InImpact: The Journal of Innovation Impact: ISSN 2051-6002 : http://www.inimpact.org
Special Edition on Innovation in Medicine and Healthcare : Vol. 6. No. 1 : pp.121-130 : imed13-012

Business Intelligence Strategy for Data Warehouse in Andalusian Health Service

Ortega A.[1], Holgado M.[1], Doña J.M.[2,3]

[1]Andalusian Health Service. Andalusian Council of Health and Social Welfare, Spain
[2]Dept of Languages and Computer Sciences. University of Málaga, Spain
[3]Health Management Area North of Málaga, Spain

Abstract

The Andalusian Health Service (SAS) has developed a strategy to establish and integrate a Business Intelligence System and a Enterprise Data Warehouse in its Information Technologies Structure which allow the acquisition of analytical information in their systems and get knowledge about the different business areas of a health environment.

The main issues for the implementation of the management information system are the quality of the information contained therein and the establishment of a methodology to provide information to business people who make the decisions, in which the methodologies, versatility and human relationships are the core values to consider. In order to get these proposals new techniques based in imputation methods, context-based exploitation and aggregation of information have been studied.

1. Introduction

The Andalusian Health Service (SAS) is one of the most important Spanish public health services for population and healthcare volume. The SAS has more than 7.9 million records with clinical data. In addition, information systems connect 3.500 pharmacies and over 17,000 health professionals. Also, the available data of users are billions and they are scattered among different systems and applications. This picture hinders the ability to generate the knowledge that supports decision making processes.

In this environment it is necessary to manage huge volume of strategic information where the following problems are produced:

- there are a large number of information systems
- with vast amounts of data
- the information arises from different areas (healthcare, financial, logistic..)
- dispersed in several resources systems

To solve these issues, the SAS has developed a data warehouse system that allows the acquisition of an analytical understanding of the information in their systems and provides new tools to generate knowledge about the different business areas. A data warehouse is a subject-oriented, integrated, time-invariant, non-updatable collection of data used to support management decision-making processes and business intelligence. It is a repository where all relevant data of an organization are placed and from which emerge the information and knowledge needed to manage the organization [8, 9]. This system is supported by a business intelligence module (BI) which allows the technological shift from database to data warehouse and on-line analysis processing (OLAP). The BI module assists decision making process on all levels of management:

- On the strategic level: BI makes possible to set and follow specific objectives.
- On the tactical level: BI system provides tools for decision making. This system is able to optimize future actions and to propose modification in areas like organisational, financial or technological policies in order to help and support the strategic objectives.

The BI and OLAP systems are developed to eliminate the gap between two levels of management: strategic and tactical vision, for creating a proactive business atmosphere and generate an integrated decision support environment (IDSS).

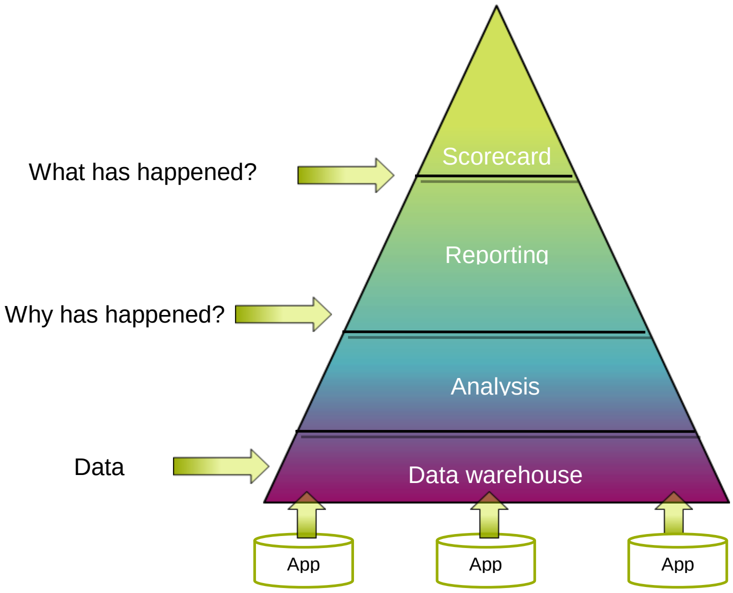

Figure 1: Analytical model

The analytical model of the IDSS system is defined using a pyramid structure of four levels (Figure 1):

- Scorecard: Provides easy access to the stored information, which is relevant to managers through key success factors.
- Reporting: Allows creating and managing different tools to report and ad hoc inquiring.
- Analysis: Enables expert users to discover patterns, generalizations, regularities and rules in data resources.
- Data Warehouse: Collects the information from different data sources and organizes it consistently.

The most important reasons that support implementation of the IDSS system are:

i. Facilitate access to applications including graphic and multimedia interfaces to provide users information in a comfortable and accessible way.
ii. Enable data integration from different systems and resources, adding processes to eliminate the inconsistency in data models. The data load processes feed the BI system through an exhaustive debugging information process.
iii. Provide data homogenization to a common standard. The source systems are on different platforms and models, which are independent from each other, and must be normalized.
iv. Improve data quality and uniqueness of the information. There is no single truth, solutions can be multiple and different. The user must be able to access the information analysis from different points of view, which allow comparing the results and achieving a higher level of objectivity.
v. Ensure the quality of the data coming from transactional systems, because they are the core of potential aggregated indicators and reports for decision making.
vi. Provide a robust technology platform to ensure correct implementation and evolution of the system, allowing users to exploit available information optimally.

The objective of this paper is to present the development of a business intelligence strategy for data warehouse in the context of a large healthcare system with large volume of data and technological diversity. The paper reviews the main techniques to create an integrated decision support environment and shows the main problems and solutions applied in the different project phases. In order to do this, the paper is set out as follows. Some general considerations about integrated decision support with the definition of main problems and solutions are presented in Section 2. Section 3 presents the main results obtained in the developing of the business intelligence system. Finally, Section 4 draws out our conclusions.

2. Integrated decision support environment. Problems and Solutions

The integrated decision support environments are usually allied with datamining tools. Most of the datamining enhanced IDSS fall into the category of specific application-oriented systems coupled with data warehouses and OLAP [15], integrating datamining into IDSS, plus methods and tools, and applying to business problems in a collaborative setting. Most IDSS integrated with data mining technology are found in medical problem domains [3, 10, 11], where the extraction, transformation, and loading (ETL) functions in a data warehouse are considered the most time-consuming and expensive portion of the development lifecycle [19]. The main problems of these systems are the missing data, the data quality, the exploitation of data warehouses, the human-computer interface and the definition of an operational and management model.

2.1 Missing data

The missing data problem is relative with (i) the extraction of data from legacy systems, (ii) transformation and preprocessing requirements to produce useful, integrated data, and (iii) the transportation of the data into the actual data warehouse structures. Often such operational systems were not designed to be integrated and data extracts are performed manually or on a schedule determined by the operational systems. As a result data in the data warehouse may reflect different states of different systems. This multiplicity of information sources makes the resulting system particularly susceptible to missing data. Missing data can significantly affect the performance of predictive risk modelling, an important technique for developing medical guidelines. The two most commonly used strategies for managing missing data are to impute or delete values, and the former can cause bias, while the later can cause both bias and loss of statistical power [5].

In order to solve this problem different imputation methods are described [5, 7], where an optimal option is the combination of fuzzy logic with ordered weighted averaging aggregation (OWA) operator [17]. These techniques must be applied in the Data warehouse level, before data analysis.

2.2 Data quality

Data quality is a major concern for many operational systems as well as data warehouses [13, 19, 20]. Validation of accuracy, timeliness, completeness, and consistency remain major problems for many organizations even in internal information systems where users are trained and managed by the organization. These problems are multiplied in information systems that are exposed to customers, vendors, and other partners. The result can be a disaster for a data warehouse that depends on such systems for its content. Mechanisms for protecting a data warehouse from poor quality data are crucial [4].

In this case a fuzzy approach in data analysis level may be the solution [12]. Through utilisation of triangular fuzzy numbers the system adds the vagueness, ambiguity and uncertainty prevalent in real word systems. The model integration include three common steps: (i) search for candidate models, (ii) benchmark candidate models and (iii) apply the selected model to evaluate the data quality.

In this phase, The ETL process is developed using Oracle Warehouse Builder (OWB). Oracle Warehouse Builder is a comprehensive toolset designed for data warehouses. This tool facilitates the consolidation of heterogeneous data sources from legacy systems, adding capabilities for metadata data modelling. OWB offers a graphical environment to build, manage and maintain data integration processes in business intelligence [16]. Figure 2 shows the basis architecture of Oracle Warehouse Builder with data quality.

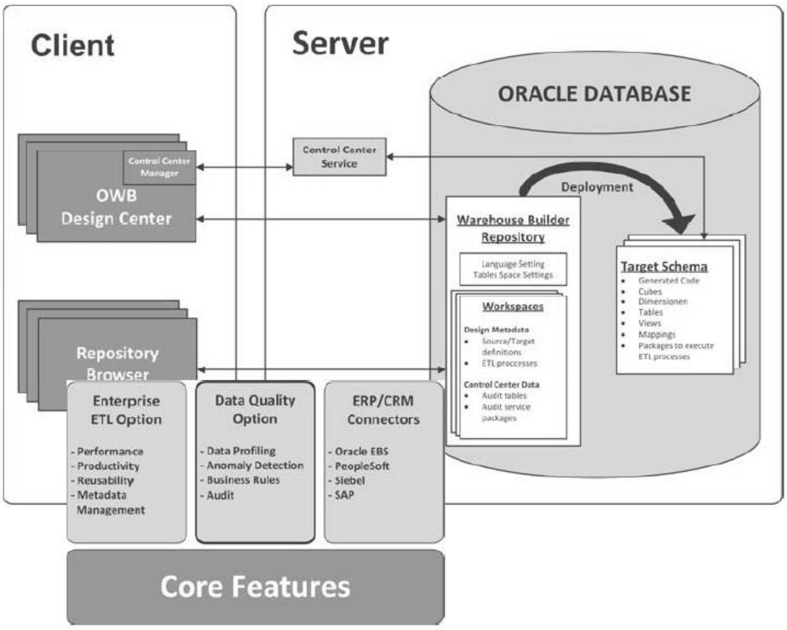

Figure 2. OWB Architecture with Data Quality Option.

2.3 Exploitation of data warehouses

In the area of OLAP exploitation of data warehouses, there is a need for organizing, reusing, sharing queries, in order to simplify and speedup the querying process [6]. Broadly speaking, it can be said that contextual information during the exploitation of data warehouses must be taken into account in analysis and reporting levels [1, 2].

To provide the system with a powerful analysis tool the Quiterian DDweb system is used together with Microstrategy 9i to provide self-service and agile advanced analytics for business users [14, 18]. The user can define and store OLAP queries in a context, where they can organize so that they are easily browsed in a subsequent session. MicroStrategy allows users to define metadata (called schema objects), such as attributes, facts, tables, and hierarchies. These metadata are mapped to the data warehouse structures and they are stored by MicroStrategy in a relational database in a proprietary format. The schema objects are used to convert user requests into SQL queries. With this tool, users can develop application objects too; these objects, as metrics, prompts and filters, are the building blocks for creating reports and documents and they are shared among applications. In addition, OLAP queries in a given context can be imported into another context to enrich the user's current analysis.

2.4 Human-computer interface

The human-computer interface is of paramount importance in the data warehouse environment and the primary determinant of success from the end-user perspective. In order to support analysis and reporting tasks, the data warehouse must have high quality data and make those data accessible through intuitive interface technologies.

The OWB, Microstrategy and DDweb provide a graphical environment to build, manage and maintain the data, with browsing tools which provide query-like access through a flexible menu-based interface, with pull-down menus representing important dimensions. These types of tools are easy to use and support some ad hoc exploration, but are usually controlled through an administrative layer that determines the data available to end-users. Figure 3 shows the final infrastructure with the Business Intelligence tools and Data warehouse. The end-user framework includes the scorecard, and the tools for reporting, analysis and datamining.

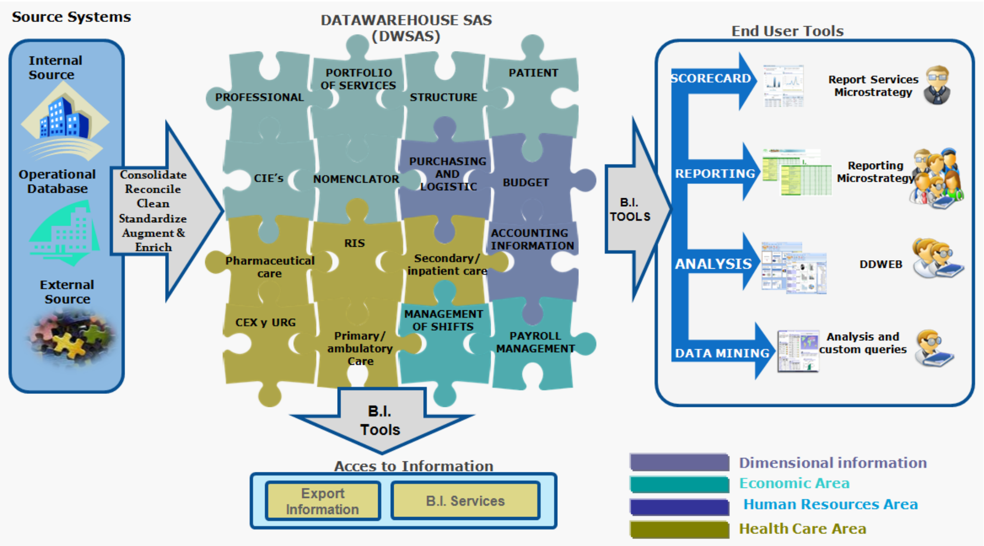

Figure 3. Final Business Intelligence System and Enterprise Data Warehouse

2.5 The operational and management model

The implementation of a BI system in an organization is a very complex process and requires large amounts of resources. To ensure the best results an agile methodology is proposed, where the following characteristics must be accomplished:

- Short development cycles: to deliver working software frequently, from a couple of weeks to a couple of months, with the shortest time interval between deliveries.
- Increase the functionality in each iteration: modifications must be allowed in all development process. The evolutive changes give an edge to decision makers.

With these restrictions, the defined operational and management model is based on the implementation of the catalogs of applications defined by the functional groups. Each catalog request generates a work order, so that these concepts may be used interchangeably. One strategy is that the size of each work order will be reduced to ensure fast delivery. The management model is defined like a continuous cycle of six phases: Business and data understanding; data preparation and modelling; evaluation and deployment; where the data warehouse is the focus of the system (figure 4).

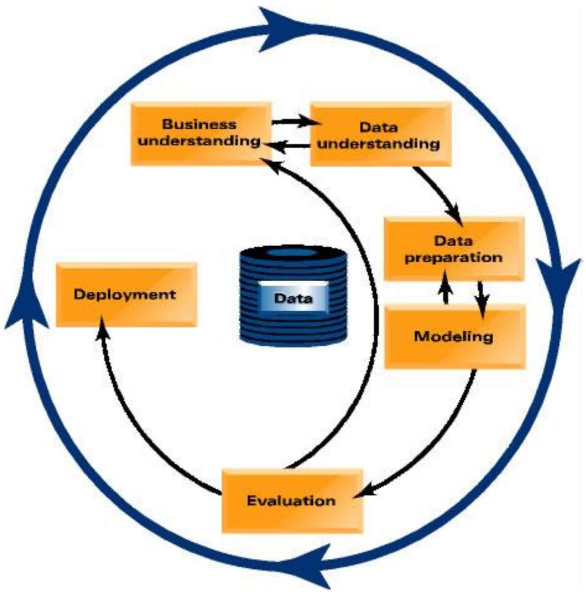

Figure 4. Phases of Change Management

This model allows many advantages:

- Perform a block prioritization
- Greater control over the allocation and work order
- Facilitate the monitoring of the implementation

3. Results

One of the main results obtained in the developing of the business intelligence system and enterprise data warehouse to generate the integrated decision support environment is the specification of the critical factors detected to create the operational and management model.

Identification of key business indicators. The success of the system is not based only on technological investment. It is needed to know the business infrastructure, the team and the business policy.

Identification of users and types of analysis. A number of users and queries under-estimated produce system failures that result in a loss of confidence in the system implanted. It is needed to make a deep study about the users and queries for the system.

Identification of information and data sources. Right identification of data sources is the basis of the BI system. It is should identify if the data will be delivered in batch or real time and what kind of "cleaning" systems will be applied to the data.

Planning and test execution by functional users. The approval of the features and information contained in the BI system is critical to the success of the system.

Technological Analysis: Not all BI tools and applications are valid for all users, It is indispensable to analyze the different types of users and assign the tools and strategies to their needs to take full advantage of the information available.

As a result of the implementation of the BI strategy the following documents are generated:

Corporate Data Catalog, which includes the data and metadata model of Data Warehouse with business entities, dimensional analysis and relationships.

Corporate Indicators Catalog, with the different business areas that will be implemented in the dashboards.

Business Rules Catalog, where the processes of Extraction, Transformation and Loading the Data Warehouse are included.

User Manual, with the definition of indicators, access profiles, reports, features and system use.

4. Conclusions

The Business Intelligence System and Enterprise Data Warehouse developed is constituted as a source of the different management information systems and data marts to cover all the necessities of analysis in the different levels of the health organization, from dashboard to top management to reporting and dynamic analysis for analyst profiles. To develop the IDSS we have studied the main techniques to create an integrated decision support environment. The problems and solutions applied in the different levels of the analytical model have been analyzed and the main solutions for managing missing data and data quality; tools for OLAP exploitation and the keys for the definition of an operational and management mode have been exposed. Also the main results obtained in the developing of the business intelligence system in the Andalusian Health Service have been presented.

With IDSS system other intangible benefits of implementing a BI strategy are obtained, with a improvement of data and better transparency of information flows and knowledge management, which enable to organisation:

- Cost reduction to increase performance of the ICT infrastructure of the organization and improve employee productivity due to the better availability and quality of information using a single repository of information for the different areas of the organization.
- Quality assurance of the information in the management information systems through consensus between functional managers.

- A global model that allows the interconnection between areas and the uniqueness of the information.
- Support the implementation of strategic business objectives.

Acknowledgement

This research is supported by the Council of health and Social Welfare of the Regional Government of Andalusia, Spain.

References

1. Akaishi, M. and N. Spyratos. Discovering implicit relationships in a web of contexts. Intuitive Human Interfaces for Organizing and Accessing Intellectual Assets, pp. 175–188 (2004).

2. Akaishi, M., N. Spyratos, and Y. Tanaka. Contextual search in large collections of information resources. EJC: European-Japanese Conference on Information Modelling and Knowledge Bases, pp. 295–302 (2003).

3. Bellazzi R, Larizza C, Magni P, Bellazzi R. Temporal data mining for the quality assessment of hemodialysis services. Artif Intell Med vol 34:1 pp 25–39 (2005).

4. Berndt, D., Hevner, A., and Studnicki, J., The CATCH Data Warehouse: Support for Community Health Care Decision Making. Decision Support Systems, Vol. 35, pp. 367-384 (2003).

5. Cismondi F, Fialho AS, Vieira SM, Reti SR, Sousa JM, Finkelstein SN. Missing data in medical databases: Impute, delete or classify? Artif Intell Med. Vol58:1 pp 63-72 (2013).

6. Dittrich, J.-P., D. Kossmann, and A. Kreutz. Bridging the gap between olap and sql. In VLDB ’05: Proceedings of the 31st international conference on Very large data bases, pp.1031–1042. (2005).

7. Doña J.M., Quintana O.P., Valesani M.E. Vallejos O.A. Analysis of Aggregation Methods in Incomplete Database Systems. 12 Int. Conf. Information Processing and Management of Uncertainty in Knowledge-Based Systems. Proceedings of IPMU'08, pp. 282-288 (2008).

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

9. Inmon, W, H. and Hackathorn, R. D., Using the Data Warehouse, New York: Wiley (1994).

10. Kurgan LA, Cios KJ, Tadeusiewicz R, Ogiela M, Goodenday LS Knowledge discovery approach to automated cardiac SPECT diagnosis. Artif Intell Med Vol 23:2 pp 149–169 (2001).

11. Kusiak A, Dixon B, Shah S. Predicting survival time for kidney analysis patients: a data mining approach. Comp Biol Med Vol 35:4 pp 311–327 (2005).

12. Lin HY, Hsu PY, Sheen GJ. Afuzzy-based decisionmaking procedure for datawarehouse system selection. Expert Syst Appl Vol 32 pp 939–953 (2007).

13. Liu, S. and Duffy, A.H.B. and Whitfield, R.I. and Boyle, I.M. Integration of decision support systems to improve decision support performance. Knowledge Information Systems, 22 (3). pp.261-286. (2010).

14. MicroStrategy 9i, Basic Reporting Guide, MicroStrategy, www.microstrategy.com.

15. Mladenic D, Lavrac N, Bohanec M, Moyle S. Data mining and decision support: integration and collaboration. Springer, Berlin. (2003).

16. Oracle, Oracle Warehouse Builder - Data Quality Option, http://www.oracle.com /technology/products/warehouse/pdf/warehouse-builder-11g-dataqualitydatasheet. pdf. Version: 2007a

17. Peláez J.I., Doña J.M., La Red D. Fuzzy Imputation Method for Database Systems. Handbook of Research on Fuzzy Information Processing in Databases. Vol 2 805-821 (2008).

18. Sallam, R.L., Richardson J., Hagerty J., and Hostmann B. Magic Quadrant for Business Intelligence Plataforms. Gartner Group, 27: 1-31 (2011)

19. Srivastava, J. and Chen, P. Warehouse Creation – A Potential Roadblock to Data Warehousing, IEEE Transactions on Knowledge and Data Engineering, Vol.11: 1 (1999).

20. Wang M,Wang H, Xu D,Wan KK, Vogal D. AWeb-service agent-based decision support system for securities exception management. Expert Syst Appl 27:439–450 (2004).