Public Healthcare and Epidemiology with Dr Warehouse

Vladimir Ivančević, Marko Knežević,
Miloš Simić, Ivan Luković
University of Novi Sad, Faculty of Technical Sciences
Novi Sad, Serbia
e-mail: dragoman@uns.ac.rs,
marko.knezevic@uns.ac.rs,
milossimicsimo@gmail.com, ivan@uns.ac.rs

Danica Mandić
Institute of Cardiovascular Diseases of Vojvodina,
Clinic of Cardiology
Sremska Kamenica, Serbia
e-mail: mandiceva88@yahoo.com

***Abstract*—The need for the systematic collection and use of epidemiological data, together with the undergoing modernization of the public healthcare system in Serbia, has motivated us to develop Dr Warehouse, an extensible intelligent software system for the collection, presentation, and analysis of data from epidemiological and public healthcare sources. The central point of the system is a data warehouse where medical data about registered disease cases and relevant demographic data are being collected. Through a web application and mobile device client, different categories of users may access data that are of interest to them, perform built-in analyses, or test their own epidemiological hypotheses. Dr Warehouse is expected to provide intuitive visualization of epidemiological data, facilitate discovery of epidemiological knowledge, and support modelling of epidemic dynamics. We discuss our motives for building such a system, the architecture of the system, our choices regarding data modelling, and the built-in functionalities. We implemented a foundation for different analyses that are expected to provide valuable insights if properly adapted and used in practice: investigation of diagnosis change over time, forecasts based on data mining, and compartmental models of disease dynamics.**

Keywords-business intelligence, public healthcare, epidemiological analysis, absenteeism, disease outbreak prediction.

I. INTRODUCTION

As a result of combining the latest advancements in business intelligence and data analysis to the domains of public healthcare and epidemiology, we present an extended version of the previously published overview of Dr Warehouse [1] – a closed source software system that supports storing of medical and epidemiological data, while offering descriptive, as well as predictive, analyses of disease cases and epidemics. There are several key issues that motivated us to develop such a solution. Despite numerous medical discoveries, lifestyle improvements, and strategies to battle epidemics, disease elimination and eradication remain as the probably most important goals in disease control [2]. Epidemiologists continue to collect outbreak data and analyse the dynamics of various ever-changing diseases in order to better understand their nature and, consequently, devise new effective countermeasures. With the proliferation of information technology, public healthcare has entered a new era with its own set of opportunities and challenges [3]. However, the increased possibilities in data collection and analysis have led to problems with the systematic treatment and use of available data [4][5]. On the other hand, the modernization of the public healthcare system in Serbia includes, among many tasks, a switch to electronic records and increase of the availability of medical information to all people involved in public healthcare. In a situation where many institutions of public healthcare conduct research separately using their in-house devised approaches, having a common point where epidemiological data could be stored could promote cooperation of these institutions and publishing of up-to-date epidemiological information to general public. As a response to these issues, we provide a potential solution in the areas of public health and epidemiology by building a software system that could help in the prevention and control of epidemics. This could be achieved by providing many procedures for different types of epidemiological research and a single data source that is tailored to the need for frequent analyses.

Within Dr Warehouse, all medical and epidemical data are stored in one such central data source, a specially designed data warehouse. Supported analyses include various data visualization techniques, statistical methods, analysis of absenteeism data, data mining algorithms, and compartmental epidemic models. Results of the analyses may be accessed through a rich web client application, which offers all of the analyses included in the system, or a mobile device client, which offers a subset of analyses primarily tailored to the needs of non-experts. Given the rapid rate of discovery of new analysis methods and epidemic models, we made the system extensible and ensured that new types of analyses and data visualization may be easily added.

The paper is organized in six sections, including Introduction. Section II offers a review of similar software systems and a comparison of their capabilities to those featured in Dr Warehouse. Section III presents our motives for building the Dr Warehouse system. In Section IV, there is an overview of the system, its architecture, featured data warehouse, and functionalities of client applications. Some

of the descriptive and predictive analyses supported by the system are presented together with sample results in Section V. Section VI includes concluding remarks and ideas for further research.

II. Related Work

There are numerous software systems for epidemiological analyses and monitoring. One group of such systems provides mostly statistical procedures that are often used in epidemiology. Open Source Epidemiologic Statistics for Public Health (OpenEpi) [6] is an example of a freely available system that may be run in a web browser [7] because it is implemented in HyperText Markup Language (HTML) and JavaScript. It focuses on statistical calculations: calculation of confidence interval and sample size, estimation of power for different types of studies, execution of various statistical tests, etc. Another free solution is WinPepi [8], which is a set of desktop applications that are similar to OpenEpi and offer many statistical procedures that are useful in epidemiology. When compared to Dr Warehouse, both OpenEpi and WinPepi are projects of a narrower scope because they ignore data storage and management. Furthermore, they put emphasis on statistics and a large number of calculation modules whose input is mostly a small set of summarized values. Unlike Dr Warehouse, they do not support data mining, visual representation of data, epidemiological maps, nor user extensions. However, the source code of OpenEpi may be directly modified to include new procedures.

The second group of epidemiological systems includes data storing and manipulation capabilities in addition to analysis procedures. Epi Info [9] is one such example of a desktop software application with a wider range of functionalities than OpenEpi and WinPepi. What sets it apart from other software systems for epidemiology is support for form creation. A user may design custom forms through an integrated editor and later use them for data entry. Besides basic and advanced statistical procedures, this system supports data import and export, as well as basic data selection and transformation. It has good data visualization capabilities and offers various types of charts, tables, and even map overlay. Its main strengths with respect to Dr Warehouse are support for form creation, direct data entry, data transformation and data import/export for various types of data sources. However, there is a conceptual difference between these two systems regarding data storage. Epi Info is a tool that may be used over any data (in the supported file or database format) and, therefore, provides transformation functions, which a user utilizes in order to prepare data for analyses. On the other hand, Dr Warehouse features a data warehouse with a fixed set of facts and dimensions, and a carefully designed ECTL process, which is automatically executed. Therefore, there is generally no need for manual data import and transformation because data preparation is done automatically. In other words, Dr Warehouse may be seen as a more specialized and more automated solution in which the data warehouse has a prominent role. Our system relies on a strong dependency between the data warehouse schema and analyses, which helps to simplify the analysis process. This, in turn, alleviates much of the burden concerning data preparation, which is usually the longest activity in analysis projects. Some of the main features of Dr Warehouse that Epi Info lacks are data mining procedures and the support for adding user extensions. We consider data mining to be an essential part of the system because, unlike most statistical procedures, it is well suited for analysing large quantities of data that are efficiently stored in a data warehouse. We may summarize this comparison by generally classifying Epi Info as a solution that offers a fixed set of analyses for any set of data attributes and Dr Warehouse as a solution that features a fixed set of data variables but an extensible set of techniques for data presentation and analysis.

The third group consists of typically web-based systems that focus on epidemiological monitoring and publicly presenting latest disease outbreak data for different regions throughout the world. They primarily rely on data from numerous Internet-related sources, which may be informal or official. HealthMap [10] provides a world map with the latest information on outbreaks by automatically collecting and integrating data mostly from several online news sources and reports from eyewitnesses and officials. There is also a mobile version of the system with similar functionalities. Another web system with a support for mobile devices is Outbreak Watch [11]. It does real-time analyses of data in social networks by evaluating keywords that are considered to be indicators of outbreaks. In this manner, the system tracks changes in the number of reports concerning relevant diseases. Google Flu Trends [12] was created as an attempt to estimate actual flu activity in various countries by analysing aggregated Google search queries that are related to flu. Since there is a relationship between an actual number of flu cases and search queries about flu, as confirmed by the overall match between the official surveillance data and the calculated estimates, this service offers near real-time results, which may help in preparing a response to a flu outbreak. Dr Warehouse is similar to these systems, as it may offer latest epidemiological data and forecasts in the form of charts, tables, and maps. In addition to supporting web access, it also features a mobile version with a selected set of services. On the other hand, the principle difference lies in the selection of data sources. The three monitoring systems use data that are available on the Internet (HealthMap and Outbreak Watch) or from web search queries (Google Flu Trends), while Dr Warehouse displays only data present in the data warehouse, which was planned to include credible data collected in healthcare institutions. However, the ECTL process in Dr Warehouse may be extended in the future to include data from public web sources.

When compared to the three aforementioned groups of epidemiological software, Dr Warehouse is a complex system that possesses traits typical of all three because: (i) it may offer any statistical procedure that has been added as an extension; (ii) data management is one of the key segments of the system; and (iii) collected data are constantly available to users via web and mobile client, which makes the system suitable for epidemiological monitoring. As a result, we consider the following two characteristics to be its major

advantages over the other solutions: (i) versatility, i.e., suitability for healthcare institutions, epidemiologists and general public; and (ii) comprehensiveness, i.e., support for data collection and storage, epidemiological analysis, data presentation, and functionality extension.

III. MOTIVATION

In addition to the prediction of epidemics and understanding of disease dynamics, we are motivated by two more specific reasons: modernization of the healthcare system in Serbia and impact of absenteeism on the economy.

As outlined in the national development strategy [13], the Serbian healthcare system is undergoing a significant transformation. Many segments of that system are being modernized to rely more on electronic records as opposed to traditional paper records. Moreover, the expected interconnection of healthcare centres would allow a better electronic access to medical data and consequently better conditions for data analyses, as in the case of the health information system (HIS) for the Serbian Ministry of Defence [14]. In such circumstances, Dr Warehouse could be integrated into the main healthcare system, which would act as a data source. After several processing steps, these data would be stored in the data warehouse within the Dr Warehouse system. Dr Warehouse has been developed also as a pilot solution that should demonstrate advantages of using a business intelligence (BI) system in the healthcare domain. It is primarily applicable within institutions that deal with disease prevention, such as institutes of public health.

Absenteeism is defined as "failing to report for scheduled work" [15]. High absenteeism has negative impact not only on colleagues and superiors, who must cope with greater workloads, but also on the profit. According to the 2009 research by the Chartered Institute of Personnel and Development (CIPD) from Great Britain [16], the most important reasons for the short-term work absence (up to four weeks) are: colds, influenza, stomach problems, headaches, migraines, injuries of the muscular and skeletal system, and pain in the lower back part. Most of these conditions, which are also a major health problem in Serbia, are preventable non-communicable diseases (NCDs). However, there is no adequate prevention and control of NCDs in Serbia [17]. We hope that, by using Dr Warehouse, valuable absenteeism patterns could be uncovered.

There are three primary groups of users who might benefit from the developed system: employees in public healthcare institutions, researchers in epidemiology, and non-experts interested in epidemiological information. Users in public healthcare institutions that are dealing with epidemiological data could utilize our software system, which is specially tailored to the epidemiological domain, instead of relying on solutions that are intended for generic statistical analyses. An expected advantage of having a domain-specific system would be an increase in user productivity. Large amounts of data that are typical of modern HISs may be well utilized owing to the well-tried approach incorporated into our system – a data warehouse for data storing and data mining for efficient analyses. The main system load may be reduced by running analyses primarily on data stored in Dr Warehouse. The second group of users are scientists whose research is related to epidemiology. By utilizing Dr Warehouse, they may create, test, and improve epidemic models through adding, running, and modifying new extensions. New visualization techniques for epidemiological data may also be employed and evaluated. Furthermore, the system may also target users who are not medical experts but are interested in latest disease trends, forecasts, or results of some specific analysis.

Owing to the adverse health of the population in Serbia and the "white space" in terms of medical services aimed at predicting disease occurrence, our decision to develop a system that would allow the use of BI technologies in such a context should be both socially and economically justified.

IV. SYSTEM OVERVIEW

In this section, we present the system and give an overview of its architecture and functionalities. Public resources concerning the system are available at [18]. The featured data warehouse, which represents a foundation for data analyses, is explained in more detail. We also elaborate on the featured web application, mobile application, and built-in support for adding new functionalities.

A. System Architecture

There are four principal components in the system: (i) database server, (ii) application server, (iii) web client application, and (iv) mobile device client application. The overview of the system is given in Fig. 1.

The database server is depicted as a rounded rectangle titled *SQL Server* (in the left portion of Fig. 1). It consists of a relational database management system (*SQL Server RDBMS*), which hosts a data warehouse containing epidemiological and medical data; services for data analyses that focus on data mining and OLAP cube analytics (*SQL Server Analysis Services*); and data integration services (*SQL Server Integration Services*) for data extraction, transformation, and loading (the ETL process) from the supported types of data sources (Excel files, various relational databases, data files, or some other external sources).

The application server is depicted as a rounded rectangle titled *Dr. Warehouse Server* (in the central portion of Fig. 1). It acts as an intermediary between the database server and the two client applications. This component communicates with the database server using the *ADOMD.NET* subcomponent, which is responsible for providing access to analytical data sources. On the other hand, web services (*WCF Services*) are used to exchange information between the application server and the two client applications. New functionalities (*Extensions*) may be added to the application server using the *MEF* subcomponent.

The web client application, which is depicted in Fig. 1 as a rounded rectangle titled *Silverlight Client Application*, may be extended in the similar manner using its own *MEF* subcomponent. It also supports reading of Serbian identity cards (ID cards) using the *Smart Card Reader* module. The mobile device client is depicted in Fig. 1 as a rounded rectangle titled *Windows Phone 7*.

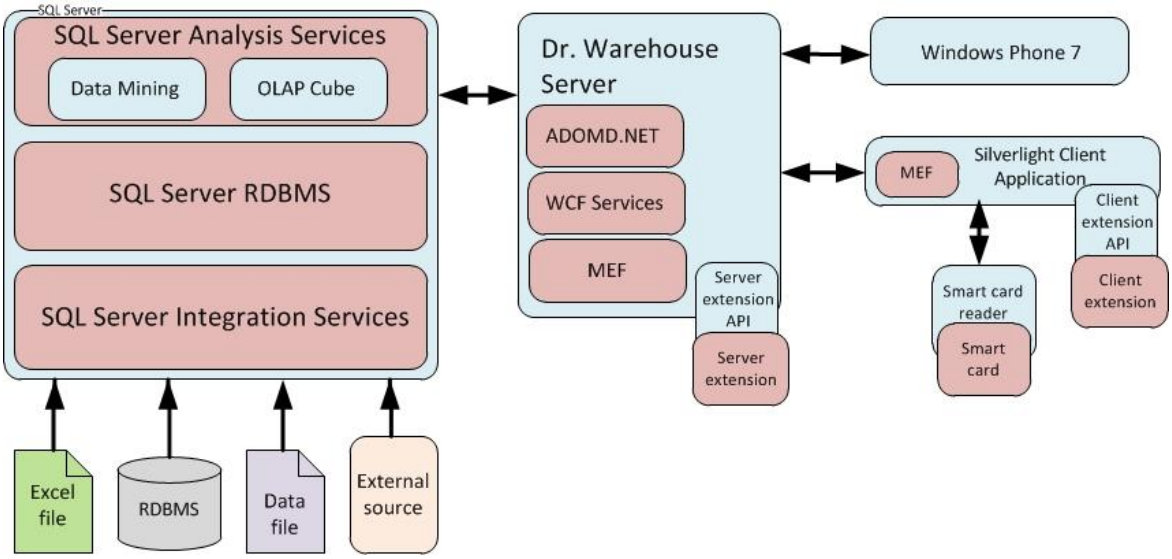

Figure 1. System overview.

The system may fit into existing HISs and provide various services to other similar solutions. This architecture allows the possibility of having the database server and application server reside at different physical locations. Furthermore, in order to increase the scalability and performance of the system, the data mining and analysis services (currently implemented using Microsoft SQL Server Analysis Services [19]) may be located separately from the database server. In future versions of the system, the architecture may be extended to include terminals that would be publicly available and offer a set of functionalities similar to those in the existing web client application.

B. Data Warehouse

The data warehouse is modelled using a star schema, which consists of eight dimensions, two of which are role-playing dimensions, and one fact table (Fig. 2). The fact table keeps track of events that lead to absenteeism, disease occurrences and time measured in days that person spent away from duty or workplace. Each dimension represents the context of disease occurrence and absence. Therefore, we can observe these events in the context of time (when an event occurred or ended), gender of the person involved, place where it happened, person's profession, data source, absence cause, person's age, and diagnosis that was established.

Dimensions concerning diagnosis, place, and time have several hierarchical levels modelled as a fully denormalized structure, which enables multi-level classification of factual data. In the time dimension, we have two hierarchies: one defined as calendar year, quarter, month, and day, and the other one as calendar year, week, and day.

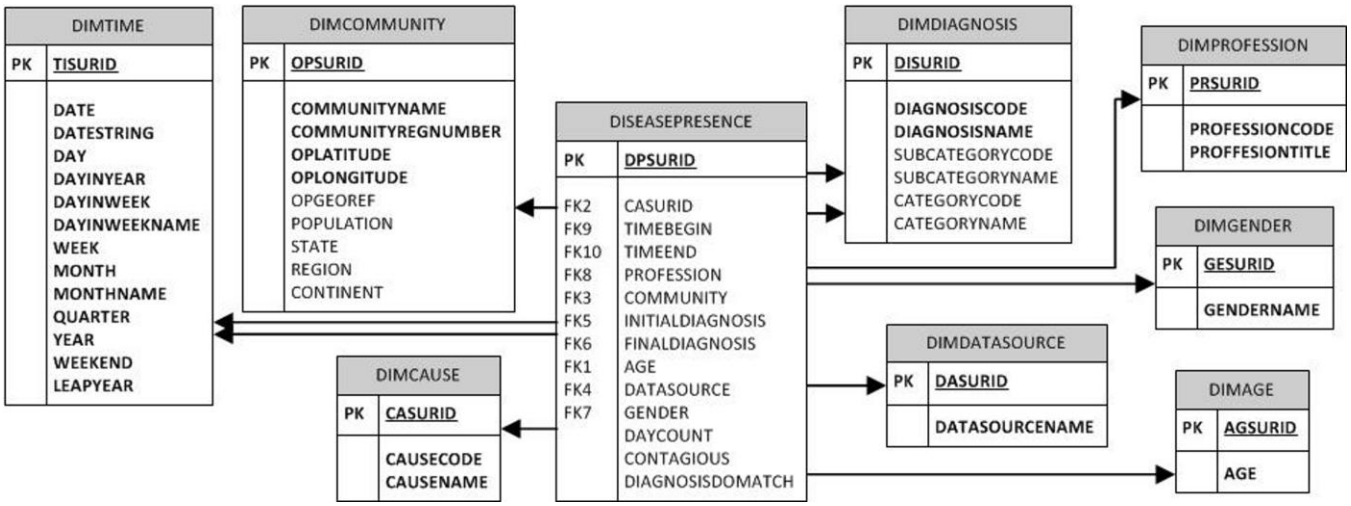

Figure 2. The star schema of the data warehouse.

The diagnosis dimension has three levels of hierarchy for diagnosis, disease subcategory, and disease category. Within the community (place) dimension, there are four levels of hierarchy for community, state, region, and continent.

Although the normalization of our schema would remove redundant data, which in turn would make the schema easier to maintain and change, our initial considerations of the schema type led us to choose the star schema. Denormalization, which is typical for the star schema, helped us to reduce the number of foreign keys and to reduce the query execution time. As the system was designed to be used by a wide variety of users, ease of use was one of our priorities. For end users, the star schema is more comprehensible than snowflake schema and less complex queries are needed to satisfy their information needs. Since this is a pilot project, advanced cost-benefit analysis of normalizing our star schema into the showflake schema is a matter of our future work. The unavailability of a larger and more complex absenteeism data set was a major reason for simplifying the initial schema design and focusing on the aforementioned fact and dimensions.

The data warehouse was implemented using Microsoft SQL Server 2008 [20]. It includes the following dimensions: *DimCause*, *DimDiagnosis*, *DimGender*, *DimProfession*, *DimCommunity*, *DimDataSource*, *DimTime*, and *DimAge*. *DiseasePresence* is the only fact table in the system. Each of these tables contains a surrogate primary key, which allows us to deal with changes in natural key in a more convenient way and track slowly changing dimensions.

Taking into consideration that the data in the system are expected to reflect the actual state of the health of a population, it is necessary to support acquisition and integration of medical data from multiple sources. We developed a solution within Microsoft Integration Services [21], which allows us to extract, clean, transform, and load (ECTL) the necessary data. The ECTL process is divided in six SQL Server Integration Services (SSIS) packages each of which covers control and data flow between sources (Excel files or databases) and corresponding fact or dimension table. Furthermore, for each package, set of actions is specified within SQL Server Agent [22] jobs. Those actions involve preparing source files, configuring connections to data flow sources and destination, executing the packages and backing up old source files. Moreover, the execution of SSIS packages is logged and completion status is emailed to an administrator.

We perform incremental extraction, i.e., we consider only data that were added to the HIS of a public health institute or uploaded to application sever after the previous extraction. Extracted data serve as an input for a series of transformations in which we detect and eliminate errors and inconsistencies: (i) different domains of semantically equivalent attributes (as in the case of the attribute *GenderName*); (ii) different encodings of textual data (*ProfessionTitle*); (iii) different granularity of semantically equivalent attributes (*DiagnosisCode*). Diagnosis codes that are used in the source HIS are shorter versions of the codes that are featured in the 10th revision of International Classification of Diseases (ICD 10) [23]. We created a transformation that relies on regular expressions to resolve this issue. In this manner, we extended disease information with the disease name, subcategory and category. Dimensions *DimGender*, *DimTime*, and *DimAge*, which are static dimensions, are not extracted from data source. They are created within the context of the data warehouse and their records are either loaded manually or generated by a custom procedure.

At the moment, there is only support for data insertion. Since the data set in the current version of the system is only a sample taken from a HIS, we decided to keep all data in the data warehouse, while leaving the implementation of a deletion policy for obsolete data and data that have little or no impact on the system output, to be included in the future version.

In order to meet the needs for efficient and flexible consumption of valuable information produced by the system, we developed an online analytical processing (OLAP) database, which contains rich metadata. The OLAP cube makes our data organized in a way that facilitates non-predetermined queries for aggregated information. As we used Kimball Method [24] to implement the dimensional model in the relational database, the OLAP design step was a straightforward translation from the existing design. The relational database serves as the permanent storage of the cleaned and conformed data, and feeds data to the OLAP database. Data mining structures and models are stored in the third database, which, together with the OLAP database, resides at the Analysis Server – the primary query server in the system.

C. Web Application

The majority of the functionalities that are available to expert users are incorporated into a web application, which is implemented in Microsoft Silverlight [25]. The communication between the application server and the web client is done via web services using Microsoft Windows Communication Foundation (WCF) [26]. At present, the client application possesses functionalities concerning: access to medical records stored in the data warehouse; upload of data files containing medical records; access to the data cube and use of some of the cube's advanced analytical operations; execution of advanced analyses and forecasts, as well as result retrieval; upload of extensions; and their invocation.

The contents of the web application are organized as a set of Silverlight pages, where each page groups a number of similar functionalities. Within the *Home* page, users may access a chart about the most common causes of absenteeism for the current month. Moreover, in order to fulfil the needs of more experienced users accustomed to traditional reports, we created various operational reports within the server-based reporting platform, SQL Server Reporting Services (SSRS) [27]. Besides data sheets, these reports include rich data visualization in form of 3D charts. Users may access reports on-demand through a web browser. After they run a report, they can export it to another format, such as Excel spreadsheet or PDF. The functionalities of other pages are presented in the remainder of the subsection.

1) General Predictions

The *Analysis* page contains functionalities regarding the execution of advanced analysis and forecasts that identify the most probable diseases (or causes of work absence) and the most frequent diagnosis mismatches. Through this page, a user is able to generate basic predictions concerning a selected subpopulation for a particular quarter of a year. The subpopulation may be specified by selecting an age group, gender, and municipality (Fig. 3). More information about these predictions may be found in Section IV-C.

2) Personalized Predictions

The *What about me?* page is a location from which it is possible to generate and retrieve results of the personalized predictions concerning the most probable diseases (or causes of work absence). In order to generate these predictions, a user is required to insert his or her ID card into the attached smart card reader and a report is automatically generated.

Card data are read with the help of the Čelik API [28], which is primarily intended for integration of ID cards into business systems. In order to access the smart card reader from a web browser, we had to enable trusted applications to run inside the browser. Moreover, as soon as health smart cards become publicly available in Serbia, we intend to adapt the system to allow data reading from these cards as well.

3) Analytical Operations

The execution of analytical operations and access to historical data is provided within the *Health Reports* page. Users may perform operations such as dice and slice in order to analytically process the available data. The *PivotViewer* control [29], which is an integral part of the page, helps users to interact with thousands of items at once and see trends and patterns that would be hidden when looking at one item at a time. Fig. 4 shows the *PivotViewer* control in which green squares represent diagnoses and the numbers of their occurrences among males of a certain age. The user is able to zoom in and select squares in order to retrieve more information on the corresponding diagnosis such as description and number of occurrences.

4) Upload of New Data and Functionalities

The *Upload* page offers functionalities regarding uploading of server and client extensions. Within the *Extensions* page users may activate and run uploaded extensions. Furthermore, through the *Upload* page users may upload Excel files containing medical records. As most HISs support exporting data in the form of Excel files, we considered it to be most suitable format for the task. Every uploaded file is stored with a unique name, which prevents file name collisions. In order to minimize the impact of the intense ECTL activity on user experience, we scheduled a server job that executes SSIS package every day at 12:00 pm.

However, we are going to reconsider this decision once we gather concrete data on the frequency of data upload and user activity within the web application. The first step in the job is reserved for preparing the uploaded file for the ECTL process and the last step for backing up the file. All steps within the job are repeated until there are no more files in a directory on the server where uploaded files are stored.

5) Diagnosis Discovery Support

The *Symptoms Checker* page is part of our more recent research, which resulted in the functionality that supports medical diagnosis discovery. The diagnostic process is based on symptom matching in a way that is sensitive to the diagnoses that are dominant in the population to which the patient/user belongs. This functionality, which is presented in more detail in [30], utilizes data sets presented in Section IV-A, and a two-phase algorithm that is based on the differential diagnosis method from medical diagnostics and predictive models for disease occurrence in a subpopulation.

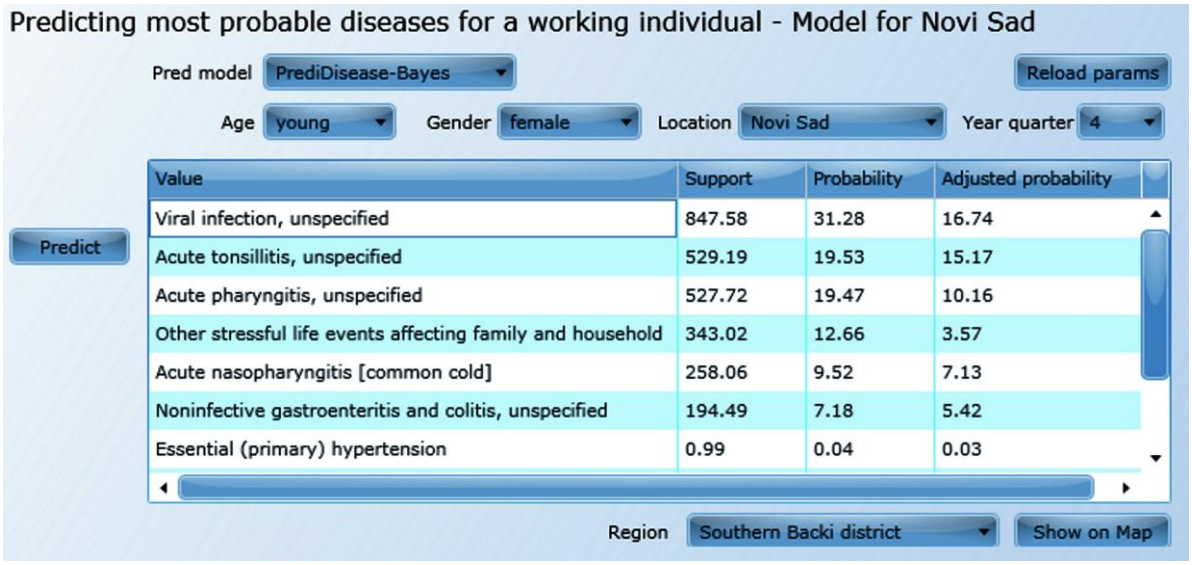

Figure 3. Section from the Analysis page in the web client.

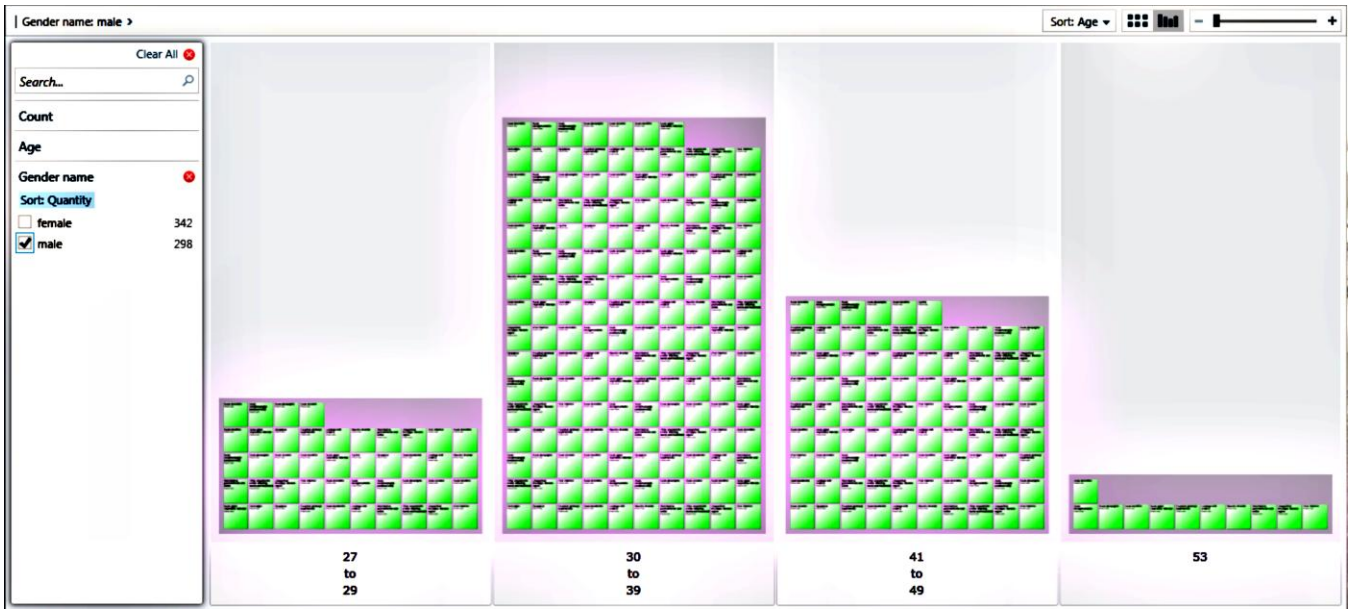

Figure 4. Data visualization using PivotViewer.

D. *Mobile Application*

Dr Warehouse Mobile is a mobile client for the Dr Warehouse system. It is an application designed for smartphones and implemented only for the Microsoft Windows Phone platform [31]. However, porting the mobile application to other platforms such as Android and iOS is a matter of future work. Given the fact that smartphones are widely used, we have chosen to offer a mobile application as a way of integrating the collected medical information and predictions from Dr Warehouse into regular activities of potential users.

The mobile application has functionalities similar to those offered through the web client. However, it has a narrower set of features that are customized for mobile users. Its functionalities are organized into several segments: *Login Process*, *Registration Process*, *Current Position*, *Position Information*, *Personal Information*, and *Symptoms*. In order to exchange data between the mobile client and application server, we developed WCF web services tailored for the mobile client. The methods for determining the common diseases and potential diagnoses are the slightly modified versions of the corresponding methods available through the web application.

Once started, the mobile application presents a login screen (*Login Process*) with the option to navigate to the user registration form. When registering in the system (*Registration Process*), a user provides personal information such as age, gender, profession, and place of residence. The provided information is used when performing a personalized analysis similar to that featured in the *What about me?* page in the web client.

After a successful login, the user is provided with the position (*Current Position*) obtained via the GPS receiver of the phone and Microsoft TerraService [32]. With this information, the user may inquire about the most common diseases at the current location (*Position Information*). For this purpose, the client issues an asynchronous call to a web service. Once the response is received, the user is informed via the toast pop-up notification and information about the most common diseases is shown on a map. Moreover, the same information, in the textual or chart form, may be obtained for any location that is designated by the user.

In the similar manner, the user may retrieve a list of diseases that are most probable for the subpopulation to which the user belongs, where the subpopulation is determined using the personal information provided during the registration, or for any other specified subpopulation, where the user needs to specify the required information: age, gender, profession, and location. The scenario of obtaining prediction according to the personal information (*Personal Information*) is illustrated in Fig. 5.

The last segment of the mobile application is devoted to providing potential diagnoses for a list of symptoms exhibited by the user (*Symptoms*). The user may create a personal symptom list by adding observed symptoms from a list of the common symptoms, which is regularly updated from the server, or by manually entering the name of a less common symptom. After the symptom list is submitted for the evaluation, the list of potential diagnoses is retrieved from the server. For each diagnosis, there is additional information such as risk factors, treatments, and aliases (Fig. 6). Furthermore, the user may access a separate description of a potential diagnosis. For each provided piece of information about a diagnosis, there is a hyperlink to a relevant resource with a more detailed description.

Figure 5. Personalized predictions within the mobile client.

Figure 6. Description of a potential diagnosis within the mobile client.

E. *Extensibility*

New functionalities may be added to the system in the form of extensions. The support for extensibility was implemented using Managed Extensibility Framework (MEF) [33]. A user may upload an extension, which then becomes immediately available for use without a need to restart the system. There are two types of extensions: (web) client extensions and (application) server extensions. Both may be uploaded to the application server through the web client. A web client extension is automatically downloaded from the application server to a web client machine, where it is then executed. This is done upon the first invocation of the extension at the client side. Such extension is actually a Silverlight web page that is embedded within the *Extensions* page in the web application. It is generally expected to act as a user interface to the built-in or user-added (via server extensions) queries and analyses. On the other hand, server extensions reside on the application server, where they are also executed upon the invocation initiated at the client side. These extensions are functions generally responsible for data operations, analyses, and epidemic models.

V. FEATURED EPIDEMIOLOGICAL ANALYSES

In this section, we present three types of epidemiological analyses that are available in Dr Warehouse: an analysis about the difference between the initial and final diagnosis during work absence, epidemiological forecasts that rely on data mining and epidemiological forecasts that rely on compartmental models. In addition to describing a data set that was used, we offer exemplary results of these analyses. These results primarily illustrate what kind of valuable information may be obtained from Dr Warehouse. Their validity is tightly coupled with the quality of available epidemiological data and the precise tuning of procedure parameters.

A. *Data – Sample, Quality, and Security*

Data set used in the testing of the system during the development was acquired from the HIS of The Institute of Public Health of Vojvodina in Novi Sad, Serbia. The obtained sample (an excerpt is featured in Fig. 7) has approximately 8,500 records about workplace absences that ended in 2009. It contains depersonalized information including: gender (represented by the variable *pol*), age (*starost*), municipality code (*opstina*), absence cause (*uzrok*), start (*prvidan*) and end date (*krajdan*) of absence, disease codes for initial (*pdijag*) and final (*zdijag*) diagnosis, and business activity code (*delatn*) of a person involved.

Gender is represented by numbers 1 and 2 referring to the male or female respectively. Business activity code is represented by a five-digit code indicating sector, division, branch and group of a business activity in accordance with the classification of activities as defined by the corresponding law of the Republic of Serbia. Municipality code is a unique identifier of the municipality in which an absence was recorded.

The cause of the absence is denoted by numbers from 1 to 12 that respectively correspond to: disease, isolation, accompanying sick person, maintenance of pregnancy, tissue and organ donor, injury at workplace, injury outside of workplace, occupational disease, nursing a child under 3 years, nursing a child over 3 years, care of other sick person, and maternity leave. Initial and final diagnosis codes are obtained by reducing the appropriate diagnosis codes defined by the 10th revision of International Classification of Diseases (ICD 10) to four characters. Codebooks of diseases, business activities, causes and municipalities may be gathered from official Internet sites of organizations that are responsible for their maintenance and distribution.

	pol	starost	delatn	opstina	uzrok	pdijag	zdijag	prvidan	krajdan
1	2	52	80220	2690	1	M543	M543	19-Jan-2009	20-Jan-2009
2	2	48	92522	2690	1	M539	M539	12-Jan-2009	23-Jan-2009
3	1	40	51340	1250	1	J42X	J42X	23-Dec-2008	12-Jan-2009
4	1	23	51530	2690	1	M549	M549	20-Jan-2009	21-Jan-2009
5	2	40	85321	1250	1	J42X	J42X	12-Jan-2009	29-Jan-2009
6	1	30	34300	2690	9	Z637	Z637	05-Jan-2009	16-Jan-2009
7	2	30	01110	1250	10	Z637	Z637	26-Jan-2009	26-Jan-2009
8	2	30	01110	1250	10	Z637	Z637	12-Jan-2009	12-Jan-2009
9	2	30	01110	1250	10	Z637	Z637	19-Jan-2009	21-Jan-2009
10	2	26	01110	1250	10	Z637	Z637	30-Jan-2009	30-Jan-2009

Figure 7. Excerpt from a data set used in the generation of predictions.

Credibility of the data depends largely on the credibility of data sources. Therefore, we rely on sources that can guarantee the integrity and validity of provided data.

Within the system, we provided different ways of presenting and making use of existing data. Dr Warehouse is not only conducive to making decisions for medical experts by means of its data rich reports, but also favourable to general population because of its easy to use components. In such way we tended to satisfy some of the secondary data quality criteria such as reliability, credibility, usefulness, added value, ease to use, and accessibility. However, the conception and measurements according to these criteria are established on primary or secondary quality properties which are mainly assessed by subjective methods [34]. The denormalization of the data warehouse schema, which is described in Section III-B, affects some of the properties corresponding to the primary data quality criteria such as efficient use of storage and response time. Although it entails the generation of redundant data, we consider it beneficial because it optimizes read performance and makes the schema more comprehensible to users.

Security in software systems that contain medical data is generally one of the top concerns of software designers. In the current version of Dr Warehouse, the featured data warehouse supports storage only of depersonalized medical records. In such design, there are generally no standard issues with patient privacy and confidentiality of patient health records. Moreover, Dr Warehouse was developed to give wide access to epidemiological data to different categories of users. However, future versions may support distinct user roles that include specific functionalities and access to different portions of contained data. The future access control is expected to follow the role-based access control (RBAC) model [35].

In the more recent development of the system, we extended our solution with a data set acquired from Freebase, an open repository of structured data [36]. We were able to downloaded JavaScript Object Notation (JSON) [37] files that contain structured information on findings consistent with diagnoses. After parsing those files, we obtained records about diagnoses, corresponding symptoms, risk factors, disease causes, treatments, and medical specialties. Those records were stored in a specially designed relational database which is presented in more detail in [30].

B. Changes in Diagnosis

The situation when the final diagnosis in an absence case differs from the initial one may be of special interest to medical experts. Analysing cases in which complications lead to a change in diagnosis or finding often misdiagnosed cases could help experts to devise strategies for the prevention or control of such situations.

For these reasons, we added support for the comparison between the initial and final diagnosis associated with a single workplace absence. For each pair of the initial (*L*) and final (*R*) diagnosis that appears in the available data set, we automatically calculate the number of matching absence cases (*Support*), their share in the whole data set (*Support %*), the percentage of cases with the initial diagnosis *L* that also feature the final diagnosis *R* (*Prob_LR*), and the percentage of cases with the final diagnosis *R* that also feature the initial diagnosis *L* (*Prob-RL*). A user may choose to view only the specified number of the most frequent diagnosis changes, and further select either the initial or final diagnosis to investigate which diagnosis changes are associated with the selected diagnosis.

An excerpt from the results of an analysis about diagnosis change is given in Fig. 8. The majority of changes between the initial and final diagnosis is observed when the supervision of normal pregnancy was substituted with the health supervision and care of other healthy infant and child. This scenario happened in 258 cases, which make up 99.61% of normal pregnancy cases. A similar trend is noticed when the initial diagnosis was officially recorded as the supervision of normal first pregnancy (50 cases). Such diagnosis changes could be considered trivial as they are numerous and do not have much informational value.

On the other hand, the diagnosis changes with lower support value could better demonstrate in what ways a diagnosis may change. In 43 cases with an unspecified abdominal pain as the initial diagnosis, there were six different outcomes (Fig. 9). In the majority of such cases (83.72%), the diagnosis remained the same. The most common change was the one to dyspepsia, which happened in 3 cases (6.98%). The remaining cases further illustrate how a set of unrelated diagnoses, such as cholelithiasis, lumbago, haemorrhoids, and gastroduodenitis, could substitute an imprecise initial diagnosis. As a result, the discovered information might indicate for which common alternative diagnoses a more thorough test could be administered in order to reduce the chance of making a wrong diagnosis.

C. Forecasts based on Data Mining

In Dr Warehouse, we utilize three classification algorithms that are supported by Microsoft SQL Server 2008 R2 Analysis Services: decision trees, naive Bayes, and neural network classifier. These classifiers are trained to estimate the individual share of each disease in all work absences attributed to the 15 most common diseases, as determined by examining the available data set, for a selected year quarter and subpopulation, as defined by age group and gender, in a selected municipality.

Description	Support %	Support	Prob-LR	Prob-RL
Supervision of normal pregnancy, unspecified -> Health su	3.07	258	99.61	83.77
Supervision of normal first pregnancy -> Health supervisio	0.59	50	94.34	16.23
Viral infection, unspecified -> Acute bronchitis, unspecified	0.06	5	1.31	2.7
Acute pharyngitis, unspecified -> Acute bronchitis, unspeci	0.05	4	0.83	2.16
Other and unspecified abdominal pain -> Dyspepsia	0.04	3	6.98	3.37
Angina pectoris, unspecified -> Presence of coronary angic	0.02	2	6.06	18.18

Figure 8. Some of the most frequent changes in diagnosis.

Description	Support %	Support	Prob-LR	Prob-RL
Other and unspecified abdominal pain -> Other and unspecified abdominal p	0.43	36	83.72	100
Other and unspecified abdominal pain -> Dyspepsia	0.04	3	6.98	3.37
Other and unspecified abdominal pain -> Other cholelithiasis	0.01	1	2.33	4.76
Other and unspecified abdominal pain -> Lumbago with sciatica	0.01	1	2.33	0.39
Other and unspecified abdominal pain -> Unspecified haemorrhoids without	0.01	1	2.33	4.17
Other and unspecified abdominal pain -> Gastroduodenitis, unspecified	0.01	1	2.33	1.39

Figure 9. Cases with an unspecified abdominal pain as the initial diagnosis.

In this manner, we may form coarse predictions of the distribution of the most common diseases in a selected subpopulation. In Fig. 10, we give a set of predictions for male employees in the city of Novi Sad who are between 40 and 61 years old. This example demonstrates how a share of some common diseases in that subpopulation may change throughout a year. These estimates are generated using the naive Bayes classification algorithm for Novi Sad.

Predicted shares indicate that essential hypertension, dorsalgia (thoracic region), and lumbago with sciatica may be causes of a larger percentage of absence in quarters 2 and 3 (spring and summer), while their share substantially decreases during quarters 1 and 4 (winter and autumn). On the other hand, viral infection is most responsible for absences in quarter 4 (autumn).

D. Forecasts based on Compartmental Models

Compartmental models are a group of epidemic models that are used to predict dynamics of an epidemic by dividing an analysed population into several compartments (subpopulations) and calculating the changes in compartment sizes given some initial conditions [38][39][40]. These conditions include sizes of compartments (generally expressed as percentages of a whole population) at a single moment in time.

Population compartments correspond to susceptible, infectious, recovered, or some other group of individuals in a population. Furthermore, there are disease-related parameters that are needed in the calculation of changes in compartments sizes: contact rate, recovery rate, birth/death rate, etc. Different models from this family feature different compartments and may be used to obtain forecasts for different diseases. The actual spread of a disease (transition of individuals between different compartments) is modelled by a system of differential equations.

Compartmental models may be used to predict epidemics. However, the modelling of a disease offers additional benefits. Compartmental models are typically used to analyse the state of equilibrium for a particular disease, i.e., the point when there are practically no more changes in the size of featured compartments. Moreover, the information from such models may be applied to control or even prevent outbreaks by using it to define an adequate vaccination strategy.

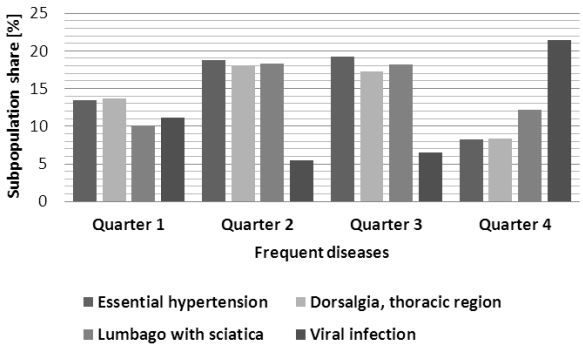

Figure 10. Example of percentage disease shares for male employees in Novi Sad aged between 40 and 61 years, as predicted using a naive Bayes classifier.

As an example of how Dr Warehouse may support standard epidemic models, we implemented five compartmental models based on the information from [41]. With these models, it is possible to describe some common diseases such as influenza, measles, and sexually transmitted diseases. Moreover, they describe with different levels of detail the nature of infection and potential immunity, which is suitable for a broad range of diseases. Both the short-term and long-term forecasts may be made owing to possibility to include basic demographic processes in these models.

The implemented models are added to the system in the form of extensions. For each model, there is a separate client extension. Each client extension is a Silverlight page within the web application. These pages are used for actions such as setting parameters, invoking model execution, and presenting results. In addition to client extensions, there is a single server extension for all implemented models. This extension contains functions that are responsible for numerically solving systems of equations which correspond to compartmental models. Our implementation approximates the solution by using the 4th order Runge-Kutta method for solving a system of ordinary differential equations. It is invoked from a client extension and provides results for the client side.

Various compartmental models may be implemented in a similar manner. The major differences in the implementation would be a change in the set of differential equations that model a disease and addition of new parameters or compartments. By adding the support for several different compartmental models in the form of system extensions, we have demonstrated that Dr Warehouse may be used to predict the rate of spread of any disease for which there is an adequate compartmental model. Given the fact that alterations of the basic models are constantly created and evaluated, the extension mechanism in the system is suitable for the timely testing of a model for a new disease or variant. In the remainder of the subsection, we give an overview of the supported models.

1) Simple SIR Model

The simple SIR (Susceptible/Infected/Recovered) model derives its name from the three compartments that are used to model a population struck by a disease: susceptible (S), infectious (I), and recovered (R). A susceptible individual from the S compartment may become infectious through contact with an infectious individual from the I compartment, while an infectious individual becomes a member of the R compartment after a recovery period and develops a lasting immunity. The rate at which a disease is transmitted from an infectious to a susceptible individual is the contact rate β, while a rate at which an infectious individual recovers is the recovery rate γ. Actual values for the rates β and γ depend on the disease that is being modelled. Some of the diseases that may be described by this model are influenza, measles, and acute hepatitis C. Three ordinary differential equations describe the dynamics:

$$dS / dt = -\beta I S, \quad (1)$$
$$dI / dt = \beta I S - \gamma I, \quad (2)$$
$$dR / dt = \gamma I. \quad (3)$$

In order to analyse the dynamics of a disease within a population, a user of the Dr Warehouse system has to specify the necessary parameters: S, I, R, β and γ. As an aid in this process, the extension may automatically provide the recommended values of β and γ for a disease that has been selected by the user from the list of the disease supported by the extension. The compartment sizes (S, I, and R) may be automatically estimated from the previous disease cases once the user specifies the coefficient describing the share of the available data set with cases in the complete population, and the periods from which the number of infected and recovered individuals should be determined

In Fig. 11, we give an example of a prediction that was generated by a chronological simulation for influenza using our implementation of the simple SIR model. The presented chart demonstrates a typical situation when equilibrium in a population is gradually reached after a peak in the number of infected individuals.

2) Generalized SIR Model

The generalized SIR model is an extension of the simple SIR model, in which demographic processes are acknowledged: the appearance of new susceptible individuals through birth and the disappearance of individuals from all three compartments (S, I, and R) because of death. The inclusion of these processes in the model is especially suitable when analysing the dynamics of a disease over a longer period. In addition to the parameters of the simple SIR model, there is the mortality rate μ, whose value is often used for the birth rate as well. The equations used in this model are slightly more complex from those in the simple SIR model owing to the terms with the μ rate:

$$dS / dt = \mu - \beta I S - \mu S, \quad (4)$$
$$dI / dt = \beta I S - \gamma I - \mu I, \quad (5)$$
$$dR / dt = \gamma I - \mu R. \quad (6)$$

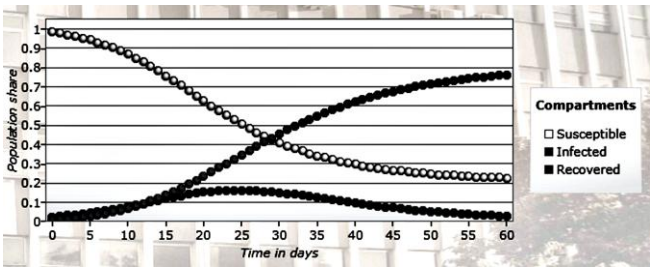

Figure 11. Example of a disease forecast obtained using the simple SIR model.

3) SIS Model

The SIS model could be considered a narrower version of the simple SIR model owing to the absence of the compartment containing recovered individuals. In this model, once an individual stops being infectious, there is no immunity period. As a result, this individual immediately becomes susceptible to the modelled disease. This scenario is typical of many sexually transmitted infections. There are two equations modelling such diseases:

$$dS / dt = \gamma I - \beta I S, \quad (7)$$
$$dI / dt = \beta I S - \gamma I, \quad (8)$$

In Fig. 12, there is an example of a prediction obtained from using the SIS model.

4) SIRS Model

The SIRS model is another variation of the SIR models, in which immunity to the modelled disease is lost after a limited period, i.e., a recovered individual eventually becomes susceptible. Because of this addition, there is a new parameter ω - the rate at which the immunity is waning. A system of equations that take into account these events and basic demographic processes include:

$$dS / dt = \mu + \omega R - \beta I S - \mu S, \quad (9)$$
$$dI / dt = \beta I S - \gamma I - \mu I, \quad (10)$$
$$dR / dt = \gamma I - \omega R - \mu R. \quad (11)$$

5) SEIR Model

The SEIR model is more complex than the aforementioned models as it features additional compartment E, which contains exposed individuals. A susceptible individual may become first exposed and only then infectious. Exposed individuals are infected but not infectious, i.e., they cannot transmit the disease for a certain period that may be expressed using the σ rate. The equations that model such dynamics include:

$$dS / dt = \mu - \beta I S - \mu S, \quad (12)$$
$$dE / dt = \beta I S - \mu E - \sigma E, \quad (13)$$
$$dI / dt = \sigma E - \gamma I - \mu I, \quad (14)$$
$$dR / dt = \gamma I - \mu R. \quad (15)$$

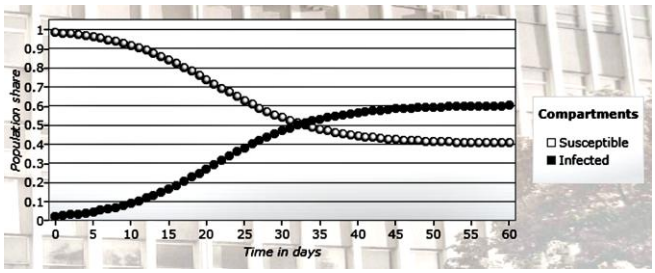

Figure 12. Example of a disease forecast obtained using the SIS model.

VI. Conclusion and Future Work

We provided an extended overview of a software system that may be used in public healthcare and epidemiology for data collection, data mining, analyses, monitoring, and research. The main contribution is the construction of a versatile and comprehensive software solution for epidemiology, as opposed to the similar existing solutions. It should act as a central point for the collection of epidemiological data, their analysis, and presentation tailored to different groups of intended users: public healthcare professionals, epidemiologists and general public. We expect that this system may have an important role in the activities concerned with epidemic control owing to the several already implemented compartmental models for the prediction of disease dynamics. The provided data visualization and reporting controls should offer better understanding of temporal and spatial disease patterns. In addition to a general description of the system's architecture, data source, and client applications, a special attention was given to the implementation of the data warehouse and data analyses. The presented examples of analyses illustrate some of the results that may be obtained through the system using the available data sample. In order to support application of the latest epidemic models and their evaluation in the context of the collected data, we incorporated an extensibility mechanism that allows addition of new functionalities to the system.

We demonstrated through various examples that the developed solution is operational and supports epidemiological monitoring, research, and prediction. These capabilities are illustrated on a data sample from an institution of public healthcare. However, the actual value brought by this solution could only be determined after its prolonged use. During that period, epidemiological data should be collected within the featured data warehouse. Since many epidemiological analyses require data sets covering large time spans, the system should be running for at least several years before it could be systematically evaluated. By creating Dr Warehouse, we have provided a software foundation for epidemiological forecasting through two predictive approaches that have been extensively and successfully used in practice: mathematical modelling of disease dynamics and data mining of epidemiological data. In addition to using available implementations of predictive procedures, epidemiologists may make customizations and even add completely new procedures to the system. This is a more probable research scenario because epidemiological characteristics of many diseases are region-dependent and may change over time, which in turn requires new ways of predicting epidemics and disease dynamics. For these reasons, the actual potential of the system may only be achieved if the system is adopted and readily used by the domain experts.

There are numerous ideas for future work and research on the presented system. We may modify the existing analyses and make them more generic so that they could support a greater number of queries. Moreover, we intend to implement and test several models that are based on cellular

automata. The data warehouse schema may be altered and extended in order to support additional analyses. Since low quality data within such solutions may cause "unnecessary anxiety, investment of time, and expensive engagements with healthcare professionals" [42], we plan a more elaborate assessment of data quality as part of future work. Furthermore, based on these assessments, we intend to improve the ETL process as it is considered a key data quality factor. Given the fact that data quality varies according to user experience, among other factors, we intend to utilize the assessment method for subjective quality properties and the Data Quality Manager (DQM) Prototype presented in [34]. The presented DQM is a prototype for the assessment of data quality within heterogeneous databases that incorporates a data quality assessment framework based on extensions of the Reference Model, Measurement Model, and Assessment Model. Moreover, it takes into account the type of information system when assessing data quality. We consider such property important because of the high data quality requirements associated with the application area.

We may also enforce a strict security policy by introducing user roles and separating the set of functionalities into subsets better suited for various user categories. A new version of the system could be implemented using open (and free) technologies, which could lead to a creation of a completely open version of the system. Due to the prominence of spatio-temporal and epidemiological data in Dr Warehouse, best practices from geographic information systems and constraint databases are topics also worth exploring in the future. Furthermore, significant additions to the system would be the construction of an epidemiological knowledge base, which could be regularly updated or consulted during data analyses, together with the inclusion of a convenient ontology. We have already undertaken some of these activities by integrating information from Freebase into Dr Warehouse. With such enhancements, the semantics may be expressed and the new version of the system could communicate with other systems that follow the idea of the Semantic Web.

ACKNOWLEDGEMENT

The research was supported by Ministry of Education and Science of Republic of Serbia, Grant III-44010. The authors are most grateful to The Institute of Public Health of Vojvodina in Novi Sad for the provided absenteeism data sample and valuable comments.

REFERENCES

[1] V. Ivančević, M. Knežević, M. Simić, I. Luković, and D. Mandić, "Dr Warehouse - An Intelligent Software System for Epidemiological Monitoring, Prediction, and Research," Proceedings of the 5th IARIA International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2013), Jan.-Feb. 2013, pp. 204-210.

[2] W. R. Dowdle, "The Principles of Disease Elimination and Eradication," Bulletin of the World Health Organization, vol. 76, Suppl 2, 1998, pp. 22-25.

[3] T. G. Savel and S. Foldy, "The Role of Public Health Informatics in Enhancing Public Health Surveillance," CDC MMWR – CDC's Vision for Public Health Surveillance in the 21st Century, vol. 61, 2012, pp. 20-24.

[4] A. B. Bernstein and M. H. Sweeney, "Public Health Surveillance Data: Legal, Policy, Ethical, Regulatory, and Practical Issues," CDC MMWR – CDC's Vision for Public Health Surveillance in the 21st Century, vol. 61, 2012, pp. 30-34.

[5] H. Rolka, D. W. Walker, R. English, M. J. Katzoff, G. Scogin, and E. Neuhaus, "Analytical Challenges for Emerging Public Health Surveillance," CDC MMWR – CDC's Vision for Public Health Surveillance in the 21st Century, vol. 61, 2012, pp. 35-39.

[6] K.M. Sullivan, A. Dean, and M.M. Soe, "OpenEpi - a web-based epidemiologic and statistical calculator for public health," Public Health Reports, vol. 124, no. 3, May-June 2009, pp. 471-474.

[7] "Open Source Epidemiologic Statistics for Public Health," http://www.openepi.com/ [Dec. 15, 2013].

[8] J.H. Abramson, "WINPEPI updated: computer programs for epidemiologists, and their teaching potential," Epidemiologic Perspectives & Innovations, vol. 8, no. 1, February 2011, pp. 1-9.

[9] "Epi Info™ - Community Edition," http://epiinfo.codeplex.com/ [Dec. 15, 2013].

[10] "HealthMap," http://www.healthmap.org/ [Dec. 15, 2013].

[11] "Outbreak Watch Social Biosurveillance Network," http://www.outbreakwatch.com/ [Dec. 15, 2013].

[12] "Google Flu Trends," http://www.google.org/flutrends/ [Dec. 15, 2013].

[13] Strategija razvoja informacionog društva u Republici Srbiji do 2020. godine [The Strategy for the Development of Information Society in the Republic of Serbia until the Year 2020], (in Serbian), Službeni glasnik Republike Srbije, vol. 51, 2010.

[14] M. Fimić, M. Radulović, I. Vulić, and S. Atanasijević, "Zdravstveni informacioni sistem Ministarstva odbrane Republike Srbije – generičko rešenje za integraciju institucija" [The Health Information System of the Ministry of Defense of the Republic of Serbia – A Generic Solution for Institution Integration], (in Serbian), Proceedings of YU INFO 2012, pp. 511-516.

[15] G. Johns, "absenteeism," in The Blackwell Encyclopedia of Sociology, G. Ritzer, Ed. Oxford, UK: Blackwell Publishing, 2007, pp. 4-7.

[16] "Absence management, Annual Survey Report 2009 - CIPD," http://www.cipd.co.uk/NR/rdonlyres/45894199-81E7-4FDF-9E16-2C7339A4AAAA/0/4926AbsenceSRWEB.pdf [Dec. 15, 2013].

[17] Đ. Jakovljević and P. Mićović, Zdravstveno stanje i zdravstvene potrebe stanovništva Srbije [Health Status and Health Needs of the Population of Serbia], (in Serbian), http://www.palgo.org/files/leaflet/brosura_zdravstvo.pdf [Dec. 15, 2013].

[18] "Dr Warehouse," http://www.acs.uns.ac.rs/sr/node/237/1429892 [Dec. 15, 2013].

[19] "SQL Server Analysis Services," http://technet.microsoft.com/en-us/sqlserver/cc510300.aspx [Dec. 15, 2013].

[20] "Microsoft SQL Server," http://www.microsoft.com/sqlserver/ [Dec. 15, 2013].

[21] "Microsoft Integration Services," http://msdn.microsoft.com/en-us/library/ms141026%28v=sql.105%29.aspx [Dec. 15, 2013].

[22] "SQL Server Agent," http://technet.microsoft.com/en-us/library/ms189089.aspx [Dec. 15, 2013].

[23] "International Classification of Diseases," http://www.cdc.gov/nchs/icd/icd10cm.htm [Dec. 15, 2013].

[24] J. Mundy, W. Thornthwaite, and R. Kimball, The Microsoft Data Warehouse Toolkit: With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset, 2nd ed., Indianapolis, IN: Wiley Publishing, Inc., 2011.

[25] "Microsoft Silverlight," http://www.microsoft.com/silverlight/ [Dec. 15, 2013].

[26] "Microsoft WCF Services," http://msdn.microsoft.com/en-us/library/dd456779.aspx [Dec. 15, 2013].

[27] "SQL Server Reporting Services," http://msdn.microsoft.com/en-us/data/ff660783.aspx [Dec. 15, 2013].

[28] "Celik API," http://ca.mup.gov.rs/Celik%20api%20Windows%20v1.1.pdf [Dec. 15, 2013].

[29] "PivotViewer," http://www.microsoft.com/silverlight/pivotviewer/ [Dec. 15, 2013].

[30] M. Knežević, V. Ivančević, and I. Luković, "A context-sensitive support system for medical diagnosis discovery based on symptom matching," Proceedings of the 5th KES International Conference on Intelligent Decision Technologies (IDT 2013), vol. 255, Amsterdam: IOS Press, Jun. 2013, pp. 1-10.

[31] "Microsoft Windows Phone," http://www.microsoft.com/windowsphone/ [Dec. 15, 2013].

[32] "TerraService.NET: An Introduction to Web Services," http://research.microsoft.com/apps/pubs/?id=64154 [Dec. 15, 2013].

[33] "Managed Extensibility Framework," http://msdn.microsoft.com/en-us/library/dd460648.aspx [Dec. 15, 2013].

[34] M. d. P. Angeles and F. J. García-Ugalde, "Subjective Assessment of Data Quality considering their Interdependencies and Relevance according to the Type of Information Systems," International Journal On Advances in Software, vol. 5, 2012, pp. 389-400.

[35] R. S. Sandhu, "Role-based Access Control," Advances in Computers, vol. 46, 1998, pp. 237-286.

[36] "Freebase," http://www.freebase.com/ [Dec. 15, 2013].

[37] "JavaScript Object Notation," http://www.json.org/ [Dec. 15, 2013].

[38] W.O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," Proceedings of the Royal Society of London, vol. 115, no. 772, pp. 700-721, August 1927.

[39] R.M. Anderson and R.M. May, "Population biology of infectious diseases: Part I," Nature, vol. 280, no. 5721, August 1979, pp. 361–367.

[40] R.M. May and R.M. Anderson, "Population biology of infectious diseases: Part II," Nature, vol. 280, no. 5722, August 1979, pp. 455–461.

[41] M.J. Keeling and P. Rohani, Modeling Infectious Diseases in Humans and Animals, Princeton, NJ: Princeton University Press, 2007.

[42] R.W. White and E. Horvitz, "Cyberchondria: Studies of the escalation of medical concerns in Web search," ACM Transactions on Information Systems, vol. 27, no. 4, pp. 23:1–23:37, 2009.