



Prototyping an Academic Data Warehouse: Case for a Public University in Kenya

Christopher A. Moturi^{1*} and Alfred Emurugat¹

¹School of Computing and Informatics, University of Nairobi, Nairobi, Kenya.

Authors' contributions

Author CAM designed the study, managed the literature review, wrote and approved the final manuscript. Author AE designed the study and the DW, performed data analysis and wrote the first draft of the manuscript.

Article Information

DOI: 10.9734/BJAST/2015/14774

Editor(s):

- (1) Hui Li, School of Economics and Management, Zhejiang Normal University, China.
(2) Qing-Wen Wang, Department of Mathematics, Shanghai University, P.R. China.

Reviewers:

- (1) Anonymous, India.
(2) Anonymous, University of Bari, Italy.
(3) Anonymous, USA.
(4) Katarzyna Rostek, Faculty of Management, Warsaw University of Technology, Poland.
Complete Peer review History: <http://www.sciencedomain.org/review-history.php?iid=1074&id=5&aid=9147>

Original Research Article

Received 20th October 2014
Accepted 10th February 2015
Published 8th May 2015

ABSTRACT

This study aimed at determining the gaps between university top decision makers in a public university in Kenya and IT personnel in accessing, analyzing and reporting data. We sought to determine an easier and quicker way to analyze data, design, and implement a solution to turn analysis into a report with little or no help from IT department. A survey showed that the University's existing systems do not allow the top management to explore data, analyze data, and create quick reports. They depend on IT personnel to access data that they need to analyze and generate reports. As result the decision makers tend to have an uneasy relationship with IT people. The study successfully implemented a data warehouse database composed of a single multidimensional cube based on subject area of Fee Payment. By implementing a Business Intelligence (BI) program based on Excel, we turned ordinary spreadsheets into a flexible, powerful, and inexpensive BI system that gives users significant power and flexibility.

Keywords: Data warehouse; University data warehouse; University data mart; business intelligence.

*Corresponding author: E-mail: moturi@uonbi.ac.ke;

1. INTRODUCTION

Research has shown that only a small fraction of data in an organization is timely available to decision makers [1,2]. A set of significant new concepts and tools have evolved into new technology that makes it possible to address the problem by providing all the key people in the organization with access to whatever level of information needed for the organization to survive and prosper in an increasingly competitive world.

Data warehousing technology makes it easier for decision makers to analyze information in an organization without the help from Information Technology (IT) people. The main objective of this study was to determine the gaps between top decision makers and IT personnel in accessing, analyzing and reporting data. The other objectives were to determine whether there is an easier and quicker way to analyze data, design, and implement a solution to be used to turn analysis into a report with little or no help from IT department.

A data warehouse is a database containing data from multiple operational systems that has been consolidated, integrated, aggregated and structured so that it can be used to support the analysis and decision making process of an organization [3].

There has been a growing interest in Business Intelligence (BI) as a solution to reduced costs, improved service quality, and better decision-making processes [4-6]). Zeng et al. [7] define BI as the process of collection, treatment and diffusion of information that has an objective, the reduction of uncertainty in the making of all strategic decisions. BI is viewed as sets of powerful tools and approaches to improving business executive decision-making, business operations, and increasing the value of the enterprise. BI allows managers to make informed and intelligent decisions regarding the functioning of their organization. BI is the set of techniques and tools for the transformation of data into meaningful and useful information for business analysis purposes. The common functions of business intelligence technologies are reporting, online analytical processing, analytics, data mining, process mining, business performance management, text mining, and predictive analytics.

The BI technology encompasses data warehousing, OLAP, and data mining. After examining how well BI capabilities are leveraged to achieve BI success, [8] suggested that technological capabilities such as data quality, user access and the integration of BI with other systems are necessary for BI success, regardless of the decision environment. [9] describes the ability of BI tools to collect data from heterogeneous source, use advance analytical methods and support multi user demands. [10] provided an analysis of the difference between business intelligence BI and knowledge management (KM). They reported that BI focuses on explicit knowledge while KM encompasses both tacit and explicit knowledge and BI should be viewed as a subset of KM. [11] reported that IT solutions are designed to address the challenges of an organization using two different approaches: structured data management (Business Intelligence) and unstructured content management (Knowledge Management). They have proposed a design of data warehouse in attempt to apply BI technology to economic environment analysis using public data sources.

A university college in Kenya, founded in the late 1940s as a training institution for corporate students from the East Africa Community, was used in this study. Over the period, the college has undergone upgrades, from a certificate college to a constituent college of a public university. The college has four faculties namely: Engineering, Computing and Information Technology, Media and Communication, and Business. The College offers various degrees, diplomas and certificates as well as short courses.

The constituent college uses both operational systems and Enterprise Resource Planning (ERPs) to manage its day-to-day operations. These systems include financial management, human resources management, student information, library, examination management, and course management. These applications typically run in an On-Line Transaction Processing (OLTP) computing environment based on business functional areas.

However, the college lacks systems that will foster decision making. The majority of businesses, including educational institutions and government agencies, realize the need for easy and quick access to the right information [12].

Different organizations face issues and challenges when implementing Data Warehouse technologies. This includes management (management commitment and support, project management, user involvement and participation, skills) and technological (selection of DW architecture, creation of the enterprise schema, data integration and scalability, data quality, design of human-computer interfaces, data mining, security and privacy risks, and networks and telecommunication) [13]. However, Data Warehousing can yield real dividends in terms of competitive advantage and business intelligence [14,15]. The university needs a data warehouse that can unlock its data in various ERPs and other operational systems [16,17]. A university needs a different approach or technique to provide a centralized source of information accessible across different academic units. The key stakeholders should quickly analyze and develop satisfactory solutions to their underlying problems as well as making better business decisions based on historical data available in the legacy databases. Research has shown universities can improve decision making in various areas using data warehousing and data mining. This includes student academic performance [18], library management [19], decision support systems [20] and ETL (Extract, Transform and Load) process [21].

The university's existing systems do not store data in easily accessible data models that can be easily transformed into complete, significant information to sustain sound decision-making. The current user interface does not allow top management to explore, analyze data, and create reports quickly without the help from IT. Moreover, reports produced are filled with errors because the source data contained errors.

The ad hoc analysis and reporting impossibilities significantly hinder the decision making process. This study therefore sought to show how BI can be the most efficient way to analyze data quickly and turn analysis into reports with little or no help from IT department.

The four research questions were as follows. What managerial gaps are there which hinder decision making process in the constituent college? What is the most valuable technology for the college users in accessing, analyzing and reporting data easily and quickly? How will the technology be implemented to narrow the managerial gaps in the college? How can the

technology be optimized to provide timely and reliably information to the college?

2. METHODOLOGY

2.1 Research Design

A survey was conducted to ascertain the gaps between top decision makers in the college and IT department in accessing, analyzing and reporting data. A sample size of 97 employees was selected from the entire population of about 430 and issued with questionnaires with 74 questionnaires returned.

2.2 Survey Analysis

An analysis was done to determine the understanding and perception of the college users on several aspects of customer service on the following items: how long it takes to create reports, if reports are produced with the help of technician, access to all required data dynamically, dissimilar reports produced later, errors in the report due to source data, reports are not developed quickly and easily, users don't have direct access to operational data, overall, rating of the quality of data analysis and reporting.

The current situation is that the user interfaces do not allow decision makers to explore, analyze data, and create reports quickly or easily without the help from IT people. The majority (75.7%) of the decision makers depend on IT professionals in producing reports. The majority of the respondents (54.1%) were of the opinion that currently, typical reports takes at least a day to prepare. Many decision makers (48.6%) understand what the data means, but they do not have easy or dynamic access. Some of the reports (13.5%) prepared either contain errors or doctored to suit its purpose. A third (33.8%) of the respondents claims that similar reports are generated several months later against similar data. The quality of the reports produced is not up to standard. Only 21.7% of the respondents feel satisfied with the quality of data analysis and reporting. Majority (93.2%) of university decision makers are competent excel users although more than half have not used Excel as a BI. The survey therefore established there are managerial gaps in the college that hinder decision making process and that need to be addressed.

2.3 Data Warehouse Design

After literature survey, we considered data warehouse as the most valuable technology to be used [22-24]. This study considered and applied star schema dimensional modeling techniques to design an integrated data structure required in a data warehouse [25]. A data warehouse is composed of subject-oriented data mart using one or more schemas to represent specific business events or processes. A star schema contains a central fact table surrounded by many dimensions tables. The design of dimension tables for the analytical needs of the college is shown in Fig. 1, with the fee payment being fact table and the dimensions comprising students, courses, employees (staff) and time.

Since data was from disparate sources, transformation process was carried out which involved conversion, summarization and filtering of data. Data Transformation Services (DTS) tool was used to extract data from various sources including Course Management Information System (CMIS) in MS SQL server, Student Management System (SMS) in MS Access 2000, data files, external files and MS Excel spreadsheet (Fig. 2). The extracted data was transformed into a format that can be loaded into the data warehouse. The staging area is used to facilitate the transformation process. The transformed data was then loaded into dimension tables before being loaded into the fact table that references the dimension.

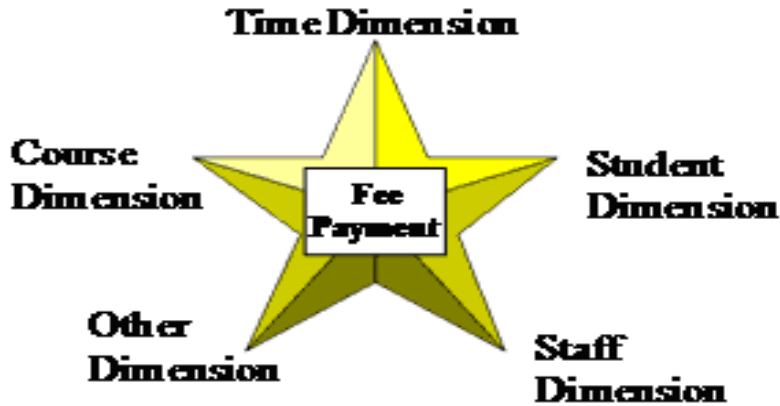


Fig. 1. Star schema design of the dimension tables

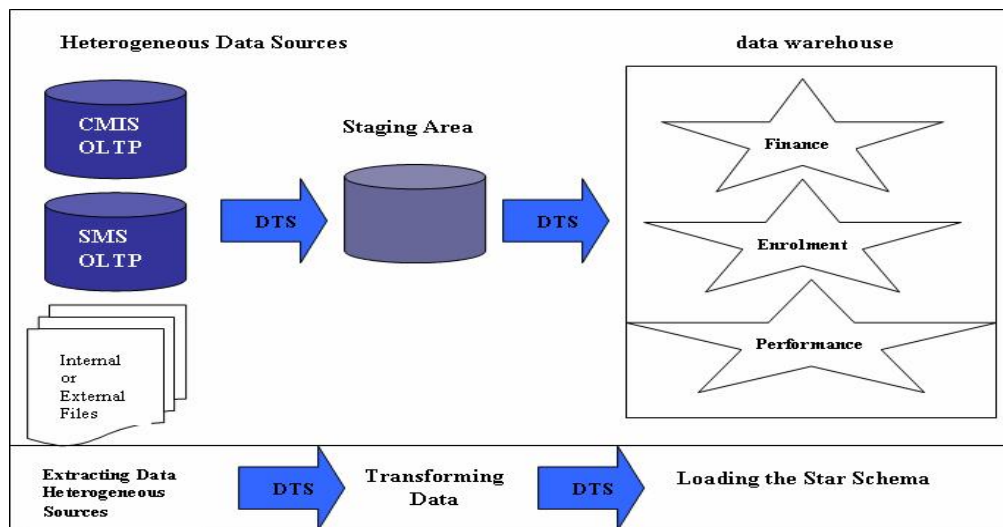


Fig. 2. University data marts

A data mart is the access layer of the data warehouse environment that is used to get data out to the users. The data mart is a subset of the data warehouse that is usually oriented to a specific business line or community of workers. A cube is an array of data understood in terms of its dimensions. After implementation, cubes are the basis for an easy, but very rich analysis and reporting experience for knowledge workers.

Online Analytical Processing (OLAP) reporting based on Excel was identified as the most suitable Business Intelligence (BI) tool. OLAP is an approach to answering multi-dimensional analytical queries and is part of the broader category of business intelligence. Typical applications of OLAP include business reporting for management reporting and business process management. OLAP reporting tool based on MS Excel gives decision makers significant power and flexibility to access, analyze and reports data.

3. RESULTS AND DISCUSSION

Data stored in a data warehouse is typically extracted from operational systems and possibly from data sources external to the institution with batch processes during off-peak operational hours [26]. Data that has been extracted is filtered to eliminate any unwanted data items and transformed to meet the data quality, security and usability requirements. It is then loaded into an appropriate data warehouse data mart to be accessed by end users who are charged with the responsibility of making informed decisions.

The proposed data warehousing technology consolidated essential historical data from the underlying college disparate databases. The source data was identified and organized into single data mart – fee payment - that tracked revenue collected from students in various courses. The fee payment cube contained only one measure called total Amount - fee collection from students taking different courses between January, 2008 and December, 2011. The dimensions contained in the cube were students, course, employees and time. The time dimension consisted of a hierarchy by year, semester and month. The dimension tables provided context of revenue collected from students in various courses. The design of the dimension tables targeted analytical needs of the business users by presenting usable, descriptive information that is easy for users to browse.

To validate the data warehouse, a Business Intelligence (BI) prototype based on Excel was implemented. OLAP Reporting Tool for Excel is an example of Excel-based BI system as shown in Fig. 3. The working environment consists of a collection of integrated windows: OLAP Report Tool for Excel main window, Command and Option window and Chart Field List, designed to support a DSS; a working Screen section - designed for dropping different fields depending report to be created; and OLAP Reporting Tool for Excel that gets its data from the data warehouse.

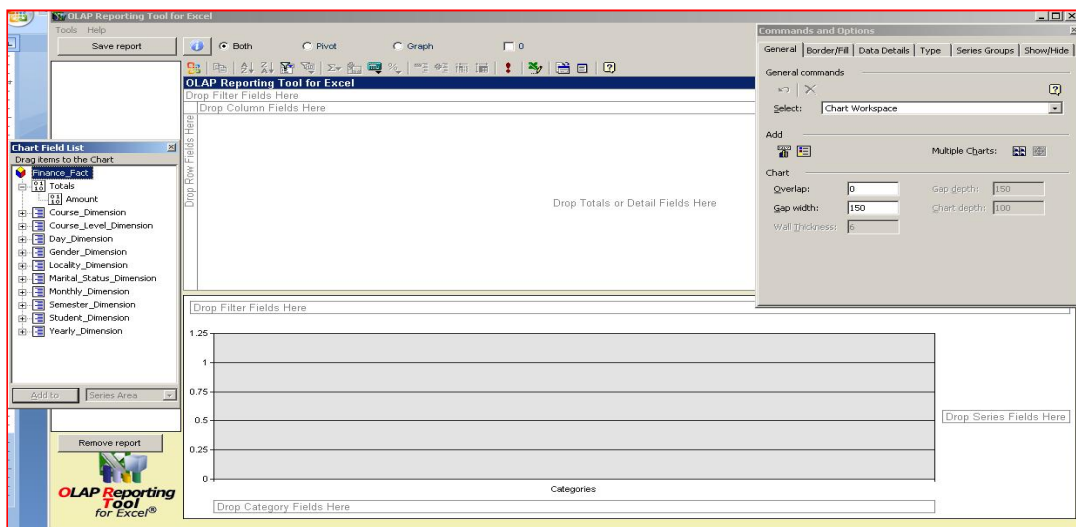


Fig. 3. OLAP reporting tool for excel

This tool can be easily deployed and give users an easy access to data from virtually any sources. An Excel spreadsheet can pull data from a cube in the BI database; send it to the add-in; receive the results; process them in the spreadsheet, and then write the results back to the BI database.

The results of the prototype implementation showed the growth in the revenue collected in the respective years. Revenue collection in 2009 and 2010 went up by 11.32% and 12.65% respectively compared to the previous respective years. However, there was sharp drop in 2011, constituting 1.78% although still more revenue was collected compared to the previous year (Table 1, Fig. 4).

The BI result shows the college top management that there has been a constant increase in revenue collection from diploma and higher diploma courses throughout the four year period. This therefore means that diploma courses are the key income generators to the university. Both certificate and degree courses showed increase

in revenue in 2009 and 2010 but a drop in 2011. The management team therefore should explore this target population of large percentage of students who do not qualify for degree programs but have the potential to continue pursuing after completing secondary education. The college courses should be restructured to consider offering courses that can generate more revenue.

Further results showed analysis of the fee collection by semester per year, fee collection per month year, fee collection per degree course name per year, most reliable and highly income generating course, and distribution of courses offered at the college.

Drill down was created to enable users retrieve from the source system. When a trend, an exception, or an error is identified, users can drill down to the transaction level to find the source data that generated this behavior. Users can navigate and analyze data from both types of databases in a single client environment.

Table 1. Fee collection per course per year

	Certificate		Degree		Diploma		Higher diploma	
	Amount	Percent	Amount	Percent	Amount	Percent	Amount	Percent
2008	748,567	2.95%	2,448,740	3.37%	21,853,923.	14.01%	2,424,550	8.72%
2009	7,593,106	29.88%	11,811,355	16.25%	33,046,219	21.18%	6,931,565	24.94%
2010	9,545,938	37.56%	33,530,481	46.13%	43,745,542	28.04%	8,212,114	29.54%
2011	7,526,186	29.61%	24,898,093	34.25%	57,391,060	36.78%	10,227,920	36.80%
Total	25,413,797	100.00%	72,688,669	100.00%	156,036,744	100.00%	27,796,149	100.00%

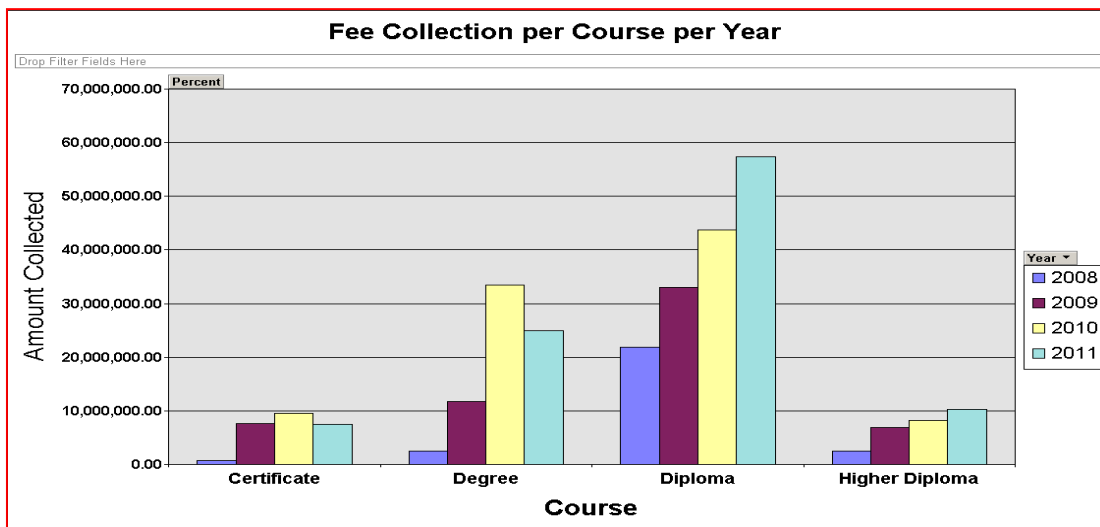


Fig. 4. Fee collection per course per year

The college should implement a data warehouse using existing resources to address managerial gaps that hinder decision making process. The study showed that existing resources or equipments, the prevailing environment, and manpower at the college are adequate to realize data warehousing system. The issue of cost should not arise but instead complexity involved in implementing a data warehouse may be looked into and be varied.

4. CONCLUSION

This research sought to establish the gaps between top management in a public university in Kenya and its IT personnel in accessing, analyzing and reporting data. We sought to determine the most valuable way to analyze data quickly and turn analysis into reports with little or no help from IT department.

The study successfully implemented a data warehouse database composed of a single multidimensional cube based on subject area: Fee Payment. This database brought together data from multiple sources providing principals and decision makers of the college greater insight into the college financial performance.

By successfully implementing a Business Intelligence (BI) prototype based on Excel, we turned ordinary spreadsheets into a flexible, powerful, and inexpensive BI system. The college management is provided with an integrated, user-friendly interface that will dynamically allow access to required data, analyze it and transform results of the analysis into a report with minimal intervention from IT department.

The study can be extended by including a larger number of data warehouse schemata such as the fact constellation schema, multi star schema, starER schema. The consistency of the data warehouse system performance can be further investigated using larger datasets for the dimension tables and with different aggregation levels.

The newest approaches or techniques in data warehousing require the use of Big Data Analytics [27]. Adoption of this approach will enable the college to analyze a mix of structured, semi-structured and unstructured data in search of valuable business information and insights.

ACKNOWLEDGEMENTS

The authors acknowledge the public university in Kenya that allowed us to use data from their student information system.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Hasan Y, Shamsuddin A, Aziati N. The impact of management information systems adoption in managerial decision making: A review. *Management*. 2013;8(4): 010-017.
2. Davenport TH. Business intelligence and organizational decisions. *International Journal of Business Intelligence Research (IJBIR)*. 2010;1(1):1-12.
3. Hobbs L, Hillson S, Lawande S, Smith P. Oracle 10 g data warehousing. Digital Press; 2011.
4. Foley É, Guillemette MG. What is business intelligence? *International Journal of Business Intelligence Research (IJBIR)*. 2010;1(4):1-28.
5. Jusoh JA, Endot N, Hamid NA, Bongsu RHR, Muda R. Conceptual framework of business intelligence analysis in academic environment using Birt. In the international conference on informatics and applications (ICIA2012). The Society of Digital Information and Wireless Communication. 2012;390-396.
6. Piedade MB, Santos MY. Business intelligence in higher education: Enhancing the teaching-learning process with a SRM system. In *Information Systems and Technologies (CISTI)*, 2010 5th Iberian Conference on (1-5). IEEE; 2010.
7. 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. 2006;6:4722-4726. IEEE.
8. Işık Ö, Jones MC, Sidorova A. Business intelligence success: The roles of BI capabilities and decision environments. *Information & Management*. 2013;50(1): 13-23.
9. Tvrdikova M. Support of decision making by business intelligence tools. In *Computer*

- Information Systems and Industrial Management Applications. 2007;364-368. CISIM'07. 6th International Conference on. IEEE.
10. Herschel RT, Jones NE. Knowledge management and business intelligence: the importance of integration. *Journal of Knowledge Management*. 2005;9(4):45-55.
 11. Albescu F, Pugna I, Paraschiv D. Business intelligence & knowledge management- Technological support for strategic management in the knowledge based economy. *Journal of information economy*. 2008;8:67-82.
 12. Akintola KG, Adetunmbi AO, Adeola OS. Building data warehousing and data mining from course management systems: A case of Federal University of Technology, Information Technology for People-Centred Development (ITePED 2011); 2011. Available: www.ncs.org.ng
 13. Al-Debei MM. Data warehouse as a backbone for business intelligence: Issues and Challenges. *European Journal of Economics, Finance and Administrative Sciences*. 2011;33:153-166.
 14. Sharma S, Jain R. Enhancing business intelligence using data warehousing: A Multi Case Analysis. *International Journal*. 2013;1(7).
 15. Alenazi SRA, Gholamdin AG, Alanzi ARA, Al-Matari EM. Prototyping data warehouse system for ministry of higher education in Saudi Arabia. *Computer and Information Science*. 2014;7(4):74-81.
 16. Heise DL. Data warehousing and decision making in higher education in the United States, PhD Dissertation, Andrews University; 2005.
 17. Wierschem D, McMillen J, McBroom R, What academia can gain from building a data warehouse, *Educause Quarterly*, 2003;1:41-46.
 18. Sreenivasarao V, Yohannes CG. Improving academic performance of students of defence university based on data warehousing and data mining. *Global Journal of Computer Science and Technology*. 2012;12(2).
 19. Laitinen M, Saarti J. A model for a library-management toolbox: Data warehousing as a tool for filtering and analyzing statistical information from multiple sources. *Library Management*. 2012; 33(4/5):253-260.
 20. Gao L, Chen Y. Application Research of University Decision Support System based on data warehouse. In *Computer Science and its Applications*. Springer Berlin Heidelberg. 2015;631-637.
 21. Mekterović I, Brkić L, Baranović M. Improving the ETL process and maintenance of higher education information system data warehouse. *WSEAS transactions on computers*. 2009;8(10):1681-1690.
 22. Di Tria F, Lefons E, Tangorra F. Research Data Mart in an Academic System. In *Engineering and Technology (S-CET), 2012 Spring Congress*. 2012;1-5. IEEE.
 23. Dell'Aquila C, Di Tria F, Lefons E, Tangorra F. Business intelligence applications for university decision makers. *WSEAS Transactions on Computers*. 2008;7(7):1010-1019.
 24. Dell'Aquila C, Di Tria F, Lefons E, Tangorra F. An Academic Data Warehouse. In *Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications*. 2007;229-235.
 25. Prasad KVKK, Data warehouse development tools. Dreamtech Press; 2005.
 26. Velicanu M, Matei G. Building a Data Warehouse Step by Step. *Informatica Economica*. 2007;11(2):83-89.
 27. Picciano AG. The evolution of big data and learning analytics in american higher education. *Journal of Asynchronous Learning Networks*. 2012;16(3):9-20.

© 2015 Moturi and Emurugat; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here:
<http://www.sciencedomain.org/review-history.php?iid=1074&id=5&aid=9147>