Improving Library Services using Business Intelligence

Danijela Tešendić*, Danijela Boberić Krstićev*
* Faculty of Sciences, University of Novi Sad, Serbia
tesendic@uns.ac.rs, dboberic@uns.ac.rs

***Abstract* - Paper describes model of data warehouse and usage of OLAP tools in order to support decision making in library. Usage of Business Intelligence technologies is presented on case study of libraries which use library management system BISIS and have considerable need for making various statistical reports. Focal point of this paper is dedicated to modeling of data warehouse using data sources from system BISIS and taking into consideration user requirements regarding reports.**

I. INTRODUCTION

Every organization has huge amount of data that are increasing every day. The performance of the organization is directly related to the ability to provide relevant information in a timely manner. Organizations must be able to transform raw data into valuable information that will enable better decision making. Every information system is built on transactional databases that allow insert, modification, deletion, and search of data relevant to the domain of the organization's business. Transactional databases have a high degree of data normalization, that is, their goal is to reduce data redundancy and achieve consistency, and therefore often have a very complex data model. The disadvantage of such databases is that they are not suitable for performing analytical queries which provide managers with necessary information to make strategic decisions. Creating such queries requires advanced IT knowledge, but at the same time it can be a very resource-intensive process that can degrade performance as well as the availability of the information system. To avoid performance degradation of information system, one of the solutions is to introduce a new way of organizing data and to develop business intelligence systems.

Business intelligence is a set of tools, methods, and applications that are used to collect, store, and analyze data from an organization. One of the basic components of a business intelligence system is data warehouse. It should be stressed that business intelligence systems and data warehouse do not exist as a ready-made solution, but need to be built in accordance with the requirements of each organization using the appropriate methodology. The data warehouse represents a centralized database storing historical data. Those data are unchangeable and obtained by collecting and processing data from various sources in order to support business decision-making [1].

This paper describes model of data warehouse and usage of OLAP tools in order to support decision making in library. Usage of Business Intelligence technologies is presented on case study of libraries which use library management system BISIS and have considerable need for making various statistical reports. The BISIS library management system is used in public libraries, faculty libraries and some specific-purpose libraries. Currently, the BISIS community comprises over 40 medium-sized libraries in Serbia [2]. The primary modules of the BISIS system comprise cataloguing, bibliography reports, circulation, OPAC, bibliographic data interchange and administration. The BISIS system has been the subject of many research papers [3][4][5][6]. In order to talk about the application of business intelligence, it is first necessary to provide an appropriate data model that can be analyzed analytically. Therefore, the central part of this paper is dedicated to the creation of a data warehouse model based on the data that exist in the BISIS system, and also taking into account the needs of the end users regarding reporting functionalities.

II. RESEARCH QUESTIONS

One of the essential functionalities of each information system, including the library information system, is to provide various statistical reports that should help the management of the library to make better business decisions. There are various reports that librarians need in everyday work, and these reports can be grouped into two groups. The first group includes statistical reports referring to the library holding (Reports showing accessions registers, Reports on the overall library holding, Reports about acquisition of books and others), while the second group includes statistical reports relating to the library members and use of library material. In this second group, two subgroups can be noticed, such as reports on members and collected fees (the number of members in the library, the number of members per category) and reports on the use of the library material (for example, the number of borrowed books in a certain period, the most popular book, ...). The reporting module in the BISIS system was implemented in the Java environment, and the JasperReports tool was used to display the report. In order to create reports on library holding, it is necessary to go through all the data in the database. This activity significantly burdens the system and reduces its performance. Because of this, reports are generated outside working hours, usually at night, and those data are stored in the XML documents. This way of creating reports significantly reduces system load and speeds up the display of reports to the end users. A more detailed description of this reporting module is provided in the paper [6].

However, this way of reporting does not meet all the needs of the library. There are reports that need to be created in real time. That group of report mainly includes reports relating to the financial aspects of the library (e.g.

the number of subscribers and the balance at the end of the day). Due to execution in real-time, these reports are ineffective and affect the performance of the entire system. Also, another disadvantage of this kind of reporting is that there is a set of predefined reports in the BISIS system, and the creation of reports that are not predefined is impossible without the prior intervention of the BISIS system developers. One of the possible solutions of the described problem is to use a data warehouse. In that case, librarians would be able to create reports by themselves using tools designed to be integrated with data warehouses. Various ways of implementing a data warehouse solution in libraries can be found in the literature [7][8].

III. METHODOLOGY

To what extent a data warehouse will be useful in making business decisions largely depends on the approach used for modeling the warehouse itself. In general, three approaches for modeling a data warehouse can be observed taking into consideration how requirements for new data warehouse are gather. These approaches can be classified in three different paradigms according to the origin of the information requirements: supply-driven, demand-driven, and hybrids of these. A supply-driven approach in modeling of data warehouse implies that the basis are data that exist in the transactional database. Those data are analyzed to determine which data are the most relevant for making business decisions and which data should be part of the data warehouse. On the other hand, a demand-driven approach implies that the data warehouse modeling is based on the end user requirements. It means that data warehouse is modeled so that can give answers to the questions asked by the end user. The third approach is a hybrid approach and it implies that in the process of building a data warehouse the two previous approaches are combined. A hybrid approach attempts to reduce the deficiencies of the previous two approaches. In the case of a supply-driven approach, it may very easily happen that the modeled data warehouse does not meet the requirements of the end users, while in the demand-driven approach, there may be no data to fill the created warehouse. Overall view of the research in the field of dimensional modeling of data warehouses is given in the paper [9].

In contrast to the previous papers dealing with the use of data warehouse in libraries and applying a demand-driven approach, this paper uses a hybrid approach to model data warehouse as it is proposed in the paper [10]. The modeling process to create a data warehouse of BISIS system using hybrid approach consisted of the following activities:

1. Analysis of existing data sources in the BISIS system and identification of possible fact and dimension tables
2. Defining user requirements regarding reporting
3. Refactoring fact and dimension tables in accordance with user requirements
4. Developing dimensional model based on star schema

IV. SOLUTION

In the BISIS system, user requirements related to analytical processing of data can be grouped into several categories. The first category contains requirements related to the library collection. Examples of the reports from this category are: the number of publications by languages for a specific period, the number of publications by departments, the number of purchased publications, the number of publications by UDC groups etc.

The second category contains requirements related to the circulation. Examples of such reports are: the number of borrowed books by the member categories, the number of borrowed books by the language of publications, the number of borrowed books by the locations of borrowing, the most borrowed books by the UDC group, member with the highest number of borrowed books for a specific period, etc.

The third category contains requirements related to the financial aspects of library membership. Some of those reports are: the number of enrolled members and the sum of member fees per day, the sum of member fees by membership categories, total amount of fees for a certain period etc.

The first step in creating a data warehouse is an analysis of existing data sources. The BISIS system uses two different data sources. Bibliographic records are stored in XML documents according to UNIMARC format. The model of that bibliographic record is described in paper [3]. Circulation data as well as holding data regarding items are stored in a relational database.

Data organized in this way are not suitable for fast and easy generation of reports that require the combination of bibliographic and circulation data. Example of such report is the number of borrowed books by the UDC groups. Information about UDC group is in the bibliographic record in the XML document while the data about book borrowing is in the relational database.

Taking into account the reporting requirements in the BISIS system as well as the existing data in the BISIS system, appropriate dimensional models are proposed in this paper. The proposed dimensional models were designed to meet all the existing requirements for analytical processing, as well as to enable a simple upgrade of the reporting system. A dimensional model based on star schema was created for each of the identified categories of reports. Only one of the three models is shown in the paper.

The dimensional model of the data warehouse for the analytical data processing of the library collection is shown in Figure 1. The data from this model are used to obtain reports about library collection. In process of generation those reports, the main role has an inventory number of copies and all reports are obtained either by counting the inventory numbers or by displaying the inventory numbers along with other data related to that copy. Therefore, the inventory number represents the measure in this dimensional model. The central table in the model is the Item table and it represents fact table. All other tables in the model are dimension tables. The Publication table represents a dimension table with the bibliographic data obtained from bibliographic records. Only the basic data from the bibliographic record, that are shown on the reports, are extracted. The Acquisition table represents a dimension table that describes the publication's acquisition data. The Location table describes the locations and departments within the library. The Status, Publisher, Language, and UDC_group tables contain information about the status of the copies, the

publisher, the language of the publication, and the UDC group to which the copy belongs. The Date and Year tables represent the time dimensions. Data in the Date table are extracted from the date of the item inventory and data in the Year table are extracted from the year of publication.

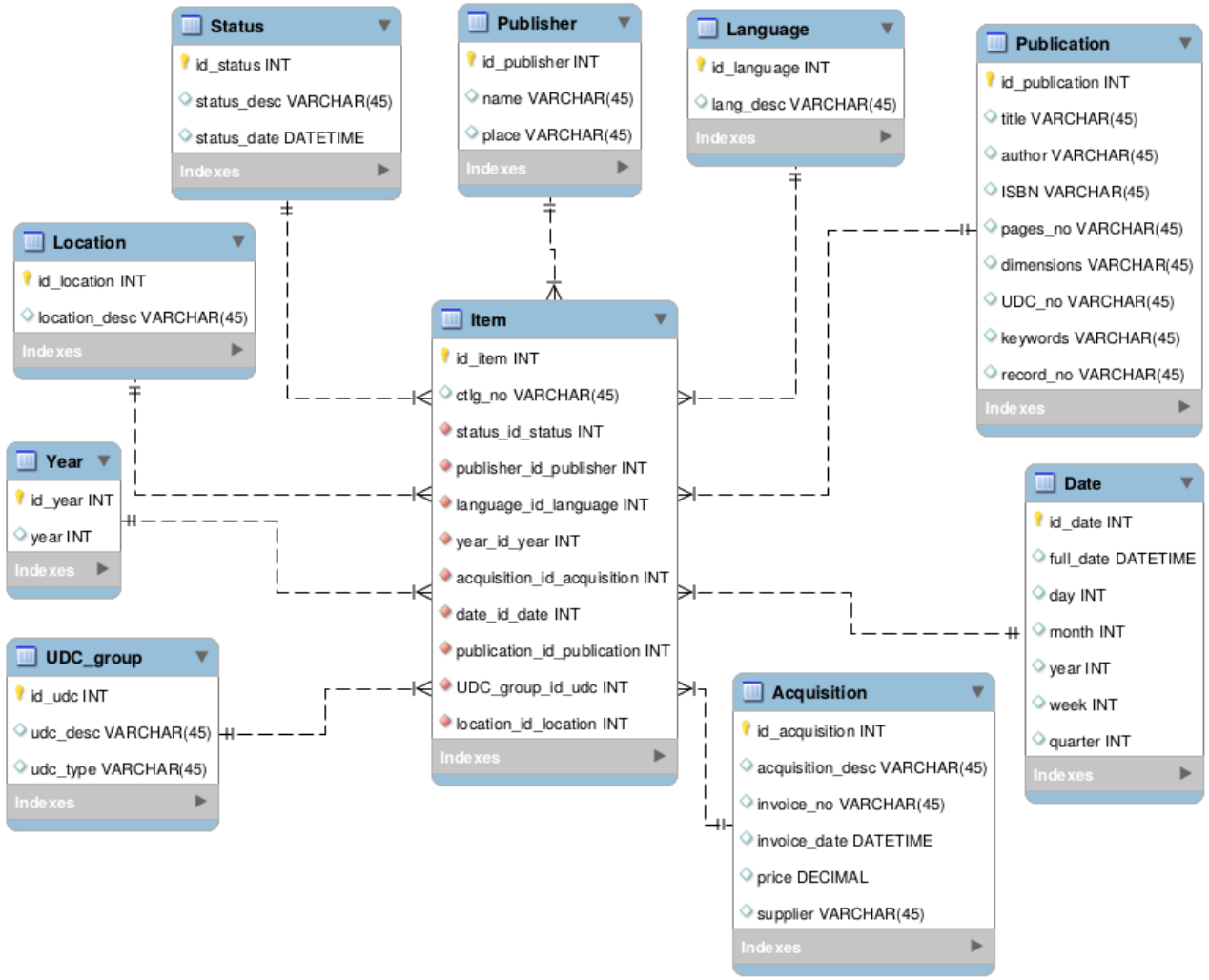

Figure 1. Data warehouse model

V. CONCLUSION

In this paper, a data warehouse model proposal is given to solve the specific problems concerning reporting in the BISIS system. The proposed model was created in accordance with the hybrid methodology of data warehouse development. It means that the design of the proposed model is based on existing data in the system as well as on user requirements related to the report process.

After creation of the data warehouse model, it is necessary to load data from different data sources into the data warehouse That task can be demanding and it should be given special attention. Also, another issue relates to data warehouse updating. In order to achieve a better performance of transactional databases, update of data warehouse is not performed in real time. In the case of libraries, those updates can be performed out of working hours so data in data warehouse can be up to date on daily basis.

However, usage of data warehouses brings great benefits to libraries. In order to achieve easier processing and access to data in the data warehouses, various OLAP tools have been developed [11]. The proposed models represent the basis for the integration of OLAP tools into the library's business. By using OLAP tools, librarians can easily generate customized reports tailored to specific needs of libraries. In this way, the librarians work in more comfortable environment, perform analytical operations independently and visualize query results without knowing the programming and the database structure.

The next step in development of BISIS system will include replacement of existing reporting module with an open-source OLAP solution which will bring great flexibility in generating statistical reports in libraries.

ACKNOWLEDGMENT

This work is partially supported by Ministry of Education and Science of Republic of Serbia, through project no. OI174023: “Intelligent techniques and their integration into wild-spectrum decision support”.

REFERENCES

[1] W.H. Inmon. *Building the Data Warehouse: Getting Started*. John Wiley & Sons. 2002.

[2] BISIS, http://www.bisis.rs/korisnici/ (last accessed 10.1.2018.).

[3] Dimić, B., Surla, D. (2009), "XML Editor for UNIMARC and MARC 21 cataloguing." *The Electronic Library*, vol. 27, no. 3, pp. 509-528.

[4] Boberić-Krstićev, D. (2013), "Information Retrieval Using a Middleware Approach", *Information Technology and Libraries*, vol. 32, no. 1, pp. 54-69.

[5] Tešendić, D., Milosavljević, B., Surla, D. (2009), "A library circulation system for city and special libraries", *The Electronic Library*, vol. 27, no. 1, pp. 162-186.

[6] Boberić D., Milosavljević B., (2009), "Generating library material reports in software system BISIS", *Proceedings of the 4th International Conference on Engineering Technologies ICET*, Novi Sad, pp. 133-137.

[7] Siguenza Guzman, L., Saquicela, V. and Cattrysse, D., (2014), "Design of an integrated decision support system for library holistic evaluation", *In Proceedings of IATUL Conferences*, pp. 1-12.

[8] Yang, Y.T. and Shieh, J.C., (2016), "Data Warehouse Applications in Libraries--The Development of Library Management Reports", In *Advanced Applied Informatics* (IIAI-AAI), 2016 5th IIAI International Congress IEEE, pp. 88-91.

[9] Romero, O., & Abelló, A., (2009), "A survey of multidimensional modeling methodologies", *International Journal of Data Warehousing and Mining*, vol. 5, no. 2, pp. 1-23.

[10] Cabibbo, L., & Torlone, R., (1998), "A logical approach to multidimensional databases", In *International Conference on Extending Database Technology*, Springer Berlin Heidelberg, pp. 183-197.

[11] OLAP Tools, https://www.softwareadvice.com/bi/olap-comparison/ (last accessed 10.1.2018.)