registration, accreditation, involvement in incidents, quality of care, faults, courses, trainings and policy renewals
- Infection control and environmental audit data
- Financial information including billing, cost and revenue
- Costs related to utilities, energy and water monitoring and consumption, cleaning costs
- Facilities utilisation efficiency
- Bed Management Administration

The indicators described above are actively used and the preparation and processing of data needed to calculate many of them require considerable effort as explained in the next section. It was decided to select a few from these metrics and indicators for the purpose of the BI system pilot implementation to demonstrate the BI software functions. The details of the indicators used in the data models designed are also described in the following sections.

5.2.2 Obtaining of the Information

The involvement of hospital managers in obtaining information varies and there are managers with specific responsibilities for obtaining information and data. They also transform the data into a presentable form or analyse the data further based on the values of the main indicators. The interviews revealed two areas of information with different levels of accessibility and the effort needed to repetitively report on certain indicators and metrics.

The financial data including the analysis of cost and revenue are available in the form of a single data file extracted from the main hospital system and processed by statistical software. The output is then presentable and available for further analysis with the use of a spreadsheet application with the pivot functionality. Due to the small number of people involved in the reading and analysis of the financial information the procedure appeared to be efficient and sufficient for the needs of the hospital managers. The spreadsheet application provides all the functionality required for the reporting of related financial indicators, the transformation of the data is done quickly thanks to routines, developed over time, containing the transformation rules and the data extracted from the hospital system includes all required financial details.

The second area is the information related to the patient administration which includes the detail of patient stays in the hospital and the treatment details. This type of information is used to produce the metrics indicating the performance of the hospital departments and individuals as well as the quality of patient care. As mentioned before the information is needed regularly for the clinical audit meetings. It was explained during the interviews that there is not sufficient detail available either directly in the hospital systems nor easily extractable from the hospital system database. For the data processed, based on the available extract, the process of transformation is manual and too laborious. This is also done using spreadsheet software. The reports and charts are however static and the pivot tables and charts are not employed in this case to allow additional flexibility and better data analysis functions. It was pointed out that the processing of the raw data extracted from the hospital system is extremely frustrating and lengthy. The data transformation is not automated in any way and in many cases some of the data needed for calculation of certain metrics has to be manually incorporated, since it is missing from the extracted data. The laborious manual processing of the data can take hours and sometimes even days when comprehensive reports are needed or metrics which involve complex calculations. If any of the indicators require a combination of data from various systems or even from different modules of the same system it is possible to create more extracts from the system. The integration of the data, which is performed manually, is either not possible or the manual integration in the spreadsheet software is too awkward.

When the BI software functionality and potential were briefly presented to the hospital managers who work with both types of information – financial and patient related – they agreed that such software could improve the availability of the information in a "ready-to-use" form with functions useful for rapid visual data analysis. It was however correctly pointed out by some of them that the pivot table functions available in Pentaho are widely available in the spreadsheet software. They would see the main benefit in the seamless automation of the data load and transformation as well as in the calculations of the metrics and indicators predefined in the analytic data models in a central data repository dedicated for analysis and reporting.
More details on the current data transfer and transformation and the procedures which were going to be applied in the BI software implementation are available in the following sections.

5.2.3 Selection of Indicators for the Pilot

From the beginning, the focus of this project was on the data recorded in the hospital systems related to patient care and patient stays in the hospital in general. Also the very first prototypes of reports and analytic data models, created in the Pentaho software, were focused on this type of information. These emerged from the review of the hospitals annual report available to the public which contained information built around the patient admissions.
The indicators selected for implementation during the pilot were selected from the list of indicators and metrics identified in the interviews. The selection of the indicators was also later adjusted due to the availability of the data required for their automatic calculation described in the "BI system implementation" section.

The following metrics were selected for the Pentaho pilot implementation:

- Number of Admissions
- Admissions with Length of Stay over 21 days
- Readmissions within 28 days
- Discharges after 10 AM
- Number of Surgeries
- Surgery Turn Around Time – from start of surgery to end of surgery
- Surgery Turn Around Time – from arrival to theatre to discharge to ward

It was planned to analyse these metrics by the following admission and surgery attributes:

- Service (Department)
- Date Of Event (Admission or Surgery)

- Admission Source
- Clinician
- Patient County
- Patient Gender
- Payer
- Theatre
- Surgeon

These metrics and attributes represent the outcome of the requirements gathering during the interviews and will be used for reporting of indicators and the data analysis.

5.3 Hospital Systems, Tools and Data Flows

The interviews confirmed that there is no system available in the hospital dedicated to reporting, data analysis or providing the functions of BI software. The existing hospital systems provide only basic reporting functionality and cannot be used for the data analysis or to produce metrics and indicators in a presentable form. Where there are reporting functions available in some of the hospital systems, these are limited and are used only to produce predefined static reports which are based only on the data from such systems. There is no integration of the systems available for the purpose of integrated reporting and the information required in the management of the hospital is scattered across various hospital systems or available only in a paper form. Spreadsheet software is used when the data is exported from a single system for further data processing and integration in order to obtain the desired metrics . It allows advanced, user friendly and visual data analysis using pivot tables and charts. This process sometimes involves people from IT department. Each time there is a new file required or new information needs to be added to the file, a request has to be made to the IT department. It also requires that the managers responsible for preparing the metrics and indicators for the management meetings possess knowledge about data processing techniques.
The hospital also does not have a data warehouse (DWH) which could be used for hospital-wide reporting. There is also no data repository (DR) which could be integrating data from all the hospital systems. The main hospital system used for patient administration consists of several modules. The data from the system individual modules is not integrated either. The data can only be exported from each module separately through export functionality from the predefined reports in the particular module and have to be integrated and processed outside the hospital system.
The main hospital system runs on a proprietary operating system and the data management used within the system is also proprietary. This means there cannot be any standard database interfaces

used to extract the data directly from the system. Additional modules would have to be purchased from the system vendor or third party tools which would enable reading of data directly from the system database using an interface such as Open Database Connectivity (ODBC). Purchasing of tools providing such interfaces is being considered in the hospital but was not available for the pilot implementation of OS BI software.

The metrics and indicators which were selected for the Pentaho implementation used data from two modules from the main hospital system. The main analytical data model and reports were based on the patient administration data including the details related to patient admissions and discharges and attributes related to patients and their stay in the hospital. A second group of information was be based on the surgery data. These two groups of information will be loaded from two separate files exported from the patient administration and surgery modules of the hospital system. The definition of the files with the required information was based on the selected metrics and indicators. Additional attributes of events available in these files were also included in the analytical data models to allow further analysis of these metrics.

The load of the data from the extracts into the reporting data repository was designed using the tools contained in the Pentaho BI Suite. The transformation of data and calculations required for the metrics are part of the data load process. If there was a data repository or data warehouse present in the hospital the data from it would either be accessed directly or the loading procedures would be getting the data from DR or DWH instead of the data files.

5.3.1 Hospital IT and Open Source Software

In the interview with the member of the IT department, the experience with OS in the hospital was discussed as well as attitudes to the use of OS software.

He explained that the TCO (Total Cost of Ownership) has to be considered with any new system which would be implemented in the hospital. Although it was accepted that the zero licence costs of the OS are an important factor, the overall cost related to running and maintenance of the OS software is perceived to be more costly due to lack of people on the labour market with the required skills and the limited number of local companies specialising in OS implementation and support, where both are believed to be more expensive than it is with commercial software specialists. It appeared from the discussion that the hospital would probably not want to implement and manage the system only with the hospital IT employees due to limited resources and would prefer a commercial edition of OS BI software with assurances of maintenance and support. It was also the IT team member who pointed out during the interview that the analysis and reporting functions of many reporting and BI tools are very similar in both commercial and open source software and the

more important aspect of the BI implementation is the ability to seamlessly integrate various data sources. It was also emphasised that the reporting tools are perceived by the non-IT managers in the hospital as something that are installed and integrate all the data automatically “out of the box”. The work involved in the data integration, design of the data structures and data models can actually be more important and can eventually represent a higher portion of the system implementation costs. The knowledge of the hospital system data structures and good data integration tools appear to be more relevant. The support and maintenance should also be split into two parts, where firstly there is the support and maintenance related to the software itself and secondly there is a need to support and maintain the content provided through the BI software in terms of the data, data loading procedures and defined data models and reports. Both contribute to the total cost of the software implementation and operation. Yet another characteristic of the software was emphasised during this interview and it was the ability to design advanced calculation and add additional data into the reporting data models which would be in responsibility of the BI system power user. This would keep down any costs related to the additional adjustments to loading procedures and data models and overall any extra involvement of IT department employees or external consultants.

5.4 BI pilot system implementation

The Pentaho BI Suite was initially intended to be installed on a server with Linux operating system so the initial testing of the system before the implementation was done using the Ubuntu Linux distribution. During the interviews it was however agreed that the operating system on the server, on which the Pentaho software will be installed, would be Window Server 2003. The platform independence of the Pentaho which is based on Java Virtual Machine, available on all major platforms, meant that the migration of the system to Windows 2003 was simply a matter of copying the system files and making minor adjustments of the configuration files. Based on the selected indicators and the data files received from the hospital IT department, it was possible to test the deployment of the Pentaho in Windows Server 2003. It was also possible to start building the loading procedures, analytical data models and sample analysis views.

The Pentaho software runs as a web service using Apache Tomcat what allows user access to the system through web browsers. This allows the Pentaho system to be accessed from any computer with a web browser on the hospital network. The Apache Tomcat is used as a lightweight application server and Pentaho was deployed together with the MySQL database system. MySQL is available for Windows Server as well as for Linux systems so the transition from the pre-project sample environment to the project development and test environment was seamless. The most important component of the software implemented was the Mondrian interface which allows the use of pivot

tables in the analysis views to summarise and analyse the data loaded into the database on the MySQL server.

The data loading procedure which included the calculations related to the selected indicators was designed in Pentaho data integration tool. The design of the tables which hold the patient admissions and surgery data was determined by the data files received from the hospital IT department and the calculations required for the indicators.

5.4.1 Data Load and Transformation

The files from which the data was loaded into the reporting database contained patient admission and surgery details where every row of data represented a single event (hospital stay or surgery) and every column was an attribute of the event. The files were extracted from two modules of the main hospital system and the connection between the data in these files was possible as a unique identification number of the hospital stay was available in every surgery record.

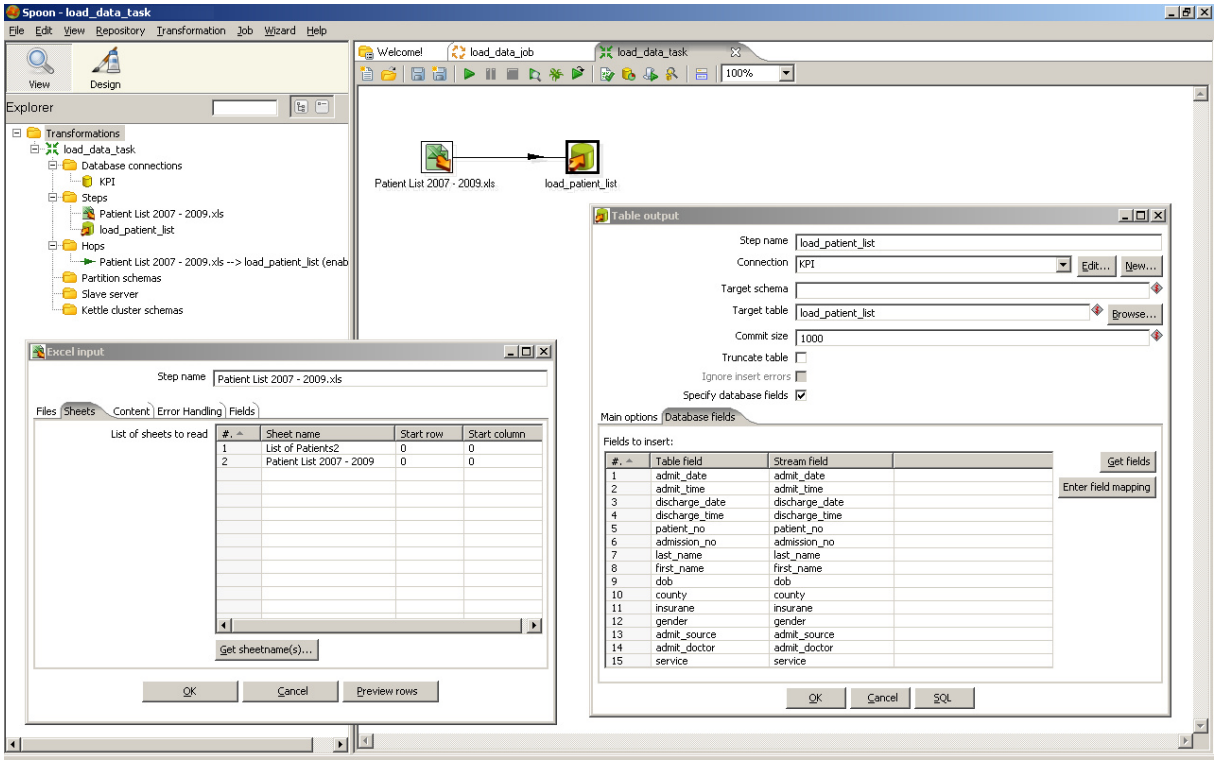


Figure 5.4: Pentaho Data Integration Tool – Definition of Data Load Task

There was only limited amount of data exported into these files for the purpose of the software pilot implementation. The patient admissions data file contained all patient hospital stays for the years 2007, 2008 and 2009. The file with the surgery records contained data for 2009 and 2010. Since all admission details were appended to every surgery record the surgery data available in the system for

reporting and analysis was only from year 2009. The data load procedure designed in the Pentaho Integration tool is captured in Figure 5.4. The screenshot of the transformation procedure in Figure 5.5 contains the Standard Query Language (SQL) expressions which were used to manipulate the data in database tables and calculate the values of attributes used in definition of metrics.

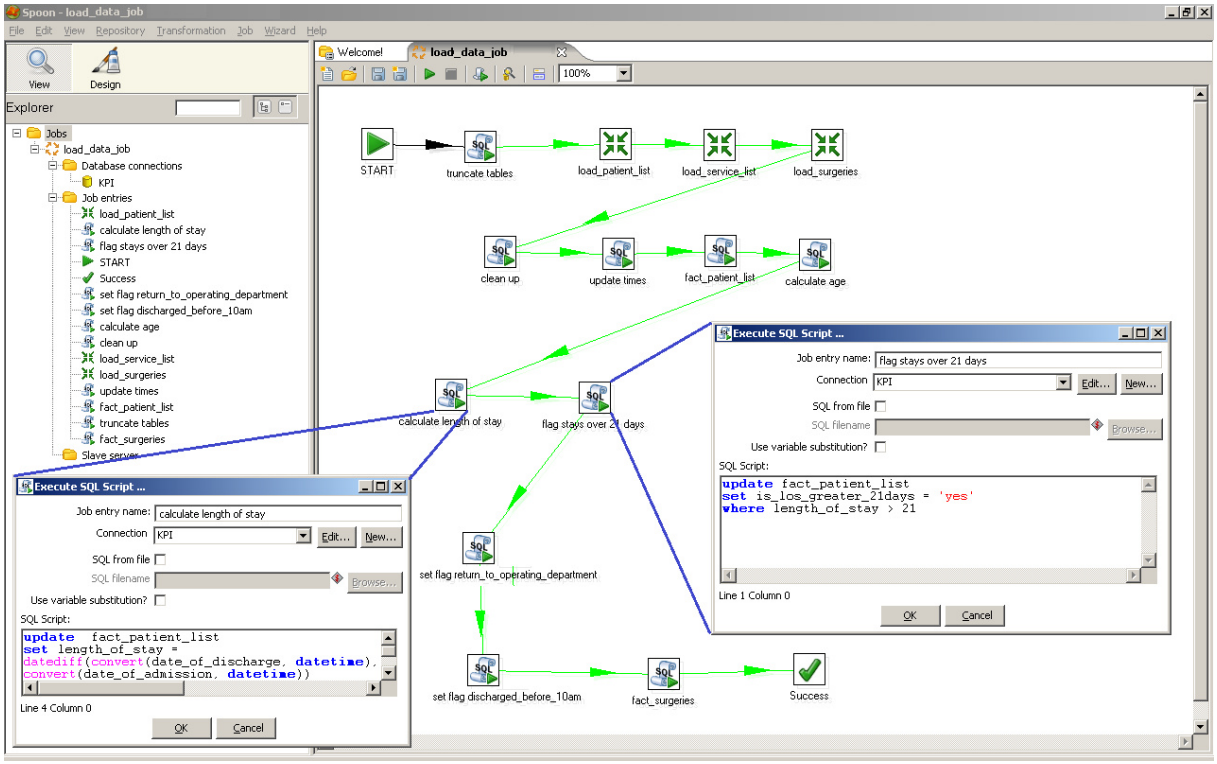


Figure 5.5: Pentaho Data Integration Tool –Data Load and Transformation Procedure

There was no plan to schedule the loading procedure in order to automatically load the latest data extracted from the hospital system. It was agreed that this process could be set up and enabled after the research project had been completed. When the data load procedure containing all necessary transformations and calculations was finalised and tested the data from the files was loaded and available in the data base tables for the design of the analytical data models. The data load and transformation procedure included the following transformation and calculations saving their output into the tables which were used for definition of data models (all attributes used in formulas were available in the source files):

- Length of Stay (LOS)

  = Date of Discharge – Date of Admission

- Length of Stay over 21 days

  = if (LOS >21 days) then "Yes" else "No"

- Return to Hospital within 28 days

  = if (patient discharged in 28 days prior to date of admission) then "Yes" else "No"

- Discharges after 10 AM

  = if (time of discharge > 10:00) then "Yes" else "No"

- Surgery Turn Around Time

  = Time of Surgery End – Time of Surgery Start

The analytical data models which are queried from the analysis views are designed in another Pentaho tool called Schema Workbench. With this tool the dimensions and measures of the analytical data models can be created and the definition of the analytical data models deployed for the use in the Pentaho analysis views. There were two data models developed allowing the querying of patient admissions and surgery data.

5.4.2 Data Modelling

The analytical data models used in the Pentaho analysis views for the querying and analysis of data through the Mondrian OLAP service interface uses the concept of measures and dimensions. Some of the selected indicators are defined by the combination of measure and dimension. There were the following analytical data models designed:

- Admission Analysis Hospital Data
- Surgery Analysis Hospital Data

The full data models descriptions were included in the "Analytic Data Models" document which was provided to the users of the system when it was deployed in the hospital. This document contained detailed definition and content of all dimensions and measures, their use and instructions on how to combine the measures and dimensions in order to obtain the desired metrics and indicators. The document could not be attached in full due to the sensitive data it contained, so only some of the dimensions and measures are described in Appendix E.

The analytical data models were based on the data base tables utilising the Mondrian OLAP interface. These data models were designed in Pentaho Mondrian Schema Workbench tool. The definition of dimensions and measures defines how the content of the database tables is accessed and grouped

when providing results for the queries made from the analysis views. The editing of the “Admission Analysis” data model definition in Schema Workbench can be seen on Figure 5.6.

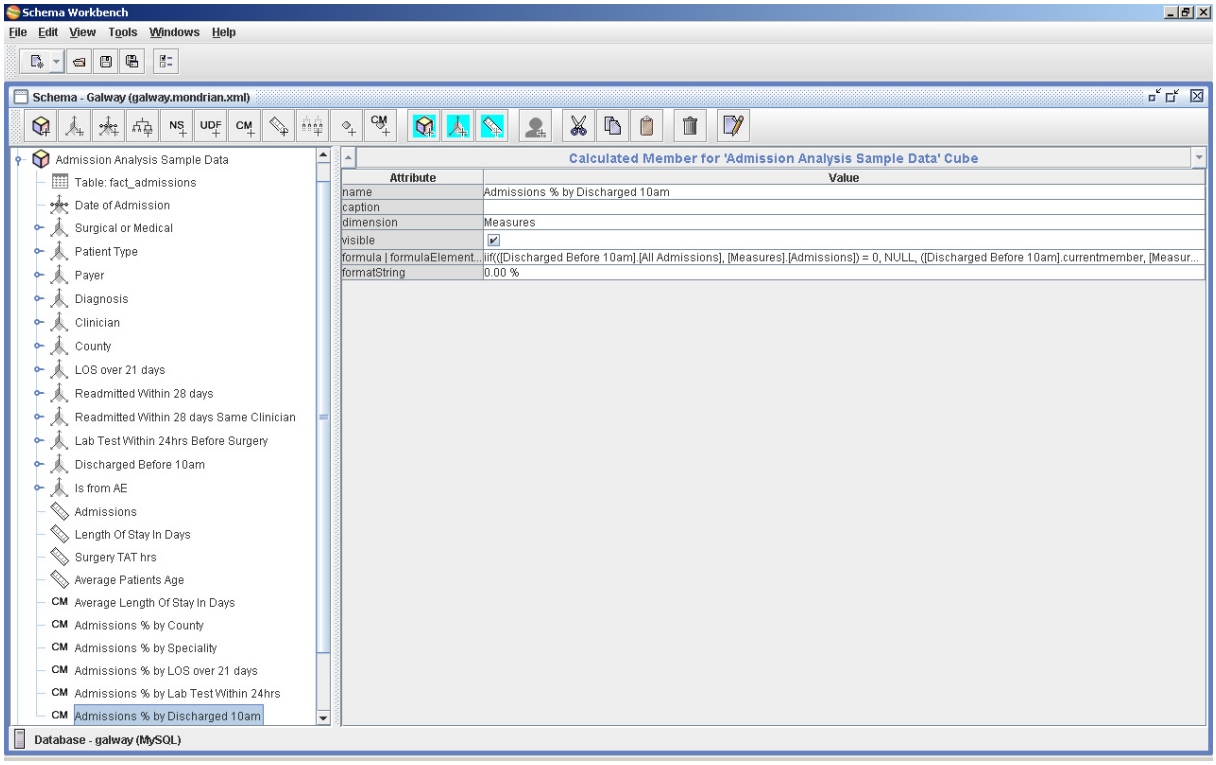
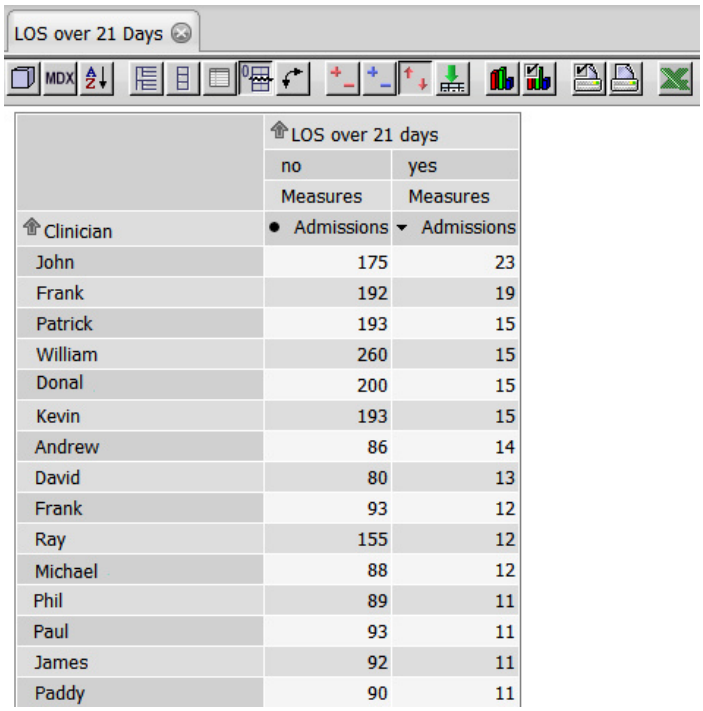


Figure 5.6: Mondrian Schema Workbench – Definition of “Admission Analysis” data model

5.4.3 Reports and Analysis Views

There were a few analysis views designed based on the analytical data models created. These analysis views demonstrated the usage of measures and dimensions defined in the data models in order to provide the selected indicators. The analysis views were configured to include both pivot tables and charts to provide the information with the use of indicators in a visually interesting manner. There were two folders created – one for each area of reports: admissions related and surgery related. Each of the folders contained reports based on the corresponding data model. The predefined views served only as a starting point for the user. Users were encouraged to use them to build their own variations of the analysis views which could be saved independently or just to use them as a starting point for data analysis. As an alternative the analysis of data or design of a new analysis view could be started from the main portal screen by connecting to one of the data models. In such cases all dimensions and measures are placed in the column area of the pivot table. The dimensions and measures can be then rearranged and moved into report rows or report filter.



	LOS over 21 days	
	no	yes
	Measures	Measures
Clinician	Admissions	Admissions
John	175	23
Frank	192	19
Patrick	193	15
William	260	15
Donal	200	15
Kevin	193	15
Andrew	86	14
David	80	13
Frank	93	12
Ray	155	12
Michael	88	12
Phil	89	11
Paul	93	11
James	92	11
Paddy	90	11

Figure 5.7: Analysis View "Length of Stay over 21 days"

The analysis view "Length of Stay over 21 days" shown on Figure 5.7 is an example of the view predefined in the Pentaho. This view can be opened from the list of views when the users logs in to Pentaho portal. This view was created to show the clinicians with most patient admissions for which length of stay exceeded 21 days in the following way:

- The "LOS over 21 days" dimension together was put to the table columns together with the "Admissions" measures
- The "Clinician" dimension was placed to the table rows
- All other dimensions were moved to the filter section without using any of them for filtering
- The view was sorted in descending order by the "Admission" measure under the "Yes" label from dimension "LOS over 21 days"

The following analysis views were prepared for the users when the Pentaho software was deployed:

- Admissions by Service Trend
- Average LOS by Years
- Discharge before 10am by Service
- LOS over 21 days by Clinician
- Readmissions within 28 days by Clinician

- "Arrival" to "Discharge to Ward" TAT by Rooms
- Surgeries by Service in 2009
- Surgeries by Year in Insurer
- Surgery TAT by Surgeon

5.4.4 BI system deployment and configuration

The development environment was running on the laptop under Windows Server 2003. The deployment of the system onto the hospital server consisted of the installation of the MySQL server, restoring database backups and copying the Pentaho files and directories onto the server. The configuration files were adjusted in order to set the server name and enable the running of the system as a Windows Server service. In this way the Pentaho service started on the server with the operating system.

Only a single user was defined for this pilot implementation of Pentaho software with access to all saved analysis views and data models listed in the previous sections. No additional users or roles were planned for this stage due to the small number of the users accessing the Pentaho portal. The access to data contained in the data models related to performance of departments and clinicians did not need to be restricted to a particular group of users accessing the system.

After the system was deployed as a copy of the testing environment the files containing the data extracted from the hospital system were copied to the server file system. The loading procedures were then executed to load and transform the data into the form queried by analysis views through the OLAP interface. The loading procedures required at this stage only minor adjustments to deal with the data validation exceptions and the specific format of the date fields contained in the extracts. The analysis views were then tested to confirm the successful execution of import procedures and the validity of the information displayed.

The Pentaho system was configured to provide the users only the functionality related to the use of analysis views utilising the Mondrian OLAP service. The use of detailed reports based on the tables from the relational database was disabled since this part of the system was not configured and was not intended to be used during this project.

5.5 System demonstration

The Pentaho software was made available on the hospital network and the managers could access the system to display and analyse the information extracted from the hospital systems. The software was then briefly demonstrated to one of the managers who had experience with the analysis of the

data using the pivot tables in the spreadsheet software. This member of the hospital management team was familiar with most of the functionality provided in Pentaho by the analysis views and felt instantly comfortable with the software controls and functions. The overview was hence focused on the information available through the data models created and the predefined analysis views. The data models were explained in detail with the focus on the description of available dimensions and measures. The functionality of the analysis views was demonstrated on the use of these analytical data model elements, pointing out their purpose and the possibilities available though their combinations. The variety of ways of looking at the information was also demonstrated using the predefined analysis views. The comprehensive overview of the portal layout, functions and features related directly to using pivot tables was explained on creating new analysis views from scratch by connecting to one of the available analytical data models.

The system was presented upon its deployment just to this single manager. The other participants of the research – the same managers who took part in the initial interviews – were not available for the system demonstration on the day of the software deployment. The person to whom the system was shown volunteered to give a brief overview of the software to his colleagues.

During the demonstration of the features of the system he expressed his positive impression of the software and although the content presented was not from the area of information he is primarily working with – financial information – he confirmed that he could see benefits in using this as a replacement of processing and analysis of data with the current combination of statistical and spreadsheet software. As he said the software features were intuitively accessible and the response of the system was very quick although working with a smaller volume of data compared to financial data. The ability to switch to the detailed view, showing the data which forms the summary figures, was according to him very useful and well implemented. More details of the feedback collected during the system presentation and in the interviews are described in the next section.

The system was intended to be tried by the hospital managers within a four to six week period. They were provided with detailed documentation of the data models and instructions for their use, together with the Pentaho analysis views user manual customised for use with the created data models. A sample of the user guide is attached in Appendix F. Standard Pentaho Analysis Views documentation was used and customised to include descriptions, examples and screenshots of the data models and analysis views created during this research project. These documents were sent to the hospital managers within the two weeks after the system was deployed, together with the questionnaire for the collection of the feedback from the system use.

The additional presentations of all system features in form of a full overview were carried out with each manager individually on the day when the final interviews were performed. This stage of the research is covered in more detail in the next section.

5.6 System Use Feedback

The feedback from the software trial was collected through the combination of interviews and a questionnaire. After a period of six weeks during which the users were asked to try the system, two more people who were involved in the management of the hospital were interviewed. This was in addition to the interview done on the day when the system was deployed and demonstrated to one of the hospital managers. The questionnaire was sent to four people, two of them were from IT department. The questionnaire was filled and returned back by three hospital managers. One of them was from the hospital IT department. The summary of the research participants and their involvement is presented in Table 5.1.

	A	B	C	D	E	F
First Interview	x	x	x	x	x	
System Demonstration Only			x			
Second Interview with Demonstration	x				x	x
Questionnaire Sent	x		x		x	x
Questionnaire Received	x		x		x	

Table 5.1 Participants involved in various stages of research

5.6.1 Interviews

The feedback collected during the presentation on the day of the system deployment showed that the analysis views and OS BI software included the majority of the functionality expected from the tools for data analysis and data presentation provided via a web-based portal. The other three interviewed people confirmed they were happy with the way the software was working and that it has exceeded their expectations. The researcher emphasised that with such a system, configured for an automatic incremental load and integration of the data, there would be no need to manually export and import any data. The time saved could be better utilised for the analysis of the actual information.

The instant access to charts and cross-tabular reports, available as an output from working with the analysis views, was also considered as a great advantage. These contained up-to date information with the option to review historical data. The access to the information through the web browser, the ability to share the saved analysis views and to configure the user access to various areas of information were also important according to those interviewed. The interviewees stated that with the use of the Pentaho software presented the times required to obtain specific information could be reduced significantly.

The outcome of the interviews was in line with the results of the study related to use of the OLAP technology in healthcare (Tremblay et al, 2007) mentioned earlier in the literature review. The Tremblay's study showed that with the use of the OLAP approach employees involved in the analysis of data and information used in planning had more time to work with the information and did not have "process the data" which was consuming a large portion of their working time before the introduction of OLAP tools.

There were however a few deficits in the software identified during the interviews, some of which reflected opinions and perceptions of the system functionality. Some of those interviewed felt that the use of the "drill-trough" functionality was not clear and could be confusing. This function is used if a summary number is selected, i.e. number of admissions in particular month for a selected department, for displaying of the individual records of these admissions containing the admissions details. Accessing this function was not obvious, the detailed information was displayed in an unexpected place and too many details were displayed at once on the screen. This might have been partially caused by the absence of the user training which had not been possible due to the lack of time of the hospital managers. The managers were provided with the user manual but probably did not have chance to review this user guide which explains the use of the "drill-through" function.

Functionality which was not in the software presented was the "what-if" scenarios where some of the calculations used in the analytic data models would be "parametrizable". Such calculations would affect the figures displayed in the analysis views. Although the Pentaho analysis views do not include such functionality directly, with the combination of the other Pentaho BI components which were not used in this research project it would be possible to provide this functionality.

Another important function which was missing according to one of the managers interviewed was the ability to build user custom calculations using the elements of the analytic data models. This feature, commonly available in similar tools which are part of commercial BI systems, is indeed missing in the software presented. The biggest disadvantage according to the same manager is the inability to include any additional data into the predefined data models and the inability to modify the definition of these data models in general by a non-technical user. The definition of the data model structure indeed requires technical knowledge of the data manipulation and query techniques, for instance the MDX or SQL language. On the other hand there is a Pentaho tool available, although not as a part of the web portal, which allows visual definitions of the data model attributes. This is the Mondrian Schema Workbench, an easily deployable tool in a form of a stand-alone Java based application which was also used in this project for the definition of the analytical data models. The definition of the data included into the analytical data models should, in a full system implementation, include all data available in the source system taking into consideration the information needs of the system end users. Once such a comprehensive data model is designed,

including the calculations of all indicators identified in the requirements gathering phase of the project, their definition should not change frequently over time unless the requirements change or additional data sources are required to be included.

With regard to the web–based interface there was a few times during the testing of the system that the software was expected to provide functionality typical of desktop applications. This is for instance the “drag-and-drop” function which was expected when defining the data models elements displayed in the analysis view in the column header, row header or in the filter. The version of analysis views available in the community edition of the Pentaho software utilising the JPivot open source component does not provide such a function. Although this function might be perceived as more user friendly the advantage of JPivot is in the simplicity, quick response, low computing intensiveness and better compatibility with various web browsers.

5.6.2 Questionnaire

The summary of the questionnaires is based on the three questionnaires which were filled and returned. The questionnaire is attached in the Appendix A. The first page of the questionnaire contained questions related to the information provided in the analysis views specified in the requirements gathering at the beginning of the research project. There were four indicators specified and available in the data models and analysis views. Responses from all three questionnaires were positive where the answers were “Agree” or “Strongly agree” for all four items.

The questions in the second section of the questionnaire asked for the general opinion of the use of the software implemented in this pilot. The questions were related to the improvement of information availability and accessibility, the assistance in decision making, the impact on the performance and the quality of provided services in the case of using the OS BI software. Almost all answers were with “Agree” or “Strongly agree”. One of the responses contained “Disagree” selected for the item related to the suitability of the designed levels of granularity available in the data models and their use in the analysis views.

The section of the questionnaire with response to “General Software Characteristics” contained mostly positive answers to all of the questions with the “Agree” and “Strongly Agree” option selected. The “web-based” nature of the software was marked as less relevant on one of the questionnaires by selecting the “Neutral” option. On one of the questionnaires the response was “Disagree” for the “Intuitiveness without need for user training” and “Displaying the detailed attributes of records ‘behind’ the summary figures”. These two questions are related and the response confirmed the importance of user training. The same responder correctly pointed out that the “user management and user rights management” functions were not present in the software

implemented. This was not because such functions are missing. It was simply due to the small scale of the project where it was not necessary to configure multiple user accounts with different access rights.

The last section of the questionnaire “Analysis View Features” contained again positive responses for almost all of the items on all three responses with “Agree” and “Strongly Agree” options selected. One of the responders disagreed on the suitability of the function for changing of level of aggregation.

The answers from all sections of all responses to the questionnaire indicated the importance of user training. During the implementation of the software for the purpose of this research it was however not possible to gain higher involvement of the participating managers in order to provide such training.

The responses showed that the users agreed with the importance of the software features listed in the questionnaire and more importantly they agreed that these functions as available in the software presented are suitable for use with the patient administration data and information.

Chapter 6: Conclusion

The purpose of this research was to find out whether hospital managers can directly or indirectly benefit from the implementation of Business Intelligence software. The focus was specifically on the parts of the BI Software which allow visual display and analysis of information without the requirement for technical knowledge related to data processing and manipulation. This is provided by the OLAP approach and the use of analysis views consisting of pivot tables and pivot charts.
The research first provided an overview of decision making and decision support systems used in management and healthcare as well as the status of the use of BI tools and OLAP covering both commercial and open source software. The research was carried out as a case study based on the pilot implementation of the OS BI software in a selected private Irish hospital. This pilot implementation focused on the patient administration data and the purpose was to deliver the analytic data models with relevant information on which features of the BI system were demonstrated using the selected set of indicators.

6.1 Findings

The literature review suggested that OS BI software is not implemented in many healthcare environments or directly in hospitals in spite of growing use of other OS software in these environments. It is difficult to estimate the number of installations as these are often not publicly known. The literature review showed many instances of positive impact of the use of BI software and OLAP specifically regarding the availability and accessibility of information, hence their incorporation into systems for decision support.
Based on the literature review the Pentaho BI software was selected for the implementation of OS BI system in the target hospital. After the pilot implementation was finalised and the system was trialled by the hospital managers for a short period of time, the feedback collected at the end of this period showed satisfaction with the functions provided by the system. The hospital managers were pleased with the simplicity of the information analysis provided by the Pentaho analysis views and the availability of data without the need for manual integration and processing. The integration and automation of the data processing appeared to be more relevant than the analysis views, nevertheless the analysis of the data using the OLAP interface to access the processed data was seen as extremely helpful too when compared to the current processes in the hospital used to achieve the same results.

The suitability of the portal based reporting system containing data and information analysis tools could be questioned, given the system is only utilised for a limited set of hospital and patient data by only a few members of the hospital management team. However, if the system was used for hospital-wide reporting and utilised by the hospital employees, at all level of management, such a system would according to hospital managers fit perfectly. It would include all features and functions of the BI software including detailed reports and dashboards which were not presented during this pilot implementation due to the resource and time restrictions.

The research question was defined as: “Can hospital managers benefit from the utilisation of the techniques and tools available in Open Source Business Intelligence software such as the On-line Analytic Processing approach for querying of the hospital data and the use of the pivot tables and charts for the interactive data analysis and presentation?” This question was answered in a positive way and the hospital managers confirmed they could see benefit from permanent implementation of the system in the hospital with the use of the analysis views based on OLAP and there in fact was a confirmed intention to continue using the software after the completion of the research.

The outcome of the interviews confirmed the availability of the information needed by the hospital managers could be improved by the system demonstrated reducing the time needed to produce the required in formation from the available data. This positively answers one of the research sub-questions: “Can the availability of the information in the managing of the hospital be improved by the use of these analytic tools?”

The second sub-question of the research was defined as: “Can this be achieved by implementing Open Source software without increasing hospital annual costs for the software licenses?” The fact that such a system can be implemented and used without incurring licence costs was appreciated which answered this the sub-question and while the licence cost for alternative commercial software can be a significant part of total cost involved in running of the system, this was not seen as a relevant advantage when comparing the Total Cost of Ownership (TCO) of Commercial and Open Source software by some of the research participants.

The limitations of the research and the factors contributing to the questioning of the suitability of OS are described next.

6.2 Limitations of the Research

Despite the acceptance of the presented OS BI solution by the hospital managers there are several factors contributing to the relevance of the findings. These can be summarised into the following points:

- The OS BI system was implemented in a hospital environment without previous use of Data Warehouse, Data repository or Business intelligence software. Due to this it cannot be concluded that the implemented Open Source Business Intelligence is better than any other Commercial or Open Source BI software. It therefore cannot be concluded that the OS BI would be the best match in terms of the hospital management information presentation and analysis needs.
- This research did not intend to carry out a comparison of OS BI tools or Commercial and Open Source Software. The Pentaho BI Software selected was only used as a representative example of OS BI software. The Pentaho BI software seemed to provide the functions of BI software in the best way among the available OS BI systems.
- The focus of the research was on proving that the use of OLAP techniques used together with the analysis views will be the main tools from which hospital managers could benefit. Since few of the managers use the pivot table functionality available in the spreadsheet application, it appears that the integration and automation of data processing including of the definition of the analytical data models might be more relevant and important. The OLAP approach would still be considered as a major improvement, the spreadsheet software could be however used as the front end tool for the users, instead of the presented web portal solution.
- Another advantage of the OS software - the zero licence cost - was questioned during the research by one of the participants. It was argued that while there are no licence related costs, the Total Cost of Ownership (TCO) of OS software can eventually be higher than that of commercial software. The research participant emphasised that more heterogeneous IT infrastructure would include a wide variety of software types and vendors including both Commercial and Open Source Software. The cost of managing such an environment is higher. For instance, this means that if most of the Operating Systems and Database Systems used on computers and servers in the hospital are from a single commercial software vendor, if there are other systems used (operating or database), these would require the IT department members to have greater set of skills, knowledge and experience which would include all the platforms used. The availability of sufficiently skilled people can be lower and these can also be more expensive in terms of labour cost.
- Another cost related aspect is connected to the previous point. There needs to be a decision made in respect of the use of OS software, namely whether the system would be managed internally by the hospital IT department or whether the management and support of OS BI

software would be out-sourced. Also in this case the TOC of running the system has to be examined carefully.

- The research suggested that the use of OS BI software as an Executive Information System is more relevant for the members of the management who are involved in the preparation and analysis of information. Some of the managers, depending on their roles and responsibilities in the management of the hospital would only use the information produced by the BI system but delivered to them in various forms – not by interacting with the system directly. This includes the reports containing tables and charts exported to PDF or Excel format, or included into the Presentations.
- The biggest limitation of this research is the size of the project. The research was carried out as a Case study and with the implementation of the software carried out solely by the researcher the involvement of the hospital staff was limited. This suggests that the results of the research might be more relevant if this project had been conducted as an Action Research project. In such a case the role of researcher could be more focused on the evaluation of the benefits. The project would need to involve a larger set of data and involve more people from hospital management. The implementation of the software itself would need to be separated from the research tasks.
- During this research the accuracy of the information recorded in the hospital systems was not discussed. As the research was focused on the technical solution and the functionality of the implemented software the data recorded in the hospital systems were considered to be accurate.

The hospital information systems use a unique identification number for each hospital patient and also for each patient visit. This allowed the integration of the data from various hospital systems and did not cause any problems. Without the unique patient and patient visit identifiers the integration of the data from various systems and system modules would not have been possible for the information presentation and analysis.

6.3 Future Work

The result of this research project shows that with the use of the Open Source Business Intelligence software the availability of the information used for decision making can be improved. The private hospital selected uses a fairly up-to-date system and yet it did not provide a central reporting and data analysis system with consolidated information from various modules of the Hospital

Information Systems. With the growing awareness of Open Source software there is a greater chance that the new implementations of BI software in hospitals could be based on the OS software. It could be used only for management or clinical information or offer hospital wide reporting.

Future research in this area should concentrate on a more thorough investigation of the use of BI software in the hospitals and OS BI specifically. The research could compare the functions of the various BI systems with the Commercial BI software in more detail. Finally if there will be OS BI software implemented in any private or public hospital as a permanent solution, there should be research carried out and focused on the detailed impact assessment comparing the pre- and post-implementation status of the information availability.

It is likely that the uptake by companies of Open Source Software will increase as a greater number of people with the right knowledge base and experience become involved in the implementation of the OS, which will provide an alternative to commercial software products offering the same or more innovative functionality at better cost.

References

Ament, L. and Hildreth, S. (2009) Hard Times in Healthcare: Business Intelligence Provides Support. Available at: http://www.b-eye-network.com/channels/1021/view/9875/ (Accessed: 24/04/2010).

Armstrong, M. and Baron, A. (2005) Managing performance : performance management in action. 1st Edition. London: Chartered Institute of Personnel and Development.

Bitterer, A. (2009) Open-Source Business Intelligence Tools Production Deployment Will Grow Five-Fold through 2012. Available at: http://www.gartner.com/resId=1210513 (Accessed: 24/04/2010).

BusinessWeek (2009) Business Intelligence for Healthcare: The New Prescription for Boosting Cost Management, Productivity and Medical Outcomes. Available at: http://download.microsoft.com/download/8/A/8/8A89DAA9-BA03-473B-BF57-AED28AFCCF98/Business_Intelligence_in_Health_Whitepaper.pdf (Accessed: 15/11/2009).

Butler, T.W., Leong, G.K. and Everett, L.N. (1996) The operations management role in hospital strategic planning. Journal of Operations Management 14 (2): 137-156.

Carnall, D. (2000) Open Source Software in healthcare. Available at: http://www.carnall.demon.co.uk/OpSrcHth.htm (Accessed: 24/04/2010).

Devlin B., M.P. (1988) An architecture for a business and information systems. IBM Systems Journal 27 (1): 60-80.

Donabedian, A. (2003) An Introduction to Quality Assurance in Health Care: Oxford University Press.

Fitzgerald, B. and Kenny, T. (2004) Developing an information infrastructure with Open Source Software. IEEE Software: 50-55.

Forgionne, G.A. and Kohli, R. (1995a) The decision value of management support systems for strategic hospital management. International Transactions in Operational Research 2 (4): 355-373.

Forgionne, G.A. and Kohli, R. (1995b) Integrated MSS effects: An empirical health care investigation. Information Processing & Management 31 (6): 879-896.

Fottler, M.D. (1987) Health Care Organisational Performance: Present and Future Research. Journal of Management 13 (2): 367-391.

Garand, B. (2009) Open Source Software: The Customer Perception from SME and SMI? School of business. Groupe ESC Rouen. Rouen. 2007-2008. MBA

Giarratano, J.C. and Riley, G. (2005) Expert systems : principles and programming. 4th Edition. Boston, Mass.: Thomson Course Technology.

Gray, P. (1994) Decision support and executive information systems: Prentice Hall College Div.

Greaney, J. (2009) Developing Key Performance Indicators to monitor healthcare quality. School of Computer Science and Statistics. TCD. Dublin. 11 September 2009. MSc.

Gruman, G. (2005) Essential Technology: Open Source CRM and BI. Available at: http://www.cio.com/article/print/14301 (Accessed: 24/04/2010).

Holland, M. (2009) Amalga Unified Intelligence: One New Solution To Many Old Problems. Available at: http://download.microsoft.com/download/4/B/5/4B5AAEE7-0A59-41F3-B9D8-E52C35E703EA/M_Holland_IDC.pdf (Accessed: 11/10/2009).

Holland M., H.J. (2009) Amalga Unified Intelligence: One New Solution to Many Old Problems. Available at: http://download.microsoft.com/download/E/4/E/E4E90982-E49A-4FEA-9DEE-A73FEF65F2B8/Health_Industry_Insights.pdf (Accessed: 09/10/2009).

Inmon, W. (1991) Building the Data Warehouse

Janamanchi, B., Katsamakas, E., Raghupathi, W. and Gao, W. (2009) The State and Profile of Open Source Software Projects in health and medical informatics. International Journal of Medical Informatics 78 (7): 457-472.

Johnson, M.B. (2008) Can preventive medicine save the American health system? Available at: http://www.acpm.org/presidents_column09-09.htm (Accessed: 01/06/2010).

Kelly, J. (2009) Open source BI can extend functionality of commercial software in recession. Available at: http://searchbusinessanalytics.techtarget.com/news/1507026/Open-source-BI-can-extend-functionality-of-commercial-software-in-recession (Accessed: 02/05/2010).

Lohr, K.N., Institute of Medicine (U.S.). Division of Health Care Services., Institute of Medicine (U.S.). Committee to Design a Strategy for Quality Review and Assurance in Medicare. and United States. Health Care Financing Administration. (1990) Medicare : a strategy for quality assurance. Washington, D.C.: National Academy Press.

Madsen, L. (2009a) Business Intelligence: Build or Buy – Considerations for Healthcare Organizations Available at: http://www.b-eye-network.com/view/11223 (Accessed: 24/04/2010).

Madsen, M. (2009b) Open Source Solutions: Managing, Analyzing and Delivering Business Information. Available at: http://www.pentaho.com/products/demos/ben_open_source_bi_report.php?&asset=ben_osbi_report (Accessed: 25/04/2009).

McDonald, C.J., Schadow, G., Barnes, M., Dexter, P., Overhage, J.M., Mamlin, B. and McCoy, J.M. (2003) Open Source software in medical informatics--why, how and what. International Journal of Medical Informatics 69 (2-3): 175-184.

Naranjo-Gil, D. and Hartmann, F. (2007) How CEOs use management information systems for strategy implementation in hospitals. Health Policy 81 (1): 29-41.

Nylund, A. (1999) Tracing the BI Family Tree: Knowledge Management.

O'Brien, J.A. (1999) Management Information Systems: Managing Information Technology in the Networked Enterprise. 4th Edition: McGraw-Hill Inc.

OpenSourceInitiative (2010) The Open Source Definition. Available at: http://www.opensource.org/docs/definition.php (Accessed.

Palaniappan, S. and Ling, C.S. (2008) Clinical Decision Support Using OLAP With Data Mining. IJCSNS International Journal of Computer Science and Network Security 8 (9): 290-296.

Pasumansky, M., Zare, R. and Whitehorn, M. (2005) Fast Track to MDX: For SQL Server 2000. 2nd Edition: Springer.

Pendse, N. (2008a) Consolidations in the BI industry. Available at: http://www.bi-verdict.com/fileadmin/FreeAnalyses/consolidations.htm (Accessed: 24/04/2010).

Pendse, N. (2008b) OLAP market share analysis. Available at: http://www.bi-verdict.com/fileadmin/FreeAnalyses/market.htm (Accessed: 25/04/2010).

Pentaho (2010) Pentaho's commercial open source business intelligence customer successes. Available at: http://www.pentaho.com/about/customers/ (Accessed: 24/04/2010).

Power, D.J. (2002) Decision support systems : concepts and resources for managers. Westport, Conn.: Quorum Books.

Power, D.J. (2007) A Brief History of Decision Support Systems. Available at: http://dssresources.com/history/dsshistory.html (Accessed: 14/02/2010).

Scamuzzo, S. (2009) Open Source Business Intelligence. Available at: http://www.ow2.org/xwiki/bin/download/Activities/EuropeLocalChapterWebinars/ELCWebinarOSBI.pdf (Accessed: 24/04/2010).

Schroeder, R.G., Anderson, J.C. and Cleveland, G. (1986) The content of manufacturing strategy: An empirical study. Journal of Operations Management 6 (3-4): 405-415.

Simon, H.A. (1960) The new science of management decision. [1st Edition. New York,: Harper.

Smith, H.F., M.; Saxberg, B. (1981) Cost Containment in health care: A model for management research. Academy of Management Review 6 (3): 397-407.

Steed, K. (2009) Business Intelligence in Healthcare. Available at: http://www.wavetwo.com/resources/Business_Intelligence_Healthcare_Whitepaper.pdf (Accessed: 09/10/2009).

Tan, J.K.H. (2001) Health management information systems : methods and practical applications. 2nd Edition. Gaithersburg, MD: Aspen Publishers.

Tan, J.K.H. (2010) ADAPTIVE Health Management Information Systems. Sadbury, MA: Jones and Bartlett Publishers.

techWeb (2010) TechEncyclopedia: BI software. Available at: http://www.techweb.com/encyclopedia?term=BI%20software (Accessed: 24/04/2010).

Tremblay, M.C., Fuller, R., Berndt, D. and Studnicki, J. (2007) Doing more with more information: Changing healthcare planning with OLAP tools. Decision Support Systems 43 (4): 1305-1320.

Wanless, S. (2007) Using Business Intelligence to Promote EHR Adoption. Available at: http://www.b-eye-network.com/channels/1021/view/6627 (Accessed: 24/04/2010).

Wanless, S. (2009) 7 Highly Elegant Business Intelligence Applications for Healthcare Providers. Available at: http://www.b-eye-network.com/view/10438 (Accessed.

Waring, T. and Maddocks, P. (2005) Open Source Software implementation in the UK public sector: Evidence from the field and implications for the future. International Journal of Information Management 25 (5): 411-428.

Weins, K. and Cope, R. (2008) A Comparison of Open Source Reporting Tools for the Enterprise. Available at: http://go.openlogic.com/pages/start/download-webinars/index.html?Campaign_Id=1321&Activity_Id=2901&rsc=Webinar.OSSReportingToolsComparison.pdf (Accessed: 24/04/2010).

Wheeler, F. (1994) Executive information systems as repositories of organizational understanding. Conference: In Proceedings of the Second European Conference on Information Systems (Baets W ed.).Location: Nijenrode University, The Netherlands Nijenrode University Press.

Wikipedia (2010a) Commercial open source applications. Available at: http://en.wikipedia.org/wiki/Commercial_open_source_applications (Accessed: 27/04/2010).

Wikipedia (2010b) Data warehouse. Available at: http://en.wikipedia.org/wiki/Data_warehouse (Accessed: 24/04/2010).

Wikipedia (2010c) Expert system. Available at: http://en.wikipedia.org/wiki/Expert_system (Accessed: 24/04/2010).

Wyszewianski, L. (1988) Quality of care: past achievements and future challenges. Inquiry : a journal of medical care organization, provision and financing 25 (1): 13-22.

Yin, R.K. (2009) Case Study Research: Design and Methods. Fourth Edition Edition: SAGE Publications.

Zahedi, F. (1993) Intelligent systems for business: Expert systems with neural networks: Belmont.

Appendix A – Questionnaire (Page 1 of 2)

Use of Open Source Business Intelligence Software in Hospital Management

Questionnaire

As a part of the research I would like to get your feedback on the use of the software and its utility in hospital management. I would like to ask you to complete the forms below. Please fill out this questionnaire based on your experience with the use of the software and your view on the benefits it brings to the managerial tasks in hospital.

Data models and predefined analysis views *For each item, please check if you agree or disagree using a rating scale from "1" to "5". A rating of "1" indicates that you strongly disagree with the statement and a rating of "5" indicates that you strongly agree and "3" is the level where you neither agree nor disagree. Alternatively please select one of the following: "Don't know" or "Not Applicable".* **By use of the designed data models with the analysis views it was possible to analyze the data and to report the following indicators:**	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree	Don't know	Not Applicable
1. Length Of Stay over 21 days (% of all admissions)							
2. Re-admissions within 28 days (% of all admissions)							
3. Surgery Turn Around Time (in hours)							
4. Discharges before 10am (% of all admissions)							

Use of OS BI software in hospital and its impact on hospital management *For each item, please check if you agree or disagree using a rating scale from "1" to "5". A rating of "1" indicates that you strongly disagree with the statement and a rating of "5" indicates that you strongly agree and "3" is the level where you neither agree nor disagree. Alternatively please select one of the following: "Don't know" or "Not Applicable".*	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree	Don't know	Not Applicable
1. The OS BI software can improve the way the information needed in the managing of the hospital are obtained and accessed in organisation where central reporting solution utilising on-line analytical processing is not used							
2. The use of presented software can improve availability of information and have positive impact on improvement of the efficiency of the hospital resource planning and allocation							
3. The use of OS BI software can provide support of decision making process in managing of hospital							
4. Information provided via OS BI software can play beneficial role in performance management and service quality monitoring							
5. Summary level of information with the option to break down the summary figures to lower level of detail suits your needs in terms of information availability							

General software characteristics *For each item, please check if you agree or disagree using a rating scale from "1" to "5". A rating of "1" indicates that you strongly disagree with the statement and a rating of "5" indicates that you strongly agree and "3" is the level where you neither agree nor disagree. Alternatively please select one of the following: "Don't know" or "Not Applicable". The statement is defined in the column header.*	**The feature of a reporting and data analysis software is relevant for the use with the hospital administration data/information:**							**Demonstrated OS BI software includes the feature in a form suitable for the use with the hospital administration data/information:**						
	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree	Don't know	Not Applicable	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree	Don't know	Not Applicable
1. Data presentation in form of cross-tabulation using pivot tables and charts														
2. Visual attractiveness - data presentation in form of charts														
3. Ability to save and share saved analysis views and organise them in folders														
4. Intuitiveness and ease of use without necessity of user training														
5. Interactivity and flexibility – predefined views easily adjustable														
6. Instant result when changing the definition of the analysis views														
7. Reliability of displayed information thanks to inability to alter the source data														
8. Central point of data – single and consistent version of truth														
9. Web based interface – easily accessible from any computer														
10. Instant access to summarized data without need to process the data by the end user														
11. User/role based access rights – password protection for access to sensitive data														
12. Ability to see "what is behind" the summary figures by displaying of the detailed data														

Appendix A – Questionnaire (Page 2 of 2)

Analysis view features	**The feature of analysis view is important and/or useful for the use with the hospital administration data/information:**							**Demonstrated OS BI software includes the feature in a form suitable for the use with the hospital administration data/information:**						
For each item, please check if you agree or disagree using a rating scale from "1" to "5". A rating of "1" indicates that you strongly disagree with the statement and a rating of "5" indicates that you strongly agree and "3" is the level where you neither agree nor disagree. Alternatively please select one of the following: "Don't know" or "Not Applicable". The statement is defined in the column header.	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree	Don't know	Not Applicable	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree	Don't know	Not Applicable
1. Dimension navigator – place or remove categories and measures to or from pivot table														
2. Dimension navigator – Set items from categories not used in the view layout as filter														
3. Sorting with option to display top X categories														
4. Visual styles selection (show parent categories, hide spans, hide empty cells)														
5. Swap axes														
6. Change of the granularity level by drill-down and drill-up functions														
7. Displaying of records which compose the summary figures by drill-through function														
8. Automatically created charts based on the pivot table content														
9. Sufficient number of chart attributes for the customisation and chart styles														
10. Export to PDF and XLS														

Name: Date: QUESTIONNAIRE - Use of Open Source Business Intelligence Software in Hospital Management 2010 | Page 3 of 3

Questionnaire Cover Letter

Dear XY,
I hope you are well.
As a part of my research I installed the software for reporting and analysing of patient administration data on the server on the network in your hospital. There have been reports and data models designed and made available in this software based on discussion I had with you related to performance & service quality indicators including:

- *Discharges before 10am*
- *Surgery TAT*
- *LOS greater than 21 days*
- *Re-admission within 28 days by same clinician*

I was wondering if you could try the system for a little time and fill in the questionnaire in attached "Questionnaire Opene Source Business Intelligence v2.pdf". Please print the document, fill in the questionnaire on paper and leave it with XY, I will ask her to collect them from all people, scan and send back to me. In case you find some of the items from questionnaire too detailed please feel free to complete only the part of the questionnaire which you find relevant.

The software runs in form of a service and is available as web portal at address:
http://pentaho/
where you can click on the Login button and enter the following to log in to the system:
user name: user
password: healthcare

The details about the use of the web portal can be found in the attached document "Pentaho_Analysis_Viewer_User_Guide.pdf" and the detailed information about the data models used in the reports and analysis views are available in attached document "Analytic Data Models.pdf"

I was hoping to get the filled questionnaires back before Friday 11th June with possibility to meet with you (based on your availability) on that day for a quick talk. Please let me know if you think this would be possible.

Please let me know in case of any questions and many thanks for your help.

Kind Regards,

Michal Miklas

Appendix B – Semi-structured Interviews Agenda

Topics for Semi-structured Interviews

Use of Open Source Business Intelligence Software in Hospital Management

Pre-implementation interviews

- Identification of information relevant for the decision making in the hospital management
- Reasons of why the information is needed
- How the information is accessed and what is the form of the information
- Required level of data aggregation
- When and how often is the information required?
- How long does it take to obtain the information if needed outside predefined periodicity and what is the procedure to achieve this?
- What is the origin of the information?
- Can the information be produced from the data recorded in the hospital systems?
- Is the information updated periodically and available in a presentable form?

Post-implementation interviews

- Data presentation - pivot tables and charts
- Ease of navigation / use of the tool
- Interactivity – available options and features to adjust the predefined report and view on the data
- Availability and instant access
- Central point of data and access – only one version of truth
- Access from any computer – web based interface
- Timeliness of the information
- Dynamic reports with regard to time – reports can be preconfigured to show last X days, weeks or months by default
- Refresh frequency, hourly, daily, weekly, on request
- Sufficiency of level of the data granularity

Technical interview

- Is there a DWH which includes the PAS data for the purpose of reporting and data analysis?
- How are the information extracted from data in DWH?
- Is the OLAP technology employed?

Appendix C – Consent Form

Consent Form

Use of Open Source Business Intelligence Software in Hospital Management

Researcher: Michal Miklas

The purpose of this research is to find out if the hospital executive members can benefit from the implementation of Business Intelligence software. It will be focused on the On-line Analytic Processing (OLAP) approach for querying of the hospital data and the use of the pivot tables and charts for the interactive data analysis and presentation. The research will attempt to prove that the availability of the information in the managing of the hospital can be improved by these analytic tools. The aim of the research is also to demonstrate that this can be achieved by implementing Open Source software without increasing hospital annual costs for the software licenses.

The research will be based on a pilot implementation of selected Open Source Business Intelligence software in your hospital. The research project will contain description of the tools utilized presently and the way of information extraction from the hospital systems and its translation and presentation. Once the Open Source Business Intelligence software is in place and used for some time by the hospital executive members the software will be assessed. The assessment of the implemented software will be carried out with the emphasis on the impact on the information availability.

The information will be acquired in the form of the semi-structured interviews. The anticipated duration of a single interview is 60-90 minutes. There will be notes taken during the interviews and also after the interview which will be later analyzed. The full content of the interviews will not be recorded in any way. Individual opinions and comments might be included in the dissertation but this will be done anonymously, so the name or the role of the interviewed person will be not included in the dissertation.

DECLARATION:
I am 18 years or older and am competent to provide consent.
I have read, or had read to me, this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
I understand that I may refuse to answer any question and that I may withdraw at any time.
I understand that my participation is fully anonymous and that no personal details about me will be recorded.
I understand that if I or anyone in my family has a history of epilepsy then I am proceeding at my own risk.
I have received a copy of this agreement.

PARTICIPANT'S NAME:
PARTICIPANT'S SIGNATURE:
Date:

Statement of investigator's responsibility: I have explained the nature and purpose of this research study, the procedures to be undertaken and any risks that may be involved. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

RESEARCHERS CONTACT DETAILS:
INVESTIGATOR'S SIGNATURE:
Date:

Appendix D – Information Leaflet

Information Leaflet

Use of Open Source Business Intelligence Software in Hospital Management

The purpose of this leaflet is to give you an overview of this research project. The aim of the research is to find out if the hospital executive members can benefit from the implementation of Business Intelligence (BI) software. It is focused on the On-line Analytic Processing (OLAP) approach for querying of the hospital data and the use of the pivot tables and charts for the interactive data analysis and presentation. The OLAP tools and techniques are used for querying of large data sets usually available in data warehouse or central data repository which integrates data from various hospital systems. These tools allow non-technical users to extracts the information using the predefined analytical data models in the form of a self-service information extraction. The OLAP in combination with the front-end tools included in the BI software provides the decision support tools required in the processes of the hospital management.

The research will attempt to prove that the availability of the information in the managing of the hospital can be improved by these analytic tools and the following can be achieved:

- make the information available in the most appropriate form in a timely manner and easy accessible
- make the reporting and analyzing data visually attractive and interactive
- improve the efficiency of the hospital resource planning and allocation

The following attributes of the BI software make the accessibility of the information simpler:

- Central data repository & Single point of access via Web portal (client-server architecture)
- Summary & Detailed information combined seamlessly
- Presentation of summary information, trends, charts, KPIs in a visually attractive way
- Flexibility: working with various levels of granularity
- Interactive: ease of modification of the report content by filtering

The aim of the research is also to demonstrate that this can be achieved by implementing Open Source software without increasing hospital annual costs for the software licenses. It was decided to focus in this research on the OS software also to prove that the OS BI software is matured in terms of simplicity of use and the user comfort as well as the seamless process of implementation.

The research will be based on a pilot implementation of selected Open Source Business Intelligence software in your hospital. The research project will contain an assessment of the current set of tools used for the extraction of the information from the hospital systems and its translation and presentation. An assessment of the implemented Open Source Business Intelligence analytic tools will be conducted when the software is deployed and will have been demonstrated to and used by the hospital executive members. While assessing the implemented software the emphasis will be put on the impact on the information availability.

The information will be acquired in the form of the semi-structured interviews.

The interview questions will be focused on the following topic:

- Identification of information relevant for the decision making in the hospital management
- Reasons of why the information is needed
- How the information is accessed and what is the form of the information
- When and how often is the information required
- How long does it take to obtain the information if needed outside predefined periodicity
- What is the origin of the information
- Can the information be produced from the data recorded in the hospital systems

Admission Analysis Hospital Data

This data models is based on the list of patients admissions exported from the hospital system. The data model consists of set of dimensions and measures. The dimensions are based on the attributes of the patients' admissions which can be used of their grouping. The measures are predefined numeric indicators which can be displayed for each group. The basic measure is the count of the items from the list which falls into a group or can be based on a calculation using numeric attribute of patient admissions. Data from years 2007-2009 is available for analysis in this data model.

The following dimensions defined in this data model are based on the basic attribute of each admission:

- Service
- Date Of Admissions
- Admission Source
- Clinician
- County
- Gender
- Payer

There were also the following dimensions created in this data model which are based on calculated attributes of patients' admissions:

- Discharged Before 10am
- Is NTPF Referral
- Is from AE
- LOS over 21 days
- Readmitted within 28 days
- Readmitted within 28 days Same Clinician
- Surgical or Medical

There have been the following measures created in the data model:

- Admissions
- Average Patient Age
- Average Length of Stay
- Admissions % by County
- Admissions % by Clinician
- Admissions % by LOS over 21 days
- Admissions % by Discharged 10am

Surgery Analysis Hospital Data

This data models is based on the list of surgeries exported from the hospital system. The data model consists of set of dimensions and measures. The dimensions are based on the surgery attributes which can be used of their grouping. The measures are predefined numeric indicators which can be displayed for each group. The basic measure is the count of the items from the list which falls into a group or can be based on a calculation using a numeric attribute of surgery. Data from year 2009 is available for analysis in this data model.

The following dimensions defined in this data model are based on the basic attribute of each surgery:

- Service
- Date Of Surgery
- Admission Source
- Clinician
- County
- Gender
- Payer
- Room
- Surgeon

There were also the following dimensions created in this data model which are based on calculated attributes of surgeries:

- Is NTPF Referral
- Is from EA
- Surgical or Medical

There have been the following measures created in the surgery data model:

- Surgeries
- Surgery TAT
- Arrival to Ward TAT

Dimension Discharged Before 10 am

This dimension is based on calculation using the time of patient admission available in the data exported from hospital system. The calculation examines the time of the patient admissions and the result can be the word Yes or No indicating whether the patient was discharged before 10 am (Yes) or the patient was discharged after 10 am (No). It contains the following categories which can be used for groping by adding the dimension to the rows or columns of the analysis view or used in the view filter:

- Yes
- No

This dimension can be used freely in combination with any other dimension. The recommended use of the dimension is with the following measures:

- Admissions
- Average Patient Age
- Average Length of Stay
- Admissions % by Discharged 10am

The combination of this dimension with measure "Admissions % by Discharged 10 am" can be used as performance or service quality indicator.

Dimension LOS over 21 days

This dimension is based on calculation using the date of patient admission and discharge available in the data exported from hospital system. The calculation examines the length of stay based on these two dates and the result can be the word Yes or No indicating whether the length of patient stay was longer than 21 days (Yes) or the length of stay was shorter or equal to 21 days (No). It contains the following categories which can be used for groping by adding the dimension to the rows or columns of the analysis view or used in the view filter:

- Yes
- No

This dimension can be used freely in combination with any other dimension. The recommended use of the dimension is with the following measures:

- Admissions
- Average Patient Age
- Average Length of Stay
- Admissions % by LOS over 21 days

The combination of this dimension with measure "Admissions % by LOS over 21 days" can be used as performance or service quality indicator.

Dimension Readmitted within 28 days Same Clinician

This dimension is based on calculation using the patient ID and historical admission records with the same patient ID considering the admitting clinician of historical admissions. The calculation examines historical admissions within 28 days prior the admission date and the result can be the word Yes or No indicating whether the an admission of the same patient by the same clinician was wound within previous 28 days (Yes) or there was no admission by the same clinician within previous 28 days (No). It contains the following categories which can be used for groping by adding the dimension to the rows or columns of the analysis view or used in the view filter:

- Yes
- No

This dimension can be used freely in combination with any other dimension. The recommended use of the dimension is with the following measures:

- Admissions
- Average Patient Age
- Average Length of Stay

The combination of this dimension with measure “Admissions” can be used as performance or service quality indicator.

Measure Admissions

The measure is defined as count of admission records available in the data exported from hospital system.

Measure Average Length of Stay

This measure is calculated using the admission and discharge dates available in the data exported from hospital system. The Average Length of Stays is calculated as difference between these dates in days.

Measure Admissions % by Clinician

This measure is calculated as percentage of admissions for particular clinician from all clinician admissions. This measure should be used with dimension Clinician displayed on rows or columns.

Measure Admissions % by LOS over 21 days

This measure is calculated as percentage of admissions Yes or No (for dimension “LOS over 21 days”**)** from all admissions. This measure should be used with dimension “LOS over 21 days” displayed on rows or columns.

Measure Admissions % by Discharged 10 am

This measure is calculated as percentage of admissions Yes or No (for dimension “Discharged before 10 am”) from all admissions. This measure should be used with dimension “Discharged before 10 am” displayed on rows or columns.

Pentaho Analysis Views

When you open any of the predefined sample analysis views or create new by connecting to the analytic data model the analysis view is opened in the main pane of the Pentaho user console. There can be more than one view opened at the same time and the user can switch between the views using the tabs. The User Interface consists of a toolbar for configuring settings related to how data is visualized and interacted with, and an interactive data grid.

} Toolbar

					Measures
Admission Source	Payer	Gender	Clinician	County	Admissions
All Sources	All Payers	All Genders	All Clinicians	All Counties	9

} Data Grid

Looking at the example above, you see the data is made up of several dimensions including Admission Source, Payer, Gender, Clinician, County and Measures. Dimensions are natural hierarchies that exist within your admission data which we can use to slice and dice the data to get different perspectives of the information or drill down for additional details. The Measures dimension is a special dimension that contains all of the numbers (or facts) can be used in hospital management. The Admission Analysis data model contains several measures such as Admissions, LOS, Average Patients Age.

Interacting with the Data Grid

Drilling allows users to expand the hierarchies within a dimension to uncover additional detail. To drill down on a dimension member, simply click the plus sign next to the member name you wish to drill on.

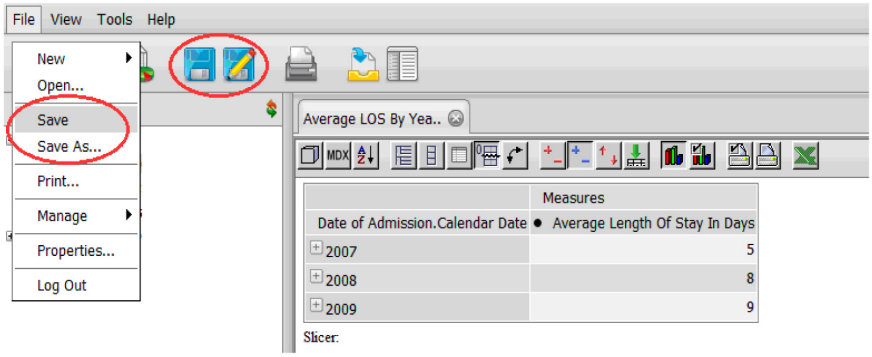
			Measures				
			Admissions				
			Payer				
Surgical or Medical			All Payers	All Payers			
(All)	Surgical or Medical	Service		BUPA	ESB	VHI	VIVAS
All Services			9	9	2	5	3
All Services	Medical		3	5	3	3	1
	Medical	MEDICAL	3	5	3	3	1
	Other		1	8	5	7	4
	Other	DENTAL	5	3			2
		DERMATOLOGY	5			5	
		UROLOGY	1	2	1	2	2
		THORACIC	8	2		6	
	Surgical		5	6	4	5	8

After drilling down on a particular member, you will notice that the plus sign becomes a minus sign. Clicking on the minus sign allows you to collapse the expanded level. In the example above, we have expanded the All Services and All Payers dimensions to reveal the members in the first level of each dimension. The toolbar contains several options for controlling the drill method including expansion drilling and member drilling. We will cover each of these options in detail later in this document.

Analysis View Toolbar

The Analysis View toolbar contains various settings for controlling the data grid display, the drill method, charting, and more. This section will describe each toolbar item in detail. There is a description associated to each toolbar icon displayed when the mouse cursor is stopped over the icon.

Save and Save As

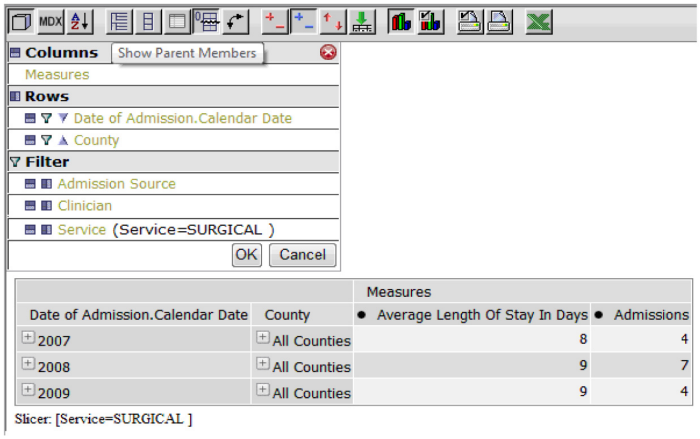


The Save button allows you to save any changes made to the current Analysis View. The Save As button will allow you to save changes as a new Analysis View.

OLAP Navigator

The OLAP Navigator is used to define the overall layout of your query like which dimensions to display as rows, which to display as columns, and which members to filter on.

In the example below, we have placed the Date Of Admission and County in the rows area of the view. We have placed the Measure dimension in the columns area of the view. Finally, we have placed (left) the Service dimension in the filters section of the view and selected to filter on the member 'SURGICAL'. This means that as users interact with the data grid, all queries will return only facts related to the SURGICAL cases.

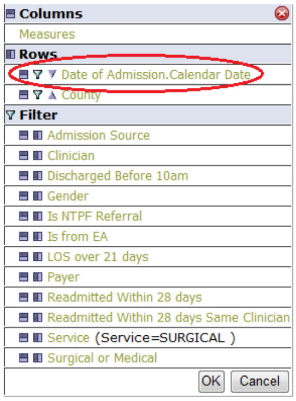


OLAPNavigatorButtons

Icon	Description
	Move Dimension to the Columns area
	Move Dimension to the Rows area
	Move Dimension to the Filters area
	Move Dimension Up or Down within the current area

LayingOutYourDimensions

Pentaho Analysis Views contain three areas in which to place your dimensions: rows, areas and filters. The Filters area, also referred to as the slicer, allows you to filter data based on a dimension member not seen in the row or column area of the view. Moving a dimension from one area of the view to another area is as simple as clicking on the appropriate icon next to the dimension you wish to move. For example, to move the DATE OF ADMISSION Dimension from the Rows area to the Columns area, click on the  button next to Time.



DATE OF ADMISSON has now moved to the top, and drills across the page.

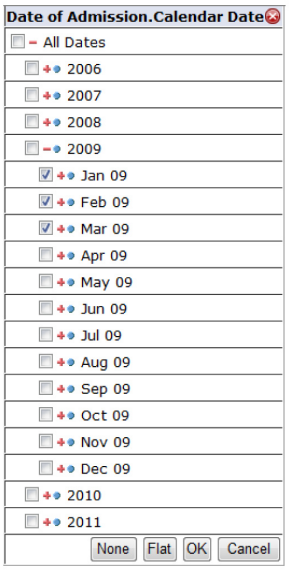
	Measures					
	Average Length Of Stay In Days			Admissions		
	Date of Admission.Calendar Date			Date of Admission.Calendar Date		
County	2007	2008	2009	2007	2008	2009
All Counties	8	9	9	44	87	44

Slicer: [Service=SURGICAL]

The order of dimensions displayed on the rows/columns can be modified by clicking one the up/down arrow displayed before the dimension name in the OLAP navigator.

BrowsingDimensions

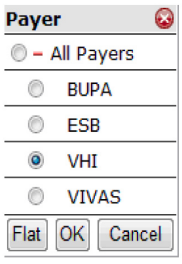
You can also define specific selections within a given dimension using the dimension browse feature of OLAP Navigator. To browse a specific dimension, simply click on the dimension name in OLAP Navigator. Here is an example of browsing the Date of Admission dimension to select first three months of the year 2009:

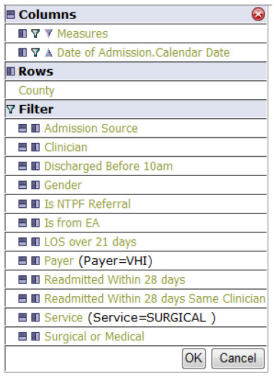


The order of selected items can be modified by clicking on the blue spot next to the plus sign for the item you want to move. Then click on the arrow sign which is displayed instead of blue spots for the item before/after which you want to move it.

CreatingaFilterExample

In this example, we will filter all data on the VHI member from our Payer dimension. In the OLAP Navigator click on the ▽ icon next to the Payer dimension in order to move this dimension into the filter area.
Next, click on the Payer Dimension to open the dimension browser. Click the + symbol next to All Payers to display its children. Then select the radio button next to VHI and click OK to finish making your selection.
Your OLAP Navigator should now look like this:





Click OK to exit the OLAP Navigator and notice that your Data Grid is updated to reflect your new layout and selections. The filter we defined can be seen in the Slicer panel just below the Data Grid.