Journal of Digital Imaging

Business Intelligence Tools for Radiology: Creating a Prototype Model Using Open-Source Tools

Luciano M. Prevedello,[1] Katherine P. Andriole,[1] Richard Hanson,[2] Pauline Kelly,[2] and Ramin Khorasani[1]

Digital radiology departments could benefit from the ability to integrate and visualize data (e.g. information reflecting complex workflow states) from all of their imaging and information management systems in one composite presentation view. Leveraging data warehousing tools developed in the business world may be one way to achieve this capability. In total, the concept of managing the information available in this data repository is known as Business Intelligence or BI. This paper describes the concepts used in Business Intelligence, their importance to modern Radiology, and the steps used in the creation of a prototype model of a data warehouse for BI using open-source tools.

KEY WORDS: Data collection, data mining, databases, data extraction, knowledge management, online analytical processing (OLAP)

BACKGROUND

R adiology administrators and managers need data from various electronic resources for use in supporting decisions and analyzing trends. This is not only true for assessing efficiency and financial performance, but increasingly useful for monitoring quality and safety to meet and exceed regulatory, compliance[1] and reimbursement[2] requirements. Radiology Information is available in multiple systems (examples include Picture Archiving and Communication System—PACS, Computerized Physician Order Entry—CPOE, Radiology Information System—RIS, Report Generation System, Electronic Medical Record—EMR, financial systems such as billing accounts receivable applications). Individually or combined, these databases can provide vital departmental metrics, but because this data is stored in multiple and disparate data silos, this poses a challenge from an informatics organizational point of view.[3] In addition, analysis of such data is difficult because: (a) detailed data is not easily accessible for quick analysis; (b) summary reports cannot be created ad-hoc (programmers and database experts are required to build specific queries); (c) conversion to one unified database is difficult or lengthy without appropriate tools; (d) the coding of data is often inconsistent across the databases; (e) frequent access to databases can negatively affect their performance and thus hamper daily operations.

In order to take full advantage of the information, all system reports should ideally be available in one single format. In addition, summary reports need to be generated within a reasonable amount of time so decisions can be made in a timely fashion.

Since it is not possible to have one single application responsible for all computerized activities of a hospital, standards of communication and integration must be adopted and utilized. Virtually
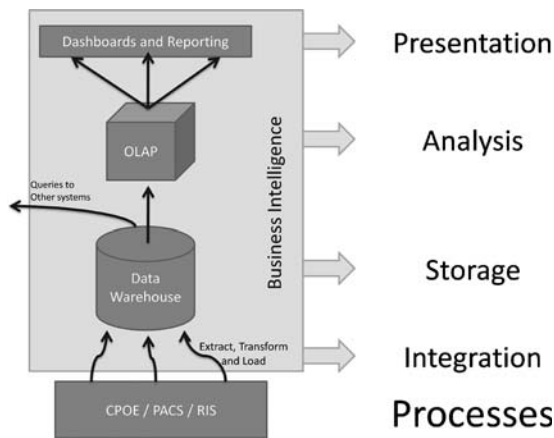
Fig 1. Diagram demonstrating the steps involved in Business Intelligence and its related processes.

queries of databases as new questions appear in an environment marked by constant change. This capability is certainly important for healthcare and these tools have become more available for medical specialties such as Radiology.

Business Intelligence refers to the tools needed to integrate, store, analyze and present data from non-integrated sources (Fig. 1). Integration is a key step in this process as this is where data from different sources are checked for consistency and subsequently converted into a unified format. The process of integration is known as Extract Transform and Load (ETL) in Business Intelligence terminology (Fig. 2). The next step is to store this organized information in a data warehouse, often using a relational database management system. Relational databases are the most popular model currently in use on commercial and open-source databases. Details are represented in tables with a set of relationships applied to the data. However, this model is not ideal for ad-hoc analysis since results are not easily understood without processing the data. Additionally, working with large databases can be time-consuming from an analytic standpoint.

Another way of organizing the data, which is much more suitable for analysis and reporting is by

merging relevant data from various sources into one single format and location could contribute to improved speed and efficiency of knowledge discovery. This concept is not new and has been used for many years in the business world through the adoption of data warehouses.[4,5] Tools under the umbrella of Business Intelligence have been developed to take advantage of data repositories making it possible to perform real-time ad-hoc

Fig 2. Diagram demonstrating the steps involved in the integration of multiple databases.
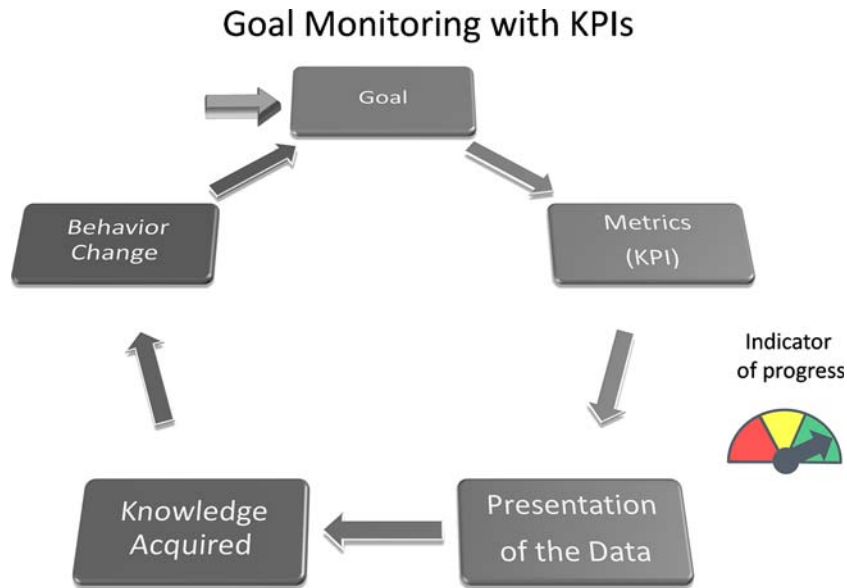
Goal Monitoring with KPIs

Fig 3. Diagram showing process improvement being monitored by Key Performance Indicators (KPIs).

using an On-line Analytical Process (OLAP). In this method, data is represented by data cubes instead of tables and organized in a multidimensional aggregated format. This format allows fast analysis by turning raw data into a format that is more easily understood by the user. Relational databases can be connected to OLAP in a variety of ways using both open-source and proprietary tools.

One of the objectives of representing the data from relational databases in OLAP tools is that it facilitates the generation of ad-hoc queries and supports real-time analyses, allowing users to perform searches in a faster and more structured fashion and to generate on-demand graphs and reports more easily. The goal is to have a more detailed knowledge of the factors that are involved in a particular process or procedure, and to possibly understand how multiple processes interact with each other. This may uncover issues impeding operations and procedures within the organization, and point to potential solutions.

The overall analysis starts with a set of proposed goals, such as improving MRI scanner throughput or reducing radiology reporting turn-around times. Objective metrics or key performance indicators (KPIs) must be defined. Next, the resource or resources that contain the relevant KPI must be understood so that reports assimilating the data and information can be generated. The knowledge

acquired through this process can then be used to adjust or change current behavior, and to help the organization move towards optimal processes or goals (Fig. 3).

Tools designed to present KPIs such as Balanced Score Cards[6] or Quality Dashboards have been used in healthcare.[7,8] These tools allow various activities to be monitored within the hospital environment. However, the generation of summary reports can be time-consuming without the appropriate infrastructure. Business Intelligence tools can help optimize this workflow.

To demonstrate the use of Business Intelligence in Radiology and to start investigating the feasibility of using new KPIs specific to this field, a data warehouse prototype environment was created using open-source tools.

METHODS

The test environment consisted of a PC running Microsoft Windows XP. Pentaho Data Integration 3.0 was installed as the Extract Transform and Load (ETL) software and MySQL server 5.0 as the supporting database. Pentaho is an open-source application with reporting, analysis, dashboard, data-mining, and ETL capabilities for Business Intelligence.[9] For the purpose of creating a
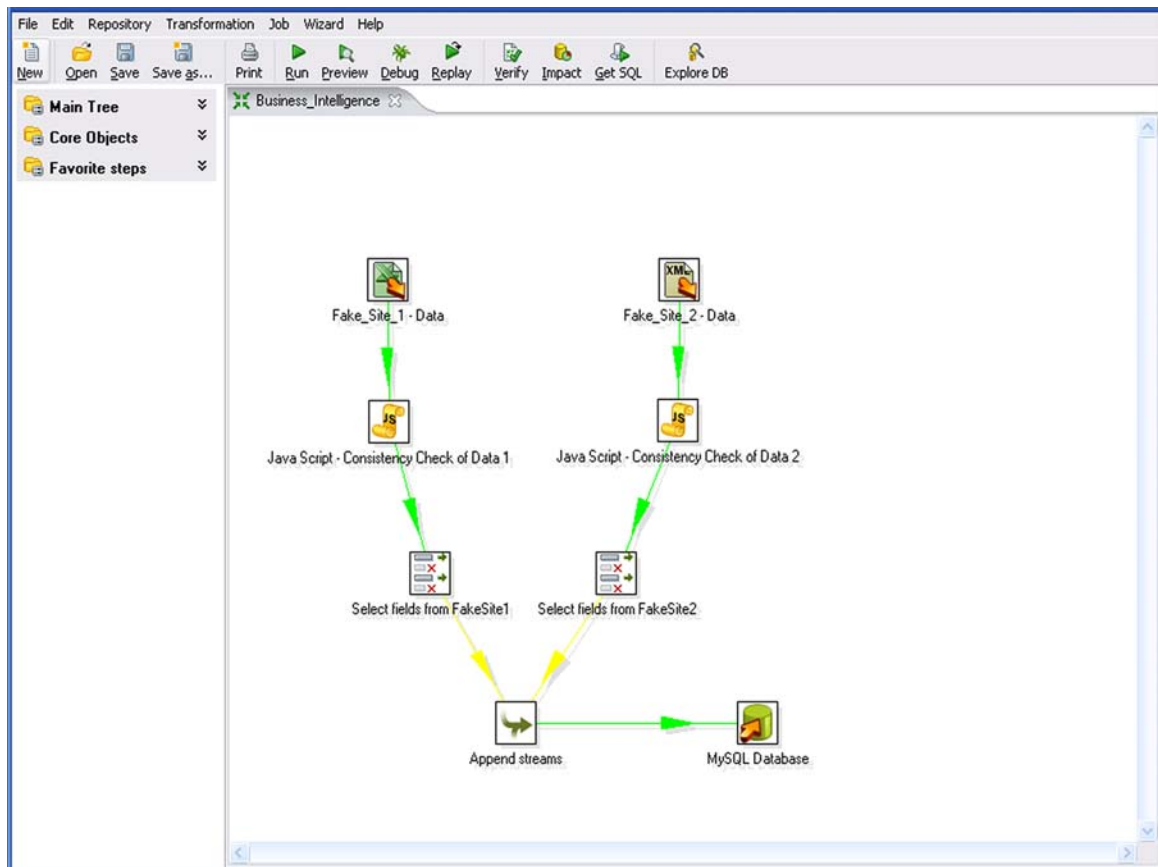
Fig 4. A screenshot of the ETL tool used to combine two databases into one single format.

Radiology Data Warehouse prototype, only the ETL component in Pentaho was used. Pentaho Data Integration is a JAVA-based application that allows administrators to create complex transformations and jobs in a graphical, drag-and-drop environment without having to generate any custom code. This application can connect to multiple commercial and open-source database platforms. An open-source database (MySQL Server) was chosen to connect to Pentaho. This integration can be seamlessly done within the application by providing the name of the server host, the database and the authentication information (administrator's username and password). The relational database (MySQL) serves as a data repository for all the transformations done inside the ETL software which will then be consumed later by other applications. Multiple options are available as input data. The input information can come from other databases such as MySQL, PostgreSQL, Oracle, Microsoft SQL Server, and others. Input can also come from separate files such as Excel, Access, or XML.

A micro-data-warehouse was created combining information from two different sources of fictitious test data. The following formats of input files were used: XML and MS Excel (XLS). The ETL software extracts specific designated fields from each of the input files, converts the data into a unified Structured Query Language (SQL) format, and loads it into one single database (Fig. 4). In this example two input files with similar content but with different configurations were used. Although both archives had the variables ID, date, modality, resource, and exam code, the name of the columns were different in both files. This was done to realistically simulate the problem that one faces when two separate databases need to be fused into one. In addition to different column headings, the date column in the database must be
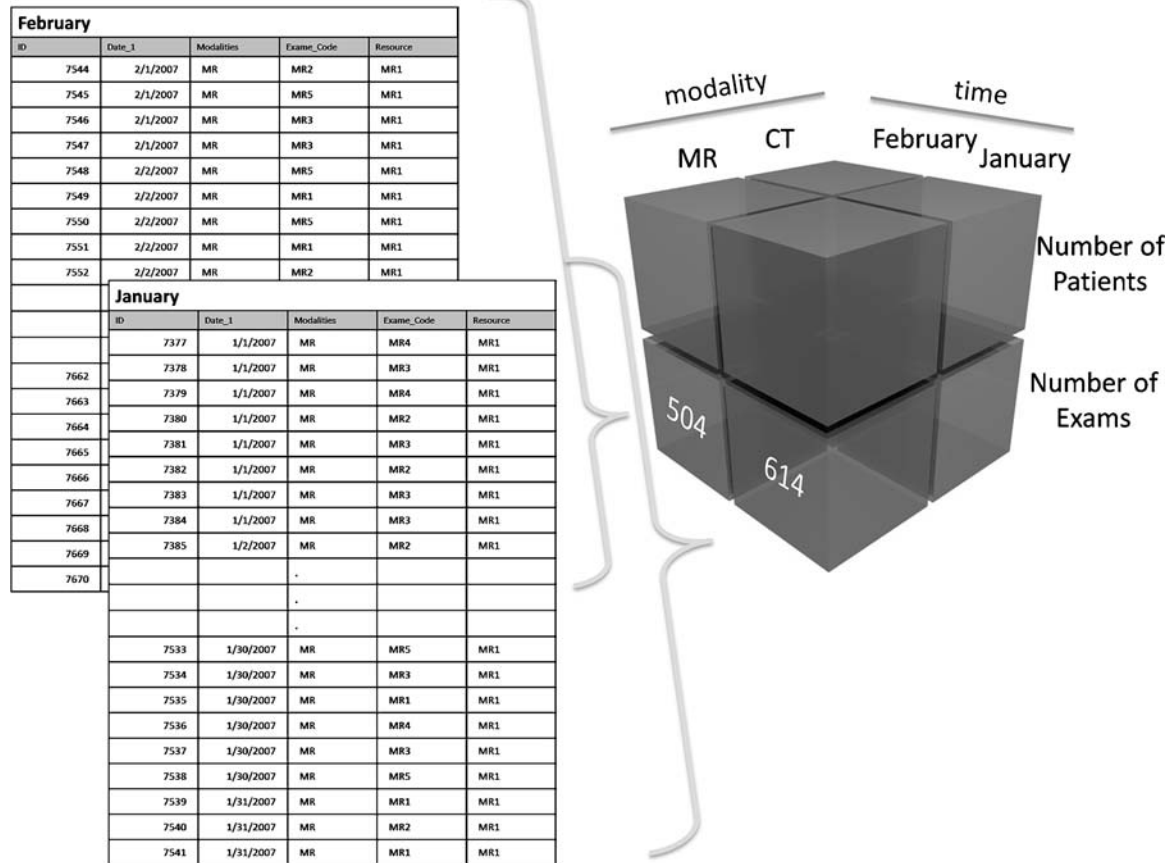
February

ID	Date_1	Modalities	Exame_Code	Resource
7544	2/1/2007	MR	MR2	MR1
7545	2/1/2007	MR	MR5	MR1
7546	2/1/2007	MR	MR3	MR1
7547	2/1/2007	MR	MR3	MR1
7548	2/2/2007	MR	MR5	MR1
7549	2/2/2007	MR	MR1	MR1
7550	2/2/2007	MR	MR5	MR1
7551	2/2/2007	MR	MR1	MR1
7552	2/2/2007	MR	MR2	MR1

January

ID	Date_1	Modalities	Exame_Code	Resource
7377	1/1/2007	MR	MR4	MR1
7378	1/1/2007	MR	MR3	MR1
7379	1/1/2007	MR	MR4	MR1
7380	1/1/2007	MR	MR2	MR1
7381	1/1/2007	MR	MR3	MR1
7382	1/1/2007	MR	MR2	MR1
7383	1/1/2007	MR	MR3	MR1
7384	1/1/2007	MR	MR3	MR1
7385	1/2/2007	MR	MR2	MR1
		.		
		.		
		.		
7533	1/30/2007	MR	MR5	MR1
7534	1/30/2007	MR	MR3	MR1
7535	1/30/2007	MR	MR1	MR1
7536	1/30/2007	MR	MR4	MR1
7537	1/30/2007	MR	MR3	MR1
7538	1/30/2007	MR	MR5	MR1
7539	1/31/2007	MR	MR1	MR1
7540	1/31/2007	MR	MR2	MR1
7541	1/31/2007	MR	MR1	MR1

(ID values: 7662, 7663, 7664, 7665, 7666, 7667, 7668, 7669, 7670)

Fig 5. Multidimensional database format (*data cube*). In this example the data is organized by modality, time and other measures (*number of exams and number of patients*). Each piece in the cube represents the total count of exams or patients (*measures*) in relation to the dimensions time and modality. A relational database is being represented on the left with tables containing the data for the months of January and February. A total of 614 MR exam counts in the January table and 504 in the February table are represented in the cube to demonstrate how the data can be "three-dimensionally" aggregated.

checked for consistency in string type. In some files, for example, the date can be saved as a text string and therefore, it has to be appropriately transformed into date/time data type. If this is not done, the date field cannot be chronologically organized by day, month, quarter, and year later on in the data aggregation process. The resulting new SQL database contains the specified fields (selected according to the relevant KPIs for the task) of each of those input files. This transformed and organized database can then be used by other applications (open source or proprietary) which can aggregate the data using OLAP in order to analyze and present generated reports. For this test environment, OLAP capabilities available in the pivot table function of Microsoft Excel were used for ad-hoc analysis and presentation of data in a dashboard-like format. The connection between MySQL database and Microsoft Excel was created using MySQL Connector/ODBC 3.51, available as an open-source application. Microsoft Excel was selected for the analysis and display of results because it is capable of handling on-demand queries; it is easily available to users in the healthcare environment and its interface is widely known. In addition, Excel also allows further calculations of the aggregated data which can then be presented in a dashboard format. By using conditional cell formatting in Excel, it is possible to change the color of the values (red, yellow or green) if they fall under certain limits. Pentaho can be an open-source substitute for this commercial solution since it also handles online analytic processes and displays results in a similar fashion.
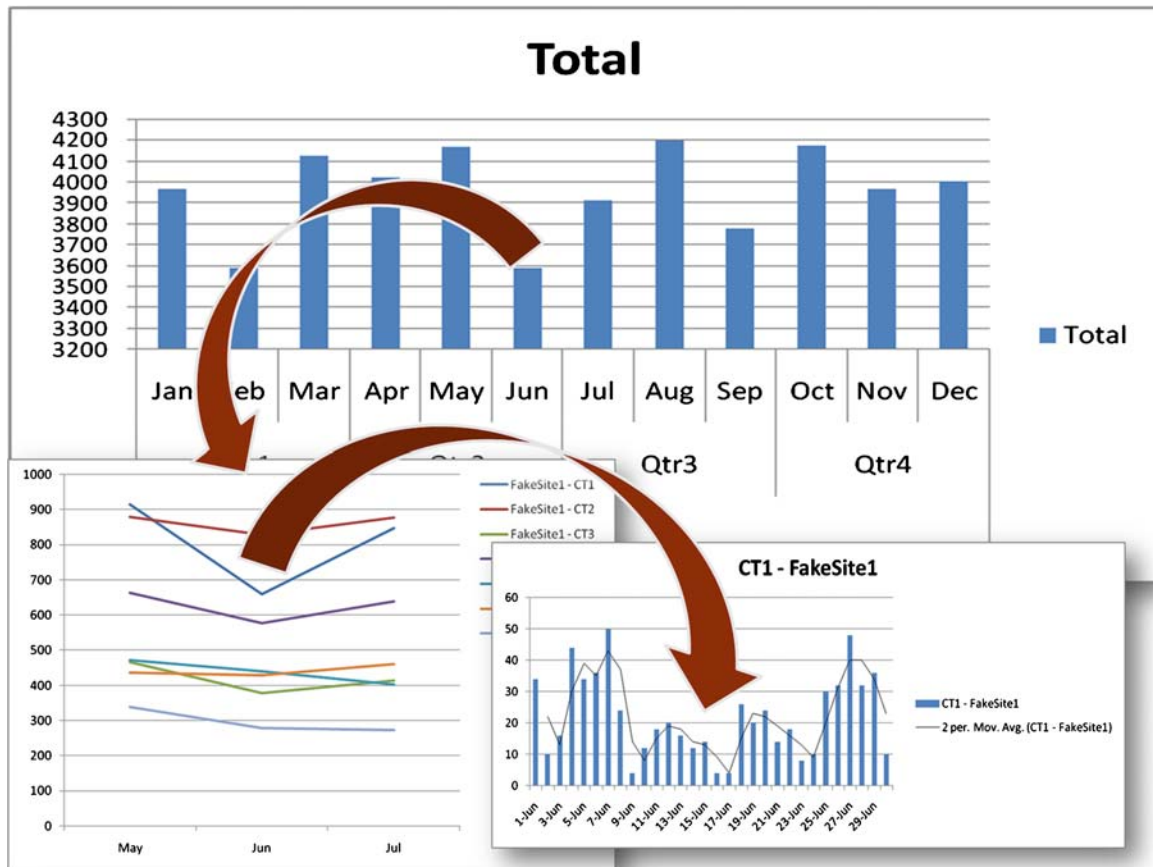
Fig 6. Graphical visualization of trends. Data can be interrogated on-the-fly. In this example, June had the lowest number of exams. It was discovered that CT1 scanner from FakeSite1 was the main variable accounting for this reduction in exams performed, with the volume on that scanner decreasing during the period between the second and third weeks of June.

RESULTS

One important observation that became clear through testing is that two distinct databases can be combined using this process. For instance, selected fields from the RIS and CPOE databases could be potentially mapped and combined into one single resource using this process. In addition, converting large relational databases into a multi-dimensional format (data cubes) can result in performance improvements when reports need to be generated.

Testing the Process

A cube was built with the dimensions time, modality, and number of exams, such that one could request the number of exams sorted by modality on a particular date (Fig. 5). The dimension time is automatically organized in a hierarchical fashion making it possible to sort the data by year, month, or day. Selecting the right variables that will represent the KPI that relates to the main goals is essential. By aggregating data, it is possible to easily analyze trends and graphically observe discrepant values. The discrepant variable can be further analyzed by drilling down into the data to that specific instance of the variable, for example a specific period of time, making the problem more easily recognizable (Fig. 6).

Results from aggregated data are important in demonstrating changes in time but additional steps may be required for final interpretation. For example, comparing the number of exams from different scanners is not as reliable as scanner capacity since hours of operations can vary between different locations (Fig. 7). Since scanner capacity data is not directly available on our

Exam_Code	(All) ▾		
Quarters	(All) ▾		
Months	Jun ▾		
Modality	CT ▾		

Row Labels ▾	Count of exams	Clinical Hours		Utilization capacity
⊟FakeSite1	1866			
CT1	660	6am - 5pm M-F	✅	92.31%
CT2	828	24/7	❌	37.91%
CT3	378	7am - 7pm M-F	❌	48.46%
⊟FakeSite2	868			
CT1	440	7am - 7pm M-F	⚠	56.41%
CT2	428	7am - 5pm M-F	⚠	65.85%
Grand Total	2734			

Fig 7. Dashboard-like spreadsheet example. Note that non-normalized data (count of exams) does not represent a good metric. Utilization capacity represents a better KPI as it allows direct comparison between scanners.

systems, additional calculations of aggregated information are required. Although it adds an extra step in the analysis process, a normalized score is more reliable for high level data visualization and comparison.

Current Use

A similar process is in place at our institution to evaluate the utilization of a teaching file application in the Radiology Department. The Medical Imaging Resource Center (MIRC) is the open-source application used for this purpose. Once installed and integrated to other systems, administrators can generate an XML file with multiple fields of the database. This database can be mapped with other resources that have information on the users and the sections to which they belong within the Radiology Department. After fusing these data the information can be mined by administrators to interactively display utilization sorted by radiologist or by section or to show which modality or body part is being used most frequently, or to bring out other trends in the data. The MIRC application does not have robust reporting capabilities, but when combined with BI tools, the database which served only for storage purposes before, can then be used for knowledge discovery. This allows administrators to understand application weaknesses and knowl-

edge gaps empowering them to better act on the solutions for these problems.

Potential Uses

It is envisioned that this methodology could be used to test feasibility and reliability of certain KPIs. In some circumstances, many aspects of the variables that will compound a KPI need to be researched in advance. For instance, if one is interested in understanding the trends in the report turn-around-time in the department, all the variables related to this KPI need to be collected, combined into a single database, checked for consistency and tested. By doing this process in a test system one can better perceive the problems that will need to be solved to implement the monitoring of such KPI in a production system.

DISCUSSION

Initially proposed for strategic management of financial organizations, the Balanced Scorecard (BSC) concept has been used by healthcare institutions. The BSC involves four strategic perspectives: financial, customer (patient), internal business process, and learning and growth measures. The use of a normalized score as was presented in this study is a well-known practice. In Radiology, research has

shown that the use of performance indicators has not been the standard practice in most academic radiology departments throughout the United States.[10] Although the reason for this finding was not within the scope of this study, it may be that technological barriers play an important role.

Generic BI concepts can be extended to Radiology in a variety of ways. However, key performance indicators cannot be selected with a business perspective only. Healthcare outcomes have different goals and costs than financial outcomes making it hard to define a straightforward strategy.[8] In addition, data structures and integration standards unique to the field must be considered. The complexity of data acquisition and the reliability of the variable to be studied should be ideally examined in a test environment. Tools such as Quality Dashboards have been used in medicine as an actionable web-based application for quality reporting and population management.[7] In radiology, the utilization of quality dashboards has been shown to be beneficial given the high complexity of data and similar overall needs.[11] As these tools become more available, radiology departments will need to prepare their infrastructure to be more easily consumed by Business Intelligence applications. Once this step is done, previously used KPIs should be more easily generated. In addition, discovery of problematic processes within the department may become more apparent with analytic tools requiring new KPIs to monitor progress of behavior change for systematic and sustainable improvements.

The methodology described in this paper can be helpful to begin investigating the use of business intelligence concepts in Radiology, but should not be seen as the optimum solution for reporting and dashboard creation. Ideally, the user interface should be web-based and the authentication system should reflect the users' roles and responsibilities in the department so data consumption can be tailored to an individual's needs. Both web-based and user-specific displays are not possible with Microsoft Excel alone. Pentaho has a web-based interface and can be a potential alternative for this purpose. Nevertheless, setting-up the web environment in Pentaho requires scripting knowledge and most likely team work for it to be appropriately installed.

The use of the method proposed here is particularly helpful when applied to cases in which data is being continuously generated and reports need to be repetitively created based on renewed data. Data processing and cleaning can be a time-consuming process, but once the commands are properly set-up in the ETL software, the program automatically does this step and data is constantly ready to be consumed by analytic tools.[12]

Multiple commercial BI solutions are available and their use may be applicable to healthcare,[13] but even with these solutions, infrastructure considerations have to be discussed before implementation. Although software applications are usually seen as an IT responsibility, setting-up such systems will likely require a deep knowledge of specific departmental processes and workflows and thus will require a multi-disciplinary and collaborative approach.[3,13]

CONCLUSION

Business Intelligence tools have been used by industry for many years. Such applications have become more available for use in healthcare environments. Open-source tools can be potentially used to create a micro-database from various departmental information sources. It is envisioned that this micro-system may be useful in the selection of new KPIs that will monitor activities and processes within the department. Business Intelligence tools can thus enable improvements in healthcare quality, safety, efficiency and financial performance once operational.

REFERENCES

1. The Joint Commission. Available at: http://www.jointcommission.org/ [Accessed September 9, 2008]

2. Pay-for-Performance | Joint Commission. Available at: http://www.jointcommission.org/PublicPolicy/pay.htm [Accessed September 9, 2008]

3. Glaser J, Stone J: Effective use of business intelligence. Healthc Financ Manage 62(2):68–72, 2008

4. Devlin BA, Murphy PT: An architecture for a business and information system. IBM Syst J 27(1):60–80, 1988

5. Luhn HP: A business intelligence system. IBM Journal 2(4):315–319, 1958

6. Kaplan RS, Norton DP: The balanced scorecard—measures that drive performance. Harv Bus Rev 70(1):71–79, 1992

7. Olsha-Yehiav M, Einbinder JS, Jung E, Linder JA, Greim J, Li Q, Schnipper JL, Middleton B: Quality Dashboards: technical and architectural considerations of an actionable

reporting tool for population management. AMIA. Annu Symp Proc 2006:1052

8. Berler A, Pavlopoulos S, Koutsouris D: Using key performance indicators as knowledge-management tools at a regional health-care authority level. IEEE Trans Inf Technol Biomed 9(2):184–192, 2005

9. SourceForge.net: Pentaho—Business Intelligence. Available at: http://sourceforge.net/projects/pentaho/ [Accessed September 9, 2008]

10. Ondategui-Parra S, Bhagwat JG, Zou KH, Gogate A, Intriere LA, Kelly P, Seltzer SE, Ros PR: Practice management performance indicators in academic radiology departments. Radiology 233(3):716–722, 2004

11. Khorasani R: Setting up a dashboard for your practice. J Am Coll Radiol 5(4):600, 2008

12. Zeng L, Xu L, Shi Z, Wang M, Wu W: Techniques, process, and enterprise solutions of business intelligence. In: Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on 6:4722–4726, 2006

13. Hedgebeth D: Data-driven decision making for the enterprise: an overview of business intelligence applications. VINE 37(4):414–420, 2007