management

Number 54, Year XV
March 2010.

Publisher
Faculty of Organizational Sciences - Belgrade

Management is being regularly reviewed in DEST
DEST - Department of Education, Science and Training of Australia
Ulrich Periodicals

Technical Editor
Milivoje Radenković
Andrej Novović

Address
Belgrade, Jove Ilića 154
Tel./fax. 381 11 3950 868
E-mail: manage@fon.rs
http://management.fon.rs/

Journal is published quarterly

Printed by
Sigra Star, Belgrade

CIP Katalogizacija u publikaciji
Narodna biblioteka Srbije, Beograd
005
ISSN 1820-0222 = Management (Engl. ed.)
COBISS.SR-ID 112265484

CONTENTS

Business Intelligence System Development Over Document Meta Data in the Oorganization

UDC 005. 21. 352 (497. 4)

Milija Suknović, Miloš Jovanović, Boris Delibašić, Milan Vukićević
Faculty of Organizational Sciences, Belgrade

Business intelligence is an active field of research and practice. Although the field is young, it offers a lot of solutions, mainly intended for better reporting and analysis, which would support the decision making process at all levels of decision making. As the decision process is part of every area of human endeavor, so are business intelligence systems becoming part of many different areas of management. Nevertheless, there are specific business topics for which there are still no record of business intelligence application. One such topic is Document Management Systems (DMS), which are also carrying informational potential for better decisions, that could be harnessed in building a business intelligence system. This paper describes such an effort, which solves the identified needs for quality information by building a business intelligence system over document meta-data within the organization.

1. Introduction

The decision-making process is the most important process in managing an organization. The theory defines management as a succession of choices, i.e., decisions and actions meant to bring the system into a desired state. Hence the importance of successful decision making is a precondition to a successful management, which is the primary objective of organizational management. As Peter Drucker, the founder of modern management, pointed out: "Making good decisions is a crucial skill at every level."

One basic way to support the decision-making process is to supply the decision makers with information. It is actually the support to the first phase of the Symon structure of the decision-making process [1]. The last two decades have witnessed the development of a number of tools and methods in the field of business operations, under the name of Business Intelligence and for the purpose of aiding the first phase of decision making. The general principle is "the right information in the right time, to the right user", which stresses the provision and presentation, rather than the method in which the available information from the environment is made use of.

Business intelligence systems are meant to enhance the improvement of a number of criteria of the decisions made: the risk reduction and the effectiveness, at significantly lower costs (of resources and time) for consuming new information and, consequently, for decision making. This has become possible due to enormous advances in the information technologies development, in the fields of both data storage and data processing.

Business intellingence actually links two sides: the management needs for information for the purpose of making better decisions and a multiple growth of the available data.

The goal of the business intelligence is, therefore, to integrate and consolidate the data from different, often isolated information systems, as well as to allow for the decision makers to access all the data available in a simple, fast and flexible manner in order that they should acquire the information that would in turn advance the decision-making process. The business intelligence systems hence make the decision-making process faster and risk freer, which are the basic qualities of decision making.

2. Busines intelligence over documentation data in the organization

The digitalization era shows that the traditional methods of data storing are inefficient. The advantages of digitalization are numerous, and most evident are the following: a more reliable and efficient warehousing, easier search, copying, sending, etc. The disadvantages are that the digital records depend on the computer and on the electronic devices that make it possible for us to use them.

As far as the media forms are concerned, they have all found their digital match, from the text, to images, to sound, to film, etc. The paper media, called *treeware* in slang, are increasingly being abandoned in many areas of modern living, so it is with business. The digitalization of business documentation, however, is slightly slower due to legislation that demands high levels of reliability and confidentiality in order that the electronic documents should gain the legal power. Hence in the provisional solutions the electronic documentation is used alongside the paper documentation, which has the probative force in case of dispute. Such redundant storage of documents is, however, still used since it has more advantages than disadvantages, which speaks in favour of the high quality of electronic documentation. The situation in Serbia is similar, the companies use the DMS systems to organize and have the insight into the documentation, whereas they keep the paper documents in case of dispute.

The Electronic Signature Act, enforced in 2004 [2], is meant to regulate the relability and confidentiality of electronic documents. For the time being, its application is limited, and the role of certification bodies is assigned only to the "PTT Serbia" and the Economic Chamber of Serbia. When the application of this act has been extended to an appropriate extent, the paper documentation will become inferior, which should improve the electrinic business operations in this country, and this is one of the objectives of this Act.

The Document Management System serves to efficiently 1) store, 2) organize, and 3) make the digitalized documentation available to the user, so that the user can browse through it. The document is transformed into a digital image, marked and indexed by meta data, which will be dealt with in the following chapters. Additionally, the fields in the document indexed in advance are entered into the electronic database, in order to facilitate the search.

The DMS users are all members of the organization with an access to the documentation, either by updating the documents within the transactions over documents, or by having an insight into the documents in other business processes. This covers a large portion of organization, therefore the improvements achieved by the DMS are not local in character; their implementation rather means benefits for the entire company. Figure 1 shows the image of the document filed in the DMS.

2.1. Documentation metadata

The metadata are the data that describe the data, referring to the data type, data purpose, values allowed, connections with other data, data history, etc.

The document is a form in which the data are exchanged in an organization. It implicitely defines the purpose of the data, as well as the manner in which the data is used. Since the document is a unit in defining business processes, each document stores the data that describe it additionally (time of creation, creation site, changes over time, users accessing to data, etc.).

It is important to differentiate between the basic data the document contains and the data describing the document itself (metadata). The basic data contains the contents, whereas metadata describes the structure of the document and of the data the document includes.

The basic data from the document are the reason the documentation is gathered and most often they are the subject of detailed analyses. The idea in this paper is to allow for the analysis of metedata on the documents in the organization, in order that new useful information should be obtained. Thus, for example, on the basis of such descriptions the reports can be made of the surveys of one type of documents in the organization, analyses of the document creation can be conducted, as well as the analyses of implementation and changes, etc. Such

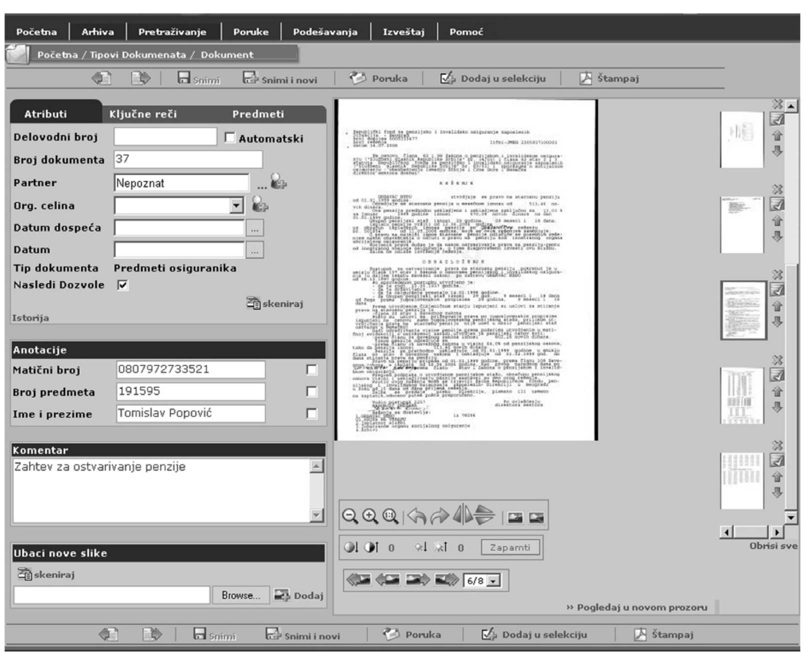

Figure 1. The image of the document filed in the DMS

information can show the state the business is in and foster the improvement of the business processes.

The DMS application also presents a generic OLTP system. The application relies on the metadata from the database that describe the fields of each defined document. Hence the base of such a system contains tables describing each of the document's attributes, whereas the concrete images of the document are stored in the tables generated by the application itself, unknown and non-existent prior to the implementation of the application in the given company. This practically means that the structure of the documents describing the business is generated by the user through the implementation of the application.

Therefore the structure of the database tables describing business operations is unknown prior to the implementation and use of the application in a concrete company, which will largely affect the creation of data warehouse, and especially the ETL procedures for warehouse loading.

The idea of developing a business intelligence system over the existing document management system came naturally. Namely, when the basic DMS application matured in use, and the databeses stored in it gathered a sufficient quantity of data, there was an opportunity to additionally use the gathered data on the documentation.

The basic DMS application offers an opportunity for browsing and insight into the documentation. It is accompanied by an application that ensures the tracking and controlling the life cycle of the document in the business processes. The users of these applications are the employees who make use of the documents in doing their business. On the other hand, the management users do not see much benefit from these applications, because they handle the documens in an infinitely granular form, on the level of the document. The management needs are viewed in dealing with the entire documentation, analysing its flows and identifying potential problems in business. Hence they are the target group in the development of new analytic functionalities through the business intelligence system.

Hence the idea for the application called *BI DMS* whose goals are the prompt, flexible, effective, all-inclusive management reporting itself, which will utilise the data from the existing DMS in use to improve the management operations.

Although the BI area is widely applied in practice, which can be seen in numerous reports and publications, the application of these systems over the data generated by the DMS was not identified in the publications available. This makes our effort all the more challenging, and its results have already earned positive criticism on being presented at a number of expert and scientific symposia [3,4].

3. Challenges solutions

The basis of the majority of BI systems is the data warehouse. It acts as a support and an adjusted environment for BI user applications for reporting and analysis. The task for the BI DMS application warehouse is to integrate data from the DMS system, transform it into suitable structure and image and make them available to user applications. In a wider context, it has to enable the integration with other information systems of the user company, so that the entiraly integrated data in the company should produce a higher synergy effect.

In building this application we created the warehouse following the suggestions and guidelines provided by the founders of this field, Ralph Kimbal and Bill Inmon, presented in the already mentioned books [5,6].

The data warehouse should include the data required for the analysis and structure them into the format suitable for this purpose. The needs for the analysis are viewed through: a) the previously required reports received as the queries of client-companies and b) potentially useful reports and analyses.

On the basis of these guidelines a data warehouse was built the description of which will be presented further on in this paper.

The problem with the initial DMS base is that, when a firm starts implementing the DMS application, the base does not contain any defined document the firm uses. Through the application, the users will define any document, with the field structure, field types and other constraints. The application then generates the tables in the base, in which the data from thus defined documents are stored.

On the other hand, the idea of the *BI DMS* application is to generate a framework for introducing business intelligence over any DMS user base, rather than developing the BI system for every user company individually. As the structure of the documents used by a given company is not known in advance, this data will be disclosed via the metadata the user utilized in defining the format of his documents.

This will be a biggest problem both in the warehouse designing and in establishing the procedures for warehouse refreshment.

3.1. Data Warehouse Schema

The warehouse schema should reflect its purpose in terms of the analyses required. In order to ensure a flexible analysis that includes the necessary and potentially useful reports, we will structure the data on the documents into the following concepts (subjects, according to Inmon):

Basic data on the document
Time of creation, change and usage of the document,
Type of the document,
Department of the company – document user
Partner with whom the document is exchanged,
Event (action) over the document.

These are the data that describe every type of document, regardless of its contents. Additionally, certain fields are stored for every document, that show the document contents and are defined by the user for the purpose of facilitating the browsing through the documents. In this way the currency amounts can be stored (in case of financial documents), as well as goods denominations (in case of sales receipts) etc. Since it is sometimes useful to include this data (contained in the given document) into the analysis, an appropriate site to do it will be a warehouse. Such data will be grouped into a separate concept:

Specific data from the document.

The architecture of the warehouse schema is presented in Figure 2.
The schema is a *star schema*, where the basic table is the *Document*, over which the measures of the dimension model are computed. The table simultaneously defines the granular structure of the warehouse as an elementary (atomic) unit of reporting. The selection of the granularity level is a critical decision of the architecture, since it restrains the possible analyses, while significantly affecting the performances.

For the purposes of the analyses in our case there are two levels of granularity that have to be covered. For the analyses of the **number of documents** (per customer, departments, partners, etc.) the atomic element is the document. For the analyses of the **usage of the documents**, the atomic element is the action (event) over a document (generation, transformation, view, etc.). The initial granulation is made possible by establishing measures over the *Document* table, whereas the second granulation is made possible by establishing measures over the *Event* table.

Another critical element of the design is defining the time dimension. In our case, time can be shown in the granulation of day, month and year, as well as in the day of the week. The time dimension is equal for both granulation levels. In the *Event* table it refers to the time a certain action over the document is conducted, while in the *Document* table the time refers to the time of the document creation (the time of creation in the organization, not the time of scanning and electronic archiving).

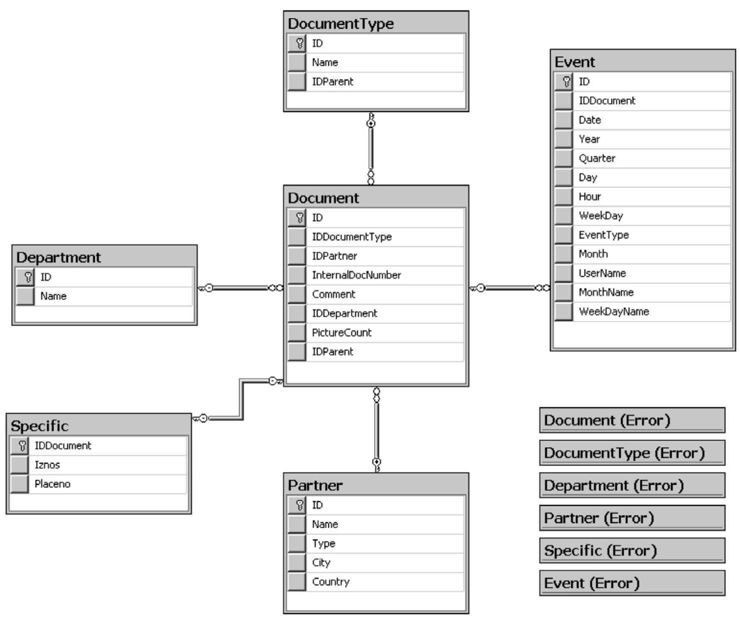

Figure 2. The schema of data warehouse BI DMS solution

The tables with the "Error" suffix are the replicas of the original tables that contain only the data that are automatically found to contain errors or inconsistencies in the process of data clearing. They are kept because certain errors are impossible to remedy automatically.

Finally, the warehouse schema is designed in the relation model, using the *MS SQL Server 2005 platform* for database management. The data model is denormalized as related to the initial model from the transactional system, which is one of the defining characteristics of the data warehouse.

3.2. Building OLAP cubes

To support the need for flexible ad-hoc reports, several OLAP cubes are defined and built. The role of the OLAP cubes is critical in obtaining reports promptly and making queries easily.

The OLAP cubes are built over the data warehouse, creating specific multidimensional structures in which the measures are defined, pre-aggregated for performance purposes. The Analysis Services tools are used from the SQL Server 2005, with the HOLAP structure as the form of storage. The example of OLAP cubes application will be presented further on in the paper.

3.3. Establishing ETL procedures

In order that the data settle in the warehouse for further analysis, it is necessary that they should be retrived from the original sources and adjusted to the data warehouse schema. The data source for the *BI DMS* warehouse is the transactional base of the *DMS* application. The base contains the preconditioned table structure for meta data in which the documents are described, as well as automatically generated tables for storing data from the concrete documents.

In the concrete execution of the ETL procedures it is necessary that the data source be defined for each data from the warehouse, as well as its required transformations. We will then proceed to systematiize the mapping of the data that require that ETL procedures be designed to load the warehouse.

As the databases of different organizations using the DMS vary by the document structure, as well as by the field names, it is necessary that specific ETL procedures should be designed for every organization that introduces the business intelligence system. Since the warehouse structure is known, it is only needed to point out to concrete fields in the database where the data required for the warehouse is stored. A special XML file is designed for the mapping which can define the data sources without having to design specific ETL procedures.

Using this XML file, the ETL procedures automatically retrieve the date fields out of the initial tables, namely those whose structure is not defined in advance. The data given in the XML file are the minimum required to carry out the ETL procedures over the concrete implementation of the DMS base.

The same XML configuration file served to define other parametres, too, such as defining the connections over bases, or determining the table filters (document types) we need in our analyses, and hence in our warehouse.

Another detail is important to mention in relation with the ETL procedures. As these procedures are long-lasting, it is important to ensure that these procedures encumber the operational base of the transaction system as little as possible. Hence the replicas of the tables from the sources are made prior to the execution of the ETL procedures, and the staging in the data transfer is designed. The replica of this base is deprived of any mechanism for referential integrity or other checks in integrity maintainance, since there are no data transformations in this interim database. This allows for the data to be retrieved from the operational base in a prompter way, and the procedures themselves continue to operate over the base replica. Additionally, these procedures are conducted at night, when the need for the operational base is minimal.

The procedures to accomplish the data mapping from the source to the warehouse are designed in the Microsoft SQL Server 2005 environment, using the *Integration Services* tools.

3.4. Quality of data

A special problem arising in the building of data warehouse is the data quality. Namely, a large number of records from the initial systems carry various errors, which, if not corrected, may affect the accuracy of the report to a large extent. As one of the founders of computer science, Charles Babbage, said: “garbage at the entrance, garbage at the exit“, which means that the quality of output reports is directly affected by the data quality in the input. The problem is that the awareness of the need to improve the data quality comes not prior to designing the data gathering system, but after the already collected data are used. The business intelligence system is therefore one of the motivating agents forf focusing upon the data quality in the initial systems.

Most typical mistakes are the the wrongly spelt words. The computer understands the words “Miloš“ and “Milos“ as two different words, although people do not find (or do not see) this to be a mistake. In data processing or in the report, such names can be interpreted as different names. Sometimes even the size of the letters can make a difference, and the computer may interpret the words "Beograd" and "beograd" as two different words. The errors in dates are also very common. They are most commonly caused by the different format of writing the date. Thus "1.20.2007." is not the same as "20.1.2007.", or "20.1.2207.". Additionally, such an error may be hidden, due to which the interpretation of the date "3.4.2007." may be ambiguous.

Errors are most commonly made when the user enters the data. The user interface that allowed the incorrect data to be entered is, however, also responsible for the arror. Although there are errors that cannot be prevented, a majority can. The concept of the Japanese firms "Poka-yoke" (Toyota) actually shows that it is possible that the systems themselves are created in such a manner that prevent any errors to be made. Well-conceptualized entry interfaces offer the user a limited input opportunity. Besides, the predefined input format sets of allowed values are designed for every input field, which largely prevents the user from making an error.

When errors are not prevented at the input, they can be identified in the system, which normally calls for a greater effort. It is of special importance that the data used in creating the report to be implemented in decision making be error-free, therefore specialsed procedures are created for clearing the identified and anticipated errors. Such procedures are most commonly activated within or after the ETL process in order that the data warehouse should be as "clean as possible". It is, of course, important to eliminate errors in the initial transactional systems as well, although this requires more effort.

These efforts reduce the errors in the business intelligence systems to a minimum. A smaller number of errors remain "hidden", i.e., correct in both the syntax and format, albeit, not in accord with reality. Such errors are identified by the report users themselves and these errors are potentially dangerous.

In conclusion, the data quality can seriously affect the quality of the business intelligence reporting, therefore it is necessary that as high as possible a quality should be achieved. This should be conducted within the business intelligence system itself ("healing treatment") as well as in the entire initial DMS system ("prevention").

4. User applications

The user applications are the ultimate visible interface of the business intelligence system towards the user. They are the reason for all the efforts made in the phases of building the PI that precede them. Directly implemented in the *BI DMS* solution are:

- predefined reports;
- dynamic reports, and
- the OLAP analysis with *drill-down* and *drill-through* options,

whereas the basis is set for further building of:

- key performance indicators (KPI)
- models for finding logics in the data (data mining),
- control boards (dashboards), etc.

The predefined and the dynamic reports offer the decisions makers a standard insight into the documentation. As the DMS is used by clients from various fields of business, the set of required reports appears to be client-specific. The situation is really more convenient as the

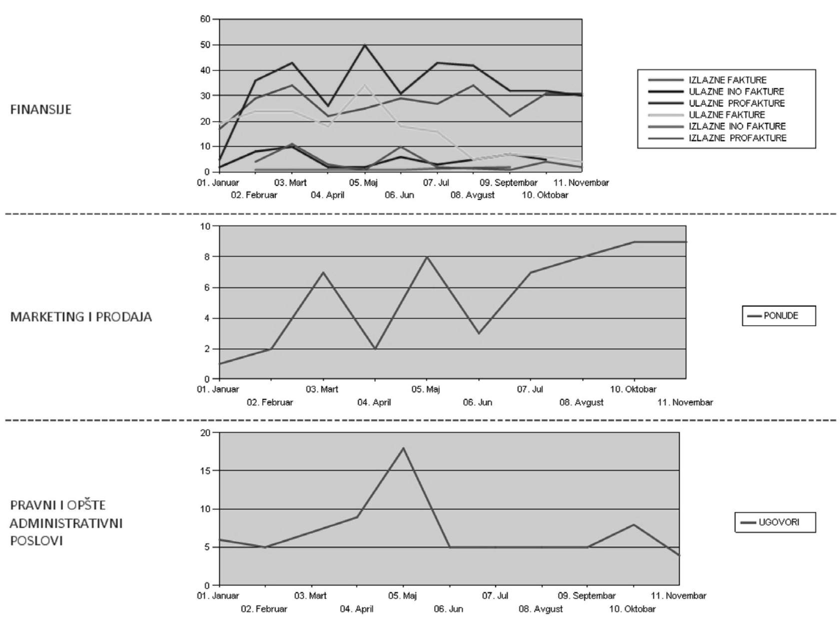

Figure 3. The report on document inflow into the company

	UGOVORI	PONUDE	TENDERSKA DOKUMENTAC
2006	49	46	5
2007	77	56	6

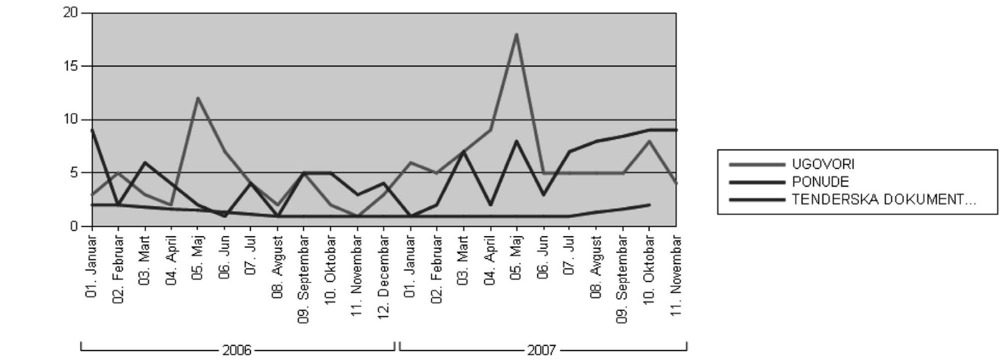

Figure 4. The report on the trends in offers, in closed contracts and in tenders

insights into the documentation and standard reports are rather similar for different users. The reports describe the conditions and the dynamics of the company's documentation, regardless of the concrete documents we have in mind. The reports built in such a way are rather useful in the future implementation of the *BI DMS* products. The examples of the predefined reports are presented in Figures 3 and 4. The reports are designed on the *SQL Server 2005 Reporting Services* platform. They show the quantity of documentation in use, as well as the trends over time.

The reports show various views on the maturity and usage of the company's documentation, classed by the document types, the time of creation/use, department, etc. The meaning of the report itself is the subject of interest of a concrete company's management, therefore they will not be described here, as they are presented as an illustration of the reporting capacity of a developed platform. It is important to stress that the reports are presented in a tabular and graphic forms simultaneously, in order that they be more comprehensive to the end user.

The OLAP analysis is another user application over the data collected in the data warehouse. The OLAP analysis helps realize the ad-hoc queries, from the overall surveys to the detailed insights into the data, in order to obtain useful information interactively. A significant advantage of the OLAP analysis is the speed at which complex reports are generated, which is one of significant problem of the classical method of reporting.

The OLAP cube, the basis of the OLAP analysis, is built by the *SQL Server 2005 Reporting Services* platform. One example of how the OLAP cube is used through the Microsoft Excel Pivot table is presented in Figure 5.

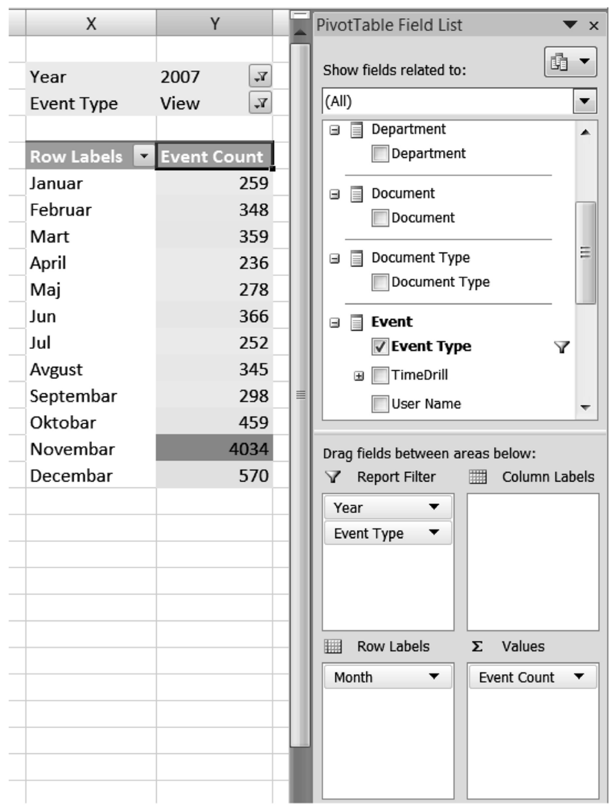

Figure 5. The example of the OLAP cube use through the Microsoft Excel Pivot table.

The OLAP analysis is used when we do not know which report will allow for us to learn something from the data. Guided by one report, the analyst creates another ad-hoc report, to get a better insight into the problem he identified in the previous report. It is in this way, from a generalized to a detailed description (*drill down*), the analyst gets to the concrete causes of problems identifiable in the data on documentation.

In addition to the *drill-down* function, the cube has the *drill-through* function that makes it possible to pass from the report on the lowest level of granulation to the transactional system (in this case, the DMS) and view the problem in the DMS base itself. For example, the reports may display a problem with a certain type of document in a certain period, and we can go directly from the report to the DMS application and see the image of the concrete document in which the problem is identified.

As mentioned above, there are other applications that can easily be built over this type of business intelligence system, such as key performance indicators and others. The system can also be extended by new data gathered from initial systems. Such superstructures are planned for further improvement of the *BI DMS* application.

5. Nonfunctional requirements in the BI DMS Solution

The basic functionality of the business intelligence system is to give the decision makers an insight into the business operations (in this case, the state of the documents), through a set of reports and analyses. In order that such a system should function, it is necessary that attention be paid to the implicit nonfunctional queries. This primarily means the **system response**, which, it is believed, should not be longer than a few second for any report, regardless of its complexity. In addition, the presence of the business intelligence system **must not endanger the operation of the transactional system and the business operations in general**. Finally, the user applications have to be **intuitive and easy to work with**, since the end users are not computer experts and may easily get demotivated to use the BI systems if these appear to be too complex.

Table 1 systematizes the efforts made to fulfill these nonfunctional requirements in the *BI DMS* solution.

6. Concluding remarks

Working on this paper we acquired a significant experience in the field of the business intelligence systems implementation over the document data in the organization. This experience further helped make new insights into the problems arising in the real implementations of the business intelligence and the solutions possible to apply in specific situations.

Finally, there remains a set of open-ended problems that can be further explored and solved, such as:

- measuring the indicators of the data quality and improvements of the automatic error identification, using the statistical tools and the tools used for processing a non-structured text;
- monitoring the BI system implementation and proposing new reports that would correspond to the identified patterns in user behaviors;
- creating a port for the access to the business intelligence functionalities;
- merging with the reporting system from other areas of business operations;
- implementation of methods and techniques for data mining, in order to identify the dependencies in the usage and creation of the company documents.

Achievements in these research paths would complete the vision of a successful implementation of the business intelligence in this area and further contribute to the theory and the practice in the business intelligence field.

Nonfunctional requests	Implementation efforts
System responds promptly to the rquest for reports	• Created indices over the critical attributes of the tables • Denormalised tables in the warehouse • Preagregation of data in the OLAP cube
Minimum encumberment of transaction systems	• “Staging” base creation prior to transformation into warehouse • Automatic starting ETL procedures at night
User-friendly applications for non-computer expert personnel	• Using MS Excel application in the OLAP analysis • Retrieving the report from the internet browser, via web application • Vizualisation of analyses in the form of graphs
Easy implementation (adjustment) for a new DMS application user	• ETL procedures are adjusted by XML file transformation • Clearing procedures are isolated as separate SQL files, easy to add new clearing procedures to

Table 1. The Nonfunctional requirements and solutions in the BI DMS solution

REFERENCE

[1] Čupić M., Suknović M. (2008), Odlučivanje, šesto izdanje, Fakultet organizacionih nauka, Beograd

[2] Zakon o elektronskom potpisu (2004), Republika Srbija, URL: http://www.parlament.sr.gov.yu/content/lat/akta/akta_detalji.asp?Id=190&t=Z

[3] Becejski-Vujaklija D., Delibašić B., Jovanović (2009), Business intelligence model for electronic documents archive, XXIII EURO konferencija, Bon, Nemačka

[4] Jovanović M., Delibašić B., Suknović M., Bečejski-Vujaklija D., Vukićević M. (2008), Primena poslovne inteligencije na oblast sistema upravljanja dokumentima SYMORG 2008

[5] Inmon William (2005), Building the Data Warehouse, 4th ed, Wiley Publishing

[6] Kimbal R., Ross M. (2002), The Data Warehouse Toolkit, John Wiley & Sons

[7] Biere Mike (2003), Business Intelligence for the Enterprise, Prentice Hall PTR

[8] Chaudhuri S., Dayal U. (1997), An overview of data warehousing and OLAP technology, ACM SIGMOD Record 26, p65-74.

[9] Larson Brian (2006), Delivering Business Intelligence with Microsoft SQL Server 2005, McGraw-Hill/Osborne

[10] Loshin David (2003), Business Intelligence: The Savvy Manager's Guide, Morgan Kaufmann

[11] Turban E., Aronson J.E., Liang TP., Sharda R. (2007), Decision support and business intelligence systems, 8th edition, Pearson Prentice Hall