5th International Conference on Control Engineering & Information Technology (CEIT-2022)
Proceedings Book Series –PBS- Vol 5, pp.1-10

Epidemiological Knowledge Mapping since the Integrating Heterogeneous Data until the Service-Oriented Data Mining Platform

Mohammed Sabri#1, Menaouer Brahami*2, Sidi Ahmed Rahal#3, Nada Matta*4

#*Department University of Sciences and Technology-Mohamed Boudiaf*

Oran, Algeria

1mohammed.sabri@univ-usto.dz

3rahalsa2001@yahoo.fr

**Department of Computer Science, National Polytechnic School of Oran*

BP. 1524 El M'naouer, 31000, Oran, Algeria

2menaouer.brahami@enp-oran.dz

**Laboratory TechCICO*

University of Technology of Troyes (UTT), France
12 Rue Marie Curie, BP 2060, 10010 Troyes, France

4nada.matta@utt.fr

***Abstract*— Knowledge management is to acquire and represent knowledge relevant to a domain, a task or a specific organization in order to facilitate access, reuse and evolution. This usually means build, maintain and evolve an explicit representation of knowledge. Knowledge mapping is graphical techniques which allows of preserving and visualizing the patrimony strategic and trades of the knowledge domains acquired over the years. Moreover, Data Warehouse is the centralized store of detailed data from all relevant source systems, allowing for ad hoc discovery and drill-down analysis by multiple user groups. It provides an infrastructure that enables businesses to extract, cleanse, and store vast amounts of data. That is, the basic purpose of a data warehouse is to empower the knowledge workers (doctors, nurses, paramedics, etc...) with information that allows them to make decisions based on a solid foundation of fact. In this paper we propose a new medical knowledge mapping approach based, on the one hand, on integration technique of heterogeneous data in the medical field by creating a data warehouse and, on the other hand, on a technique of extracting epidemiological prediction rules from medical data by choosing a technique of service-oriented data mining of APESS platform whose objective to exploit these predictive rules to automatically improve the Boolean model of the medical knowledge mapping through CARTOCEL system. The proposed approach of epidemiological Knowledge Mapping would be new direction towards as a decision support tool, whether individual or collective in the public health for management and monitoring of some pathology (chronic diseases).**

***Keywords*— Knowledge Management, Data Warehouse, Data mining, Services Web, APESS Platform, Decision Support System.**

I. Introduction

The approach of knowledge management, business intelligence and service-oriented data mining was used as theoretical technologies in order to build an intelligence enterprise framework. Knowledge management (KM) implementations in the healthcare sector depend on the methods and techniques to manage acquiring, structuring, transfer, and utilizing tacit and explicit knowledge. The continuous improvement to the epidemiological services and to the chronic diseases services requires adequate implementations to overcome the obstacles occurring in this service. Based on the interactions between those various public health services, the amount of knowledge is massive while all of knowledge details created by any of those services are necessary for a successful healthcare service delivery. Based on the interactions between those various public health services, the amount of knowledge is massive while all knowledge details created by any of those services are necessary for a successful healthcare service delivery. Since the business intelligence process can create additional customer value (Health Departments) through knowledge creation with integrating heterogeneous data, business intelligence can provide users (doctors) with reliable, accurate information and help them make decisions. In this sense, effective Business Intelligence system allows data collection from all departments within the company, their analysis, preparation the necessary reports through data mining and addressing those users who are most needed.

The doctor and these collaborators need to know its epidemiological data, which implies a strong collaboration between health trades actors and interoperability between the systems being used in the epidemiological services. Given the complexity of the epidemiological and chronic diseases field, we encounter several problems such as:

- The diversity of the distributed data sources and their heterogeneous in the different services.

ISSN: 2961-6611

- The difficulties of accessing to relevant information for the disease monitoring are related to the dispersion of epidemiological information on different hospital information systems (HIS) which are often autonomous and heterogeneous.
- The difficulty to understand the mechanisms of acquisition and interpretation of the medical experiences and epidemiologic diagnostic reasoning.
- The modelling of the information does not allow facilitate access to distributed mining in order to Increase data reliability, dissemination and use of medical knowledge.

Subsequently, we present case studies which illustrate ―good practices‖ in either managing knowledge for, from and about a large healthcare organization. Likewise, the subsequent section presents the case of Business Intelligence development at large companies, followed by discussion and conclusions for understanding Business Intelligence (BI) systems evolution in practice and research. Thereafter, we utilized a service-oriented data mining platform (APESS) for business intelligence whose objective is to integrate of many services which coordinate and communicate to one another for their respective goals, thus enabling simplified data delivery and low-latency analytics. In addition, this APESS platform is composed of two main components, on the one hand, web services, which implement data mining algorithms and on the other hand a user web interface, which can be used for modelling applications that use the services of the platform..

II. THEORETICAL FRAMING

In order to understand the existing state of knowledge management and the knowledge extracted from the heterogeneous data using the different techniques of Business Intelligence (BI) and the service-oriented data mining platforms in healthcare organizations, it is essential to review similar efforts made by other researchers.

A. Knowledge Management: Concepts and Case Studies

Since KM is a relatively new discipline, and object of study of many researches, this research focuses the attention in the different approaches and models given to the topic. To be considered appropriate and relevant by the author of this research, initially will be taken the contribution of [7], who introduce the topic of the importance of knowledge management for organizational performance has been widely recognized and acknowledged in management literature. In this context, knowledge management is assumed to create value for organizations by applying their accumulated knowledge to their products and services outputs. Knowledge management is the formal decision-making process; the decision is made to grasp and make use of the new knowledge, and then to ensure that useful value is created for staff [10]. For Jennex [16] KM is the practice of selectively applying knowledge from previous experiences of decision making to current and future decision making activities with the express purpose of improving the organization's effectiveness. He also considers a Knowledge Management System (KMS) as that system which is created to facilitate the capture, storage, retrieval and reuse of knowledge. KM and KMS holistically combine organizational and technical solutions to achieve the goals of knowledge retention and reuse to ultimately improve organizational and individual decision making [16]. Plessis [18] considers knowledge management as a structured and planned approach to manage the creation, sharing, harvesting and how to use leverage as a result of organizational knowledge, capacity, speed and efficiency in the provision and delivery of goods and services to customers. According to Zack [24], knowledge management involves the process through which knowledge is acquired and edited, shared among the members of the organization and implemented to innovate and improve the functioning of the organization. For Montoro [19], it is the "discipline that is required to examine the design and implementation of systems whose main goal is that all the tacit, explicit, individual, internal and external knowledge involved in the organization can transform and become organizational knowledge.

However, practicing successful KM strategies within the organization can lead to achieve the following advantages [28]: (1) enhance customer services; (2) apply various quality methods to gain faster innovation processes; (3) improve communication and collaboration between all stakeholders; and finally (5) achieve financial goals. In this research, the conceptual definition of KM practices contains three dimensions; that is ―knowledge creation, knowledge storage and knowledge sharing. Therefore, Wulantika [9] argues that (1) knowledge Management (KM) is the organization's activities in managing knowledge so in the end become an asset, (2) a lot of knowledge of the strategies used by people in a short time so that they can interact with each other, and (3) apply knowledge in a variety of everyday work in order to improve organizational performance. Various academics and business practitioners began to apply to grow and develop knowledge management through research and application in business practices. For Yaghoubi [25], KM promotes an integrated approach for identifying, capturing, retrieving, sharing and evaluating all enterprises' information assets. These information assets may include databases, documents, policies, procedures, as well as the uncaptured tacit expertise and experience stored in individuals' heads. Another study by Perez-Soltero [1], analyzed the knowledge management in small and medium enterprises in the restaurant industry in northern Mexico, in order to identify areas for improvement in their production processes. They analyzed the stages of identification, storage, creation, distribution, use and measurement of knowledge, and concluded that the use or application of knowledge is found to be more developed because past experiences are used to make better decisions and improve tasks, processes and services.

To sum up, a number of studies by Abas & Jali [38] highlighted an organizing review of Knowledge Management and conceptual review and empirical evidence of the three main themes of Knowledge Management, namely knowledge creation, knowledge transfer and knowledge application. Based on the empirical analysis, this paper concludes that

ISSN: 2961-6611

huge number of prior studies only focuses on the knowledge transfer theme. This is because knowledge transfer is said to be more visible and easier to observe as compared to creation and application. Specifically, this paper highlights that the emerging concept of innovation process is about the capability of managing knowledge; as well as also displays past contribution of Knowledge Management towards new technology development. Al-qarioti [20] examined the impact of specific KM dimensions on organizational performance of (245) small size business owners and managers at a management-level in their firms from (86) enterprises in Isfahan. The study results showed that some knowledge resources are directly related to organizational performance, while others are not.

Finally, Christozov & Toleva-Stoimenova [3] delineate four forms of knowledge, which we will use in our later discussion:

- Explicit-individual (concepts): Examples include engineering formula calculation and basic spreadsheet manipulation.
-Tacit-individual (skills): Examples include managing teams and troubleshooting unusual exceptions.
-Explicit-group (stores): Examples include formalized processes and patents.
- Tacit-group (genres): Examples include corporate culture and norms of communication.

The process model associated with knowledge management consists of well-defined activities which: 1) help to ensure the quality of the data and information used by knowledge workers (Trades actor's and Managers), 2) assist in the refinement of data and information into knowledge (knowledge map and knowledge engineering methods), 3) allow the efficient storage and retrieval of metadata and knowledge (knowledge map and knowledge engineering methods), 4) promote the timely dissemination and distribution of knowledge, and 5) support the tailored sharing of knowledge (e-learning, community of practice, and knowledge servers).

B. Business Intelligence: Concepts

The value of information increases with the number of users who can access that information, multiplied by the number of business areas in which the user works companies desperately need timely and relevant information and knowledge [17].

Since the literature and the operating people have different definitions of Business Intelligence, we are going to show the different experts definition of the concept. All of the experts agreed with each other that the definition of BI is a tool to make the right decisions in an organization. This by, collect data from different systems to create strategic, operational and tactical decisions [22]. Furthermore, one of the experts mean that definition of BI can be viewed in two different ways: (1) A system oriented way which collect, store and present data in different tools, and (2) a process oriented way, which he describes that the users need some type of knowledge as well. Processes are collected and the data are presented and the next thing is to do the data analysable for the end users.

In other words, the aim of BI is to provide the decision makers with analytical tools and information to make good decisions, so they can improve the correctness and quality of inputs in to the decision making process [22]. For Negash & Gray [29], an ideal BI system should provide the decision maker with data delivered in the right time, at the right location and in the right form. This would improve the timelines and value of the decision process. Curko [8] in the chapter has explored the problems of capturing different types of structured and unstructured data relate to, filtering, grouping, cleansing, and enhancement. Business Intelligence (BI) is used to represent the tools and systems that play a vital role in knowledge sharing and dissemination at organizations [34]. According to Nemati [5], Business intelligence tools are software tools which allow the retrieval, analysis and reporting of data.

Based on earlier literature, BI is a set of techniques of gathering, accessing and analysing a big amount of data, while KM is a set of practices for the creation, development and application of knowledge to create a better performance in the organization [22]. For Davenport [31], there are several differences in the concepts. KM is mainly about human subjective knowledge, thus no data or objective information. According to Khan and Quadri [27], companies need both KM and BI as an integrated system to get value from explicit and tacit knowledge. For his part, Panian [40] has discussed the characteristics and benefits of Service-oriented Architecture (SOA) and suggested to use BI solutions as Web services in an SOA environment. Furthermore, Nguyen [32] have presented a real-time Business Intelligence architecture called SARESA with the aim of providing continuous, real-time analytics in order to enable proactive responses to a business environment for effectively managing and controlling time sensitive business processes. The authors have introduced Sense & Respond loops and a service-oriented architecture that is able to detect situations and exceptions, perform complex analytical tasks and reflect on the gap between current situations and desired management goals.

In conclusion, an understanding BI system enables any organization to implement an analytical approach that transforms data into information, information into knowledge and then knowledge into decisions. Hence, KM is about Knowledge sharing, extraction, communication, application and innovation by captures, stores, organizes, and distributes knowledge, while BI converts data into knowledge for the need of the end user by identifying trends for new business strategies.

C. Service-Oriented Data Mining: Concepts and Platforms

Data mining is a complex process, which can be deployed by means of multiple approaches. Otherwise, health, Science, and pharmacy fields often need to analyze very large datasets maintained over geographically distributed sites by using the computational power of distributed and parallel systems. The grid can play a significant role in providing an effective computational support for distributed knowledge discovery applications. For this, Georgescu [33], was discusses how to design and implement data mining applications by using the

ISSN: 2961-6611

Knowledge Grid tools starting from searching grid resources, composing software and data components, and executing the resulting data mining process on a grid. Some performance results are also discussed.

Nevertheless, several authors say that the distributed nature of data and the extension of information sharing make the SOA a suitable scenario in which data mining applications can be executed. For this and after a literature search, we can observe that there is an attempt to design architecture for performing data mining on the Grid, especially in the research work of [14]. The authors have presented the design of a Knowledge Grid architecture based on the non-OGSA-based version of the Globus Toolkit. This architecture extends the basic grid services with services of knowledge discovery on geographically distributed infrastructures. Furthermore, Brezany [26] have described the service oriented architecture and its components implemented in the GridMiner application. Several data mining and OLAP services have been already deployed and are ready to perform the knowledge discovery tasks. Moreover, it is essential that the system provides a powerful, flexible and simple to use graphical user interface (GUI) which hides the complexity of the Grid but still offering possibilities to interfere during the execution phase, control the task execution and visualize results. AlSairafi [30] have developed the Discovery Net architecture for building grid-based knowledge discovery applications. This architecture enables the creation of highlevel, re-usable and distributed application workflows that use a variety of distributed resources. It is built on top of standard protocols and standard infrastructures such as Globus but also defines its own protocols such as the Discovery Process Markup Language for data flow management. Perez and Pëna [23] have proposed both a novel architecture for Data Mining Grid, named DMGA, and the implementation of this architecture, named WekaG. The DMGA (Data Mining Grid Architecture) is a vertical and generic architecture which is based on the main data mining stages: pre-processing, data mining and post-processing and usage patterns for their composition in a real scenario. However, the implementation of WekaG architecture is based on Weka, a well-known tool form developing machine learning algorithms, which can be used for solving data mining problems. Finally, Beynona et al., [15] have described a middleware framework, called DataCutter that is designed to provide support for subletting and processing of datasets in a distributed and heterogeneous environment. In DataCutter, data intensive applications are represented as a set of filters. A filter is a user-defined object with methods to carry out application-specific processing on data. Both the filtering and indexing services use the data access service to read data and index information from files stored on archival storage systems. The indexing service manages the indices and indexing methods registered with DataCutter. The filtering service manages the filters for application-specific aggregation operations.

Similarly, Talia et al., [4] have presented Weka4WS, a framework that extends the Weka toolkit for supporting distributed data mining on Grid environments. Weka4WS adopts the emerging Web Services Resource Framework (WSRF) for accessing remote data mining algorithms and managing distributed computations. The Weka4WS user interface is a modified Weka Explorer environment that supports the execution of both local and remote data mining tasks. On every computing node, a WSRF-compliant Web Service is used to expose all the data mining algorithms provided by the Weka library. Wu et al., [37] have developed a support platform, called CGSP, is a grid middleware for the construction of the ChinaGrid. ChinaGrid aims at building a public service system for Chinese education and research. Function modules of CGSP for system running are Domain Manager, Information Center, Job Manager, Data Manager, Service Container and Security Manager. Furthermore, (Stankovski et al., 2008) have designed the DataMiningGrid system according to three principles: service-oriented architecture (SOA), standardization, and open technology. SOA promotes the sharing of geographically dispersed business functions in an "exible way". Its main features include high performance, scalability, "exibility, ease of use, conceptual simplicity, compliance with emerging grid and data mining standards, and the use of mainstream grid and open technology. Finally, Podpecan et al., [35] have proposed a novel Service-oriented Knowledge Discovery framework and its implementation in a service-oriented data mining environment Orange4WS (Orange for Web Services), based on the existing Orange data mining toolbox and its visual programming environment, which enables manual composition of data mining workflows. The new service-oriented data mining environment Orange4WS includes the following new features: simple use of web services as remote components that can be included into a data mining workflow; simple incorporation of relational data mining algorithms; a knowledge discovery ontology to describe workflow components (data, knowledge and data mining services) in an abstract and machine-interpretable way, and its use by a planner that enables automated composition of data mining workflows.

III. PROPOSED APPROACH

The objective of our approach, illustrated in Figure 1, is to propose a new medical knowledge mapping approach based, on the one hand, on integration technique of heterogeneous data in the medical field by creating a data warehouse and, on the other hand, on a technique of extracting epidemiological prediction rules from medical data by choosing a technique of service-oriented data mining of APESS platform [21] whose objective to exploit these predictive rules to automatically improve the Boolean model of the medical knowledge mapping through CARTOCEL system [11, 12]. The proposed approach of epidemiological Knowledge Mapping would be new direction towards as a decision support tool, whether individual or collective in the public health for management and monitoring of some pathology (chronic diseases).

ISSN: 2961-6611

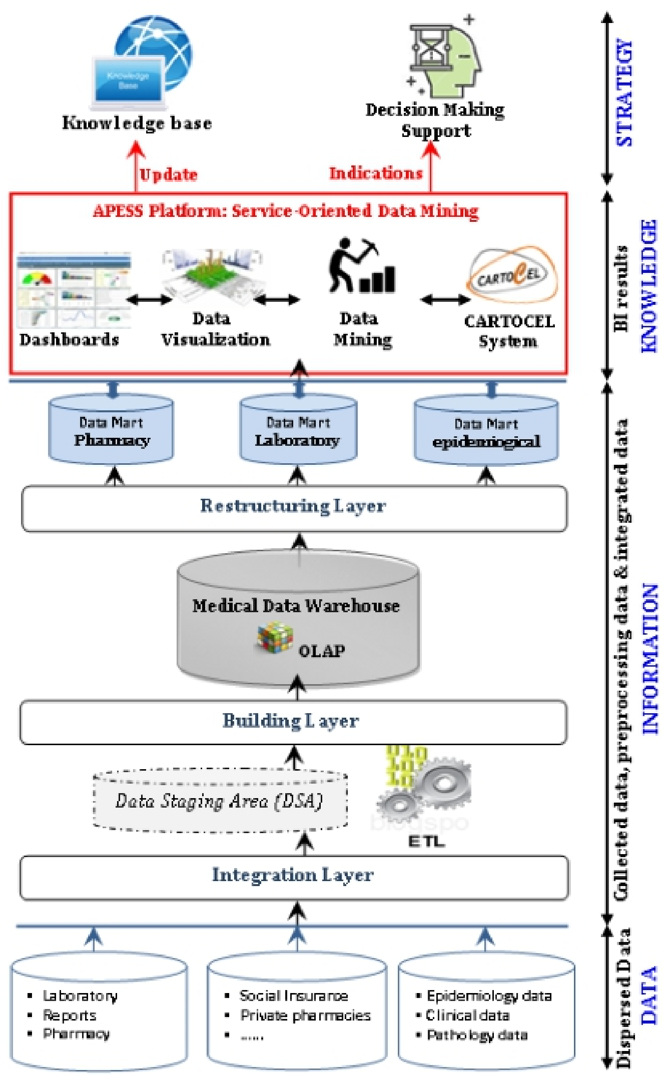

Fig. 1 Architecture of the approach proposed.

A. Data Warehousing

The first phase of our approach is the design of the data warehousing, with the objective of obtaining a unique source of data to carry out the data mining services web tasks.

Our data sources are the records of sales from the public and private pharmacies. To be usable, all data from distributed systems must be organized, coordinated, integrated, and finally stored to give the user an overview of information. The architecture of the steps processing of pharmaceutical data warehouse (see Figure 2), is articulated around three axes:

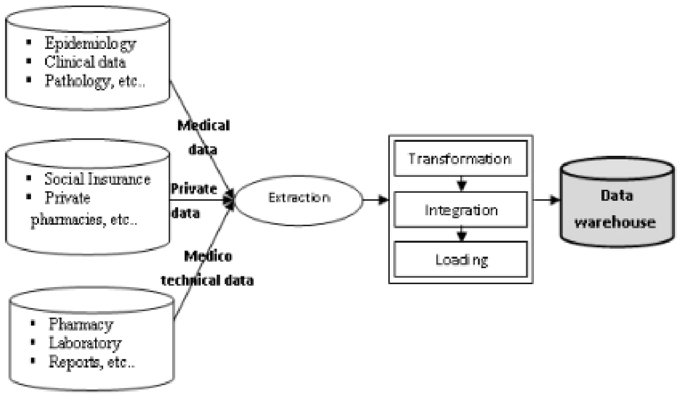

Fig. 2 Steps of processing building of medical data warehouse.

Integration: This first step consists of extracting and gathering the data coming from the various databases of public and private pharmacies and the external sources. These databases are supported by the same relational DBMS, they are identical from the point of view of their structures, and they are installed in different sites where no connection exists between these sites, neither the existence of a centralized system. The source recovered databases (files) are coded and stored in the file system.

Building: It consists of extracting the relevant data and copying it in the warehouse. Consequently, our data warehouse will constitute a centralized collection of materialized and historical data, available for data mining services web. The data related to the drug sales and the characteristics related to the sold products, are taken into account within the scope of this study, and the other data, such as the purchases, are neglected. During the transformation of data, we encountered several types of conflicts that are each treated separately: conflict of classification, descriptive conflict, structural conflict, and data conflicts.

Restructuring: This step consists of reorganizing the data, in data marts, to support the data mining services web; a specific data mart is created, in relation to the information concerning the chronic diseases selected out of all the retail sales and the characteristics of the patients belonging to the essential variables for the data mining services web.

Construction of the warehouse schema: according to Zerf Boudjettou [39], the result of processing of the data warehouse building is to define a global schema providing an integrated view of the sources the will be exploited later in the process of extracting knowledge from data. For this, our data warehouse is based on the star model (see Figure 3) and contains all the information about retail sales, products and places where the pharmacies are located. The source data available is commercial data (basic sales records) to carry out medical research (epidemiologic), and we chose a traditional multidimensional modeling for our data warehouse.

ISSN: 2961-6611

5th International Conference on Control Engineering & Information Technology (CEIT-2022)
Proceedings Book Series –PBS- Vol 5, pp.1-10

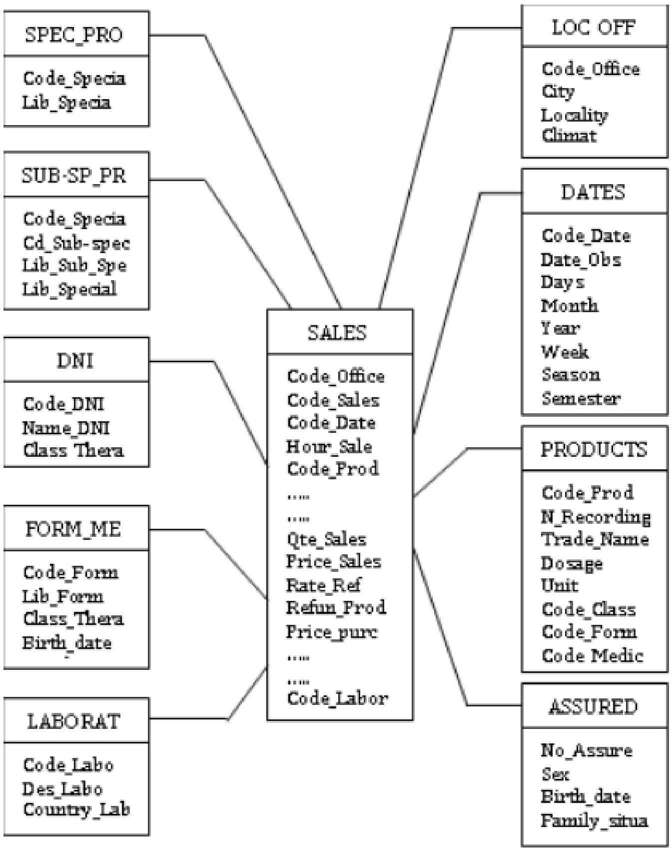

Fig. 3 Global schema of our data warehouse (dimensions and fact tables)

B. Data Mart Epidemiological

The data mart "**MACHR**" is focused and driven by the needs of our system. It has the same purpose as the medical data warehouse (provide architectural decisions), but it aims to solve our problems with a smaller number of users [39].

The data mart "**MACHR**" store is specialized for the epidemiological study of chronic disease: asthma, high blood pressure, and diabetes, where extractions are performed on the data warehouse, it is not taking that sale for these pathologies. From which extractions are made on our data warehouse it is taking sales for these pathologies. Recall that the data mart "MACHR" is part materialized on the data warehouse. MACHR is modeled as a star schema and implemented in the ORACLE database, and for launching data mining tasks on multiple views and depending on the selected dimensions. The MACHR data is as follows (see Figure 4):

- The fact table ―SALES‖; contains sold quantity (gross), selling price, etc.

- The dimension tables:

 – Localization of selected pharmacies ―LOCALIZATIONS_OFFICINES‖.
 – Dimension date ―DATES‖.
 – Table of the handled products ―PRODUCTS‖ includes commercial name, etc.
 – Specialties of the various existing products in database ―SPECIALITIES_PDT‖.
 – ―DNI (Denominations Nonproprietary International)‖ and corresponding disease.
 – Laboratories manufacturers products

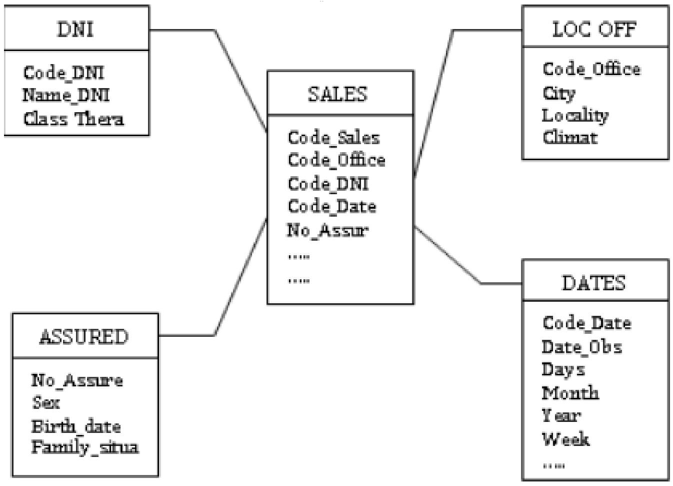

Fig. 4 The star model of MACHR (Data Mart epidemiological)

C. Extracting Knowledge from Data by APESS Platform

Before launching the process of extracting knowledge from data, we thought it useful to give a brief presentation on APESS (Assistance Platform for Epidemiological Searches and Surveillance) platform.

Let's remember that, The APESS platform's concept has been proposed on service-oriented architecture based on data mining and applied to Epidemiological Searches and Surveillance [21]. The Data mining tasks of the APESS Platform include several basic services. Each of these services is adapted to a specific usage context. The process of services selection and composition is based a priori on predefined rules. According to Sabri & Rahal [21], the APESS platform is a runtime environment that is designed and implemented according to a multilayer structure: Data Access Services Layer, Data Mining Services Layer and User Services Layer. (see figure 5).

In this part of the system, we'll create a process for extracting epidemiological knowledge from the data warehouse (data mart "MACHR") constructed in the previous section. The process we have used is divided into several phases of APESS platform [21]: data selection, data preparation, use of an intelligent data mining method applied to the processed data and finally the evaluation and validation of models.

ISSN: 2961-6611

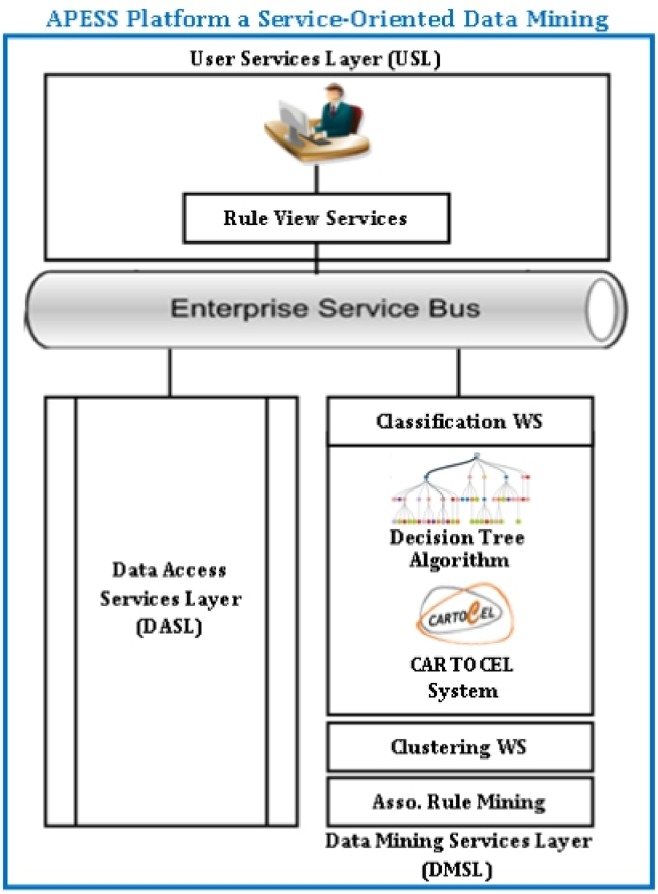

Fig. 5 Architecture of APESS platform for service-oriented data mining [21]

D. Pre-processing

Data from the warehouse is very varied and is not necessarily all exploitable by the data mining techniques. Most of the used techniques process only data tables in the traditional lines/columns. The objective is to prepare lines/columns tables; in other words, tables of individuals/variables, obtained by the following stages (see Table1).

TABLE I
EXAMPLE OF A LEARNING SAMPLE - DISEASES

Font Size	Appearance (in Time New Roman or Times)		
	Regular	**Bold**	**Italic**
8	table caption (in Small Caps), figure caption, reference item		reference item (partial)
9	author email address (in Courier), cell in a table	abstract body	abstract heading (also in Bold)
10	level-1 heading (in Small Caps), paragraph		level-2 heading, level-3 heading, author affiliation
11	author name		
24	title		

Data selection: It is carried out on the data which already exist in the data warehouse and which are in tabular form. It is then a question of applying filters which will enable us to select a subset of lines or columns. Data selection is based on the following information:

- From the fact table ―SALES‖, we will take the sold quantity, taken first in its ―gross‖ state and aggregated according to the selected dimensions.
- From table ―LOCALIZATIONS_OFFICINES‖, the attribute ―LOCALITY‖.
- The date dimension ―DATES‖ in order to carry out the data mining on a time interval. In our case, we proceed by the period ―MONTH‖.
- From the table ―DNI‖, we take information about present diseases. A filter is then applied to keep the records related to the selected diseases only.
- Finally, the patients, present in the table ―ASSURED‖ and from which we take the gender and age attributes (recommendations of the experts).

Cleaning and enrichment of the data: A stage of cleaning of the data is essential in order to process the missing data (suppression of records). Besides, enrichment by external sources was carried out during the creation of the data warehouse.

Transformation and reduction of dimension: This is about transforming an attribute (**A)** into another (**A'**) which would be more relevant to match the objectives of the study. For example, patients' dates of birth have been transformed to obtain age within intervals.

E. Data Mining by Decision Tree

After data storage and pre-processing, the phase of data mining may start. To illustrate this step, consider the problem of Epidemiological study of chronic diseases.

In this case, the patient population affected by the problem of learning is a set of tuples consisting of the four predictor variables X1, X2, X3, X4 (Locality, Season, Age, and Gender) and their classes (Asthma and diabetes), from these examples, we construct a tree said decision.

For these reasons, suppose our learning sample is composed of 14 patients. s0 the initial partition has a single element denoted s0, which includes all the learning sample with nine (09) individuals (patients) belonging to the class « diabetes» and five (05) belonging to the class «asthma» (see Table 1).

For the construction of the decision tree, we used the algorithm of ID3 method in our ARESS platform [21]. ID3 (Iterative Dichotomiser 3) [6] is a heuristic tree to construct a decision tree. Its principal consists in generating a succession of partitions by splitting nodes of the tree. Its objective is to optimize a criterion of information gain. From the sample of learning method ID3 symbolic processing begins for the construction of the decision tree [6]. Thus, the decision tree can be then exploited to: extract the classification rules concerning our target attribute « Diseases ».

ISSN: 2961-6611

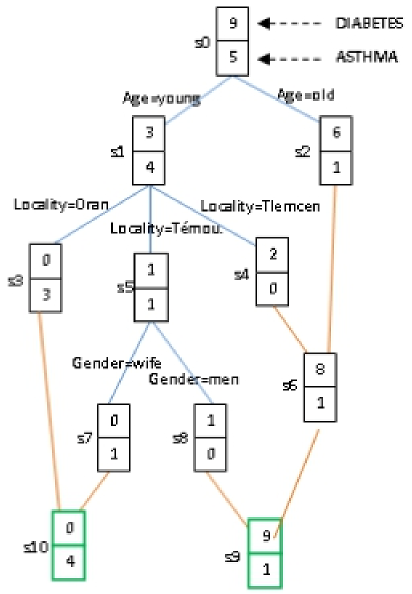

Fig. 6 The induction graph of the table xx sample realized by APESS platform

Finally, we exploited the tree before (see figure 6) to extract five rules R1, R2 … and R5 of inductions epidemiogical of some pathologies (chronic diseases) also on the target attribute « Diseases » useful and bearing knowledge critical have not been explicit in advance and which are of the form: If Condition then Conclusion. Where Condition is a logical expression composed summits which will be called Premise. And Conclusion would be the majority class in the summits described by the condition.

1. **If** (Locality = Oran **and** Age = young) **then** Asthma
2. **If** (Locality = Tlemcen **and** Age = young) **then** Diabetes
3. **If** (Locality = Temouchent **and** Age = young **and** gender = female) ***Alors*** Asthma
4. **If** (Locality = Temouchent **Et** Age = young **and** gender = male) **then** Diabetes
5. **If** (Age = old) **then** Diabetes

F. Exploitation of induction rules

For the experimentation phase, we used the CARTOCEL tool [11], [12], [13] that has been integrated into our APESS platform for extraction and the Boolean modeling the rules of epidemiogical prediction.

Furthermore, we launch the validation phase across the BV module (Boolean validation) on induction rules epidemiogical of some pathologies (chronic diseases) presented in the previous section, using the same Boolean basic principle from Boolean engine inference BIE, and the same transition functions δfact and δrule that exists in CARTOCEL Tool [11], [12], [13].

The figure 7 shows how the Boolean knowledge base extracted starting from the rules induction epidemiogical of some pathologies (chronic diseases) is modeled by layers CELFACT and CELRULE with the same principle CELSUMMIT and CELARC [11], [12], [13]. Note that in this step; the two incidence matrices of input (RE) and output (RS) are generated.

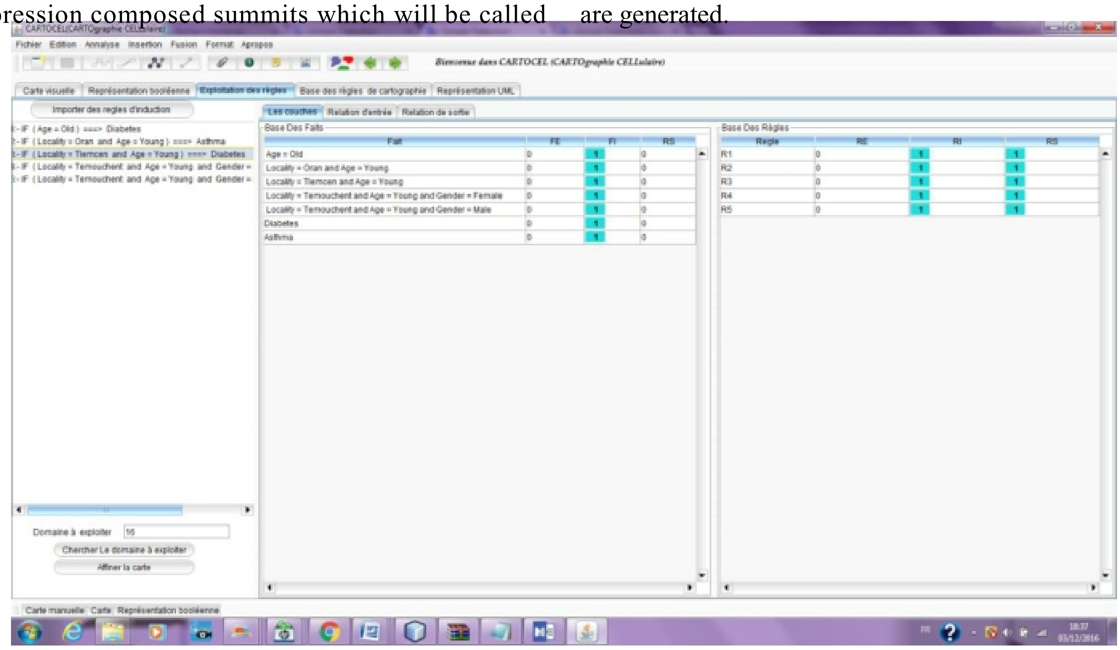

Fig. 7 The induction graph of the table xx sample realized by APESS platform

Nevertheless, we want to highlight the fact that the data warehousing has represented a major task in the implementation of the project, especially the data collection. We have nevertheless be able to put in APESS platform over thirty million data records of sales, spread out between January 2010 and April 2016, and related to 230 pharmacies distributed over 10 departments. Further, it should be noted that these records represent raw data of sales on which no form of aggregation was carried out, to finally obtain, after the pre-processing and the task of identification of the

ISSN: 2961-6611

characteristics of the patients (Gender and Age), nearly 500,000 sales transactions for the selected diseases (Asthma and Diabete). Furthermore, our experiment action is related to a sample of 80,112 sales transactions.

Finally, the objective and the automatic improvement the knowledge mapping critical epidemiological of SEMEP [11], [12], [13] guided by integration technique of heterogeneous data and the oriented-services data mining (APESS platform) in order to use it as a decision support tool, whether individual or collective in the public health for management and monitoring of some pathologies (chronic diseases).

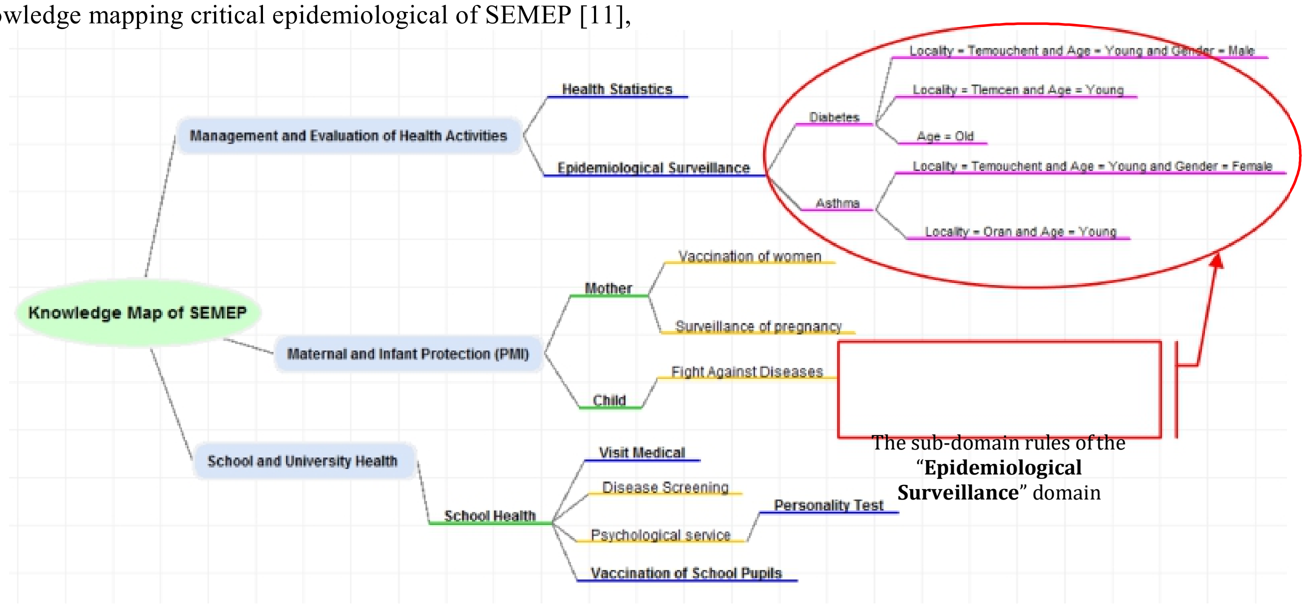

Fig. 8 The Knowledge map of SEMEP refined by a process of knowledge extraction by APESS platform

IV. Conclusions

Knowledge management has aimed at extraction of useful knowledge from such sources and its representation in several forms. Likewise, the health institutions have realized the importance of huge amount of data and information medicals that are collected to make enhanced decisions concerning chronic diseases. Moreover, we recorded and we are convinced of the interest of applying the techniques of data warehouse and oriented service data mining on several types of information stored in various media (Database, ERP, CRM, SCM, XLT, etc..) to improve the extraction process and knowledge management process and, in particular the knowledge mapping of the epidemiological services.

In this context, this paper aimed to propose a new medical knowledge mapping approach based, on the one hand, on integration technique of heterogeneous data in the medical field by creating a data warehouse and, on the other hand, on a technique of extracting epidemiological prediction rules from medical data by choosing a technique of service-oriented data mining of APESS platform whose objective to exploit these predictive rules to automatically improve the Boolean model of the medical knowledge mapping through CARTOCEL system. Finally, this approach of epidemiological knowledge mapping would be new direction towards as a decision support tool, whether individual or collective in the public health for management and monitoring of some pathology (chronic diseases).

Acknowledgment

This project is launched in collaboration between the health actors, our research LSSD laboratory (Laboratory Signal, System and Data) in USTO-MB University, the National Polytechnic School of Oran and TechCICO laboratory in university of Technology of Troyes (France). The authors also acknowledge the pharmacists team for her assistance on to finalize this project.

References

[1] A. Perez-Soltero, V. Leal, Barceló, M. & J. León, ―A diagnostic of knowledge management processes at the restaurant industry SMEs to identify improvements at their productive processes‖, Intangible Capital (IC), 9 (1), 2013, p.153-183.
[2] D. Christozov, & S. Toleva-Stoimenova, The role of information brokers in knowledge management. Online Journal of Applied Knowledge Management, 2 (2), 2014, p.109-119.
[3] D. Christozov, & S. Toleva-Stoimenova, ―Knowledge diffusion via social networks: The 21st cen-tury challenge. International Journal of Digital Literacy and Digital Competence, 4(2), 2013, p.1-12.
[4] D. Talia, P. Trunfio, and O. Verta, ―Weka4WS: a WSRFenabled Weka Toolkit for Distributed Data Mining on Grids‖. Proceeding of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2005.
[5] H. Nemati, D. Steiger, L. Iyer, & R. Herschel, ―Knowledge warehouse: an architectural integration of knowledge management‖, decision support, artificial intelligence and data warehousing. Decision Support Systems, 33(1), 2002, p.143-61.

ISSN: 2961-6611

[6] J.R Quinlan. Induction of decision trees Wadsworth Statistics/Probability,Berkeley, 1986.

[7] L. C. Cámpiz Mercado, ―Influence of Critical Success Factors of Knowledge Management on the Innovation Performance of Colombian Organizations‖. 8th LACCEI Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2010) ―Innovation and Development for the Americas‖, June 1-4, 2010, Arequipa, Perú, p. 1-10.

[8] K. Curko, V. Vuksic, & A. Lovric, The Role of Business Process Management Systems and Business Intelligence Systems in Knowledge Management. International Journal of Computer and Communications, 3,(2), 2009, p.17-24.

[9] L.Wulantika, ―Knowledge Management Dalam Meningkatkan Kreasi Dan Inovasi Perusahaan‖, Majalah Ilmiah UNIKOM, Vol.10, No. 2. P. 263 -270.

[10] M.B. Lloria, ―A review of the main approaches to knowledge management‖, Knowledge Management Research and Practice, 6, p.77-89 (2008).

[11] M. Brahami, B. Atmani, & N. Matta, ―Using rules to enhance evolution of knowledge mapping: Application on healthcare‖. *International Journal of Computer Science Issues (IJCSI)*, *10* (3), 2013, p.261-270.

[12] M. Brahami, Boolean Knowledge Mapping guided by Data Mining, PHD Thesis, Department of Computer Science, University of Oran, 2014.

[13] M. Brahami, B. Atmani, & N. Matta, ―An Approach to Dynamic Fusion of the Knowledge Maps of an Activities Process: Application on Healthcare‖. International Journal of Information Systems in the Service Sector, 7(4), 2015, p.1-26.

[14] M. Cannataro, & D. Talia, The knowledge grid, Communication of the ACM 46 (1), 2003, p.89–93.

[15] M. D. Beynona, T. Kurcb, U. Catalyurekb, C. Changa, A. Sussmana, & J. Saltza,.‖Distributed processing of very large datasets with DataCutter‖. Clusters and computational grids for scientific computing, 27 (11), 2001, p. 1457–1478.

[16] M.E. Jennex, I. Zakharova, Knowledge Management Critical Success Factors.

[17] M. Gadu, & N. El-Khameesy, ―A Knowledge Management Framework Using Business Intelligence Solutions‖. International Journal of Computer Science Issues, 11 (5), 2014, p.102-106.

[18] M. Plessis, ―Knowledge management: what makes complex implementations successful?‖, Journal of Knowledge Management, Vol. 11, 2, p.91 – 101.

[19] M. Montoro, ―Gestión del conocimiento en las organizaciones: fundamentos, metodología y praxis‖, España: Ediciones Trea, S.L .

[20] M. Q. A. Al-qarioti, ―The impact of knowledge management on organizational performance: an empirical study of kuwait university‖. Eurasian Journal of Business and Management, 3(4), 2015, p. 36-54.

[21] M. Sabri, and S. A. Rahal, APESS - A Service-Oriented Data Mining Platform: Application for Medical Sciences, International Journal of Information Technology and Computer Science, 2016, 7, 36-42.

[22] M. Sonesson, & R. Storgren, The relationship between Knowledge Management and Business Intelligence: How the concepts are affected by each other inside organizations. Master thesis, 2016, Department of Informatics, Lund University School of Economics and Management.

[23] M. S. Perez, A. Sanchez, V. Robles, P. Herrero, & J. M. Pëna, Design and implementation of a data mining grid-aware architecture. *Future Generation Computer Systems*, *23* (2007), p.42–47.

[24] M. Zack, J. McKeen, and S. Singh, ―Knowledge management and organizational performance: an exploratory survey‖, Journal of Knowledge Management, Vol. 13, (6), p. 392-409.

[25] N. M. Yaghoubi, J. Moloudi, and S. Ali Banihashemi, The relationship between knowledge management and agile supply chain management: Case study of Jihad-e-Agriculture Organization. African Journal of Agricultural Research, Vol. 8(17), 2013, p. 1700-1708.

[26] P. Brezany, I. Janciak, A. Wöhrer, & A. M. Tjoa, GridMiner: A Framework for Knowledge Discovery on the Grid - from a Vision to Design and Implementation. Proceedings of the Cracow Grid Workshop, December 2004, Cracow, Poland.

[27] R. A. Khan, and S.M. K. Quadri, Business Intelligence: AN Integrated Approach, Business Intelligence Journal - January, 2012 Vol.5 No.1, p. 64-70.

[28] S. Wamundila, and P. Ngulube, ―Enhancing knowledge retention in higher education: a case of the University of Zambia‖, South African Journal of Information Management, 13(1), p.439-448.

[29] S. Negash, & P. Gray, ―Business intelligence‖. In F. Burstein, & C. W. Holsapple (Eds.), Decision support systems, 2008, p. 175-193, Berlin: Springer.

[30] S. Alsairafi, M. Ghanem, N. Giannadakis, Y. Guo, D. Kalaitzopoulos, M. Osmond, A. Rowe, J. Syed, P. Wendel, ―The design of discovery net: Towards open grid services for knowledge discovery‖. International Journal of High Performance Computing Applications 17(3), 2003, p.297–315.

[31] T. H. Davenport. Competing on analytics. *Harvard business review*, *84*(1), 2006, p.98.

[32] T. Nguyen Manh, J. Schiefer, and A.M. Tjoa, ―Data warehouse design 2: sense & response service architecture (SARESA): an approach towards a real-time business intelligence solution and its use for a fraud detection application‖, Proceedings of the 8th ACM International Workshop on Data Warehousing and OLAP, DOLAP '05, ACM Press, New York, NY.

[33] V. Georgescu, WSRP-Enabled Distributed Data Mining Services Deliverable over a Knowledge-Driven Portal. 7th WSEAS International Conference on Applied Computer & Applied Computational Science (ACACOS '08), April 6-8, Hangzhou, China.

[34] V. Lomte, & S. Shah, ―The Survey Paper on Importance of Integration of Knowledge Management and Business Intelligence‖, International Journal of Science and Research (IJSR), 3 (6), 2014, p.1789-1791.

[35] V. Podpecan, M. Zemenova, & N. Lavrac., ―Orange4ws: environment for service-oriented data mining‖. The Computer Journal, 55 (1), 2011, p. 82-89.

[36] V. Stankovski, M. Swain, V. Kravtsov, T. Niessen, D. Wegener, J. Kindermann, & W. Dubitzky, ―Grid-enabling data mining applications with DataMiningGrid: An architectural perspective‖. Future Generation Computer Systems, 24(4), 2008, p.259–279.

[37] Y. Wu, S. Wu, H. Yu, & H. Hu, ―Introduction to ChinaGrid Support Platform‖. Proceedings of 7th International Conference on Asian Digital Libraries, ICADL 2004, December 13-17, Shanghai, China.

[38] Z. Abas, and M. N. Jali., ―Understanding Knowledge Management in Developing Emerging Concept of Innovation and Technology Into Business: Conceptual Review and Empirical Evidence‖. International Academic Research Journal of Business and Technology 1(2) 2015, p. 149-164.

[39] Z. N. Boudjettou, F. Nader, & R. Chalal, ―Medical Knowledge Management since the Integration Heterogeneous Data until the Knowledge Exploitation in a Decision-making System‖. Recent Advances in Environmental Science and Biomedicine, Proceedings of the 3rd International Conference on Health Science and Biomedical Systems (HSBS '14), November 22-24, 2014, Florence, Italy.

[40] Z. Panian, ―Business Intelligence inSupport of Business Strategy‖, Proceeding of the 7th WSEAS International Conference on Mathematics & Computer in Business Economics, Cavtat (Croatia), 13-15 June, 2006, p. 204-215.

ISSN: 2961-6611