Reengineering the eVM Mortality Surveillance Portal
Information Collection, Analysis and Publication

Filipe Alexandre Clemente Cardoso
Under supervision of
Mário Silva/Instituto Superior Técnico/University of Lisbon
Cátia Pinto/Direção Geral da Saúde

November 2016

Abstract

The eVM (electronic Mortality Surveillance) is an informatic application of the SICO (Death certificates information system), from the DGS (Directionate-General of Health), the Portuguese public health authority. It provides, in near-real-time, information on several characteristics of mortality in Portugal, enabling public health stakeholders to respond timely. This dissertation presents the reengineering of eVM. Besides the restructuring of the software architecture and reorganization of the development process, the new eVM adds functionalities for analysis and a revamped interface for visualization of mortality information. A preliminary evaluation of the new interface showed improved user satisfaction and performance tests indicate that eVM response times improved significantly.

Keywords: Mortality Surveillance, Public Health Informatics, Near-real-time surveillance

1. Introduction

Surveillance of the factors with highest negative impact on public health is a powerful tool in the development of policies to increase health status. A robust way to measure the general health status of a population is by studying mortality levels and causes. Death certificates, for instance, can be used for this study. They have existed for a long time. However, they were registered on paper. Although a large amount of data was available, their potential for analysis was wasted, as there was no way to properly treat or analyze that information. With the rise of information technology and data-processing algorithms, it is easier than ever to collect mortality data in digital form. Furthermore, we can also derive useful information that could be used to lower the incidence of diseases and circumstances that kill, through the application of national programs of treatment and prevention. Ultimately, the health status of a population can be raised.

This potential was identified at the DGS, the Portuguese national public health authority. Since 2013, DGS, together with the SPMS (Serviços Partilhados do Ministério da Saúde) has focused on deploying SICO (Sistema de Informação de Certificados de Óbito), the death certificates information system, and eVM (Vigilância Electrónica da Mortalidade) a client application for electronic mortality surveillance. SICO and eVM complement each other to fill these surveillance needs for the analysis of national mortality and its causes. SICO enables doctors to register, in a central database, death certificates as soon as a person is confirmed dead. To guarantee data completeness, the registration of deaths through SICO became mandatory by law since 2014. It therefore covers the total of deaths in Portuguese territory. SICO came to simplify the process of writing death certificates for the doctors and, at the same time, obtain all data in the same format for easier computational analysis and study.

The eVM application was created to analyze these data. It provides information of deaths, organized by different characteristics such as age, region and manner of death, and is displayed in a publicly accessible webpage. The eVM is unique at a worldwide level as it accesses and processes the mortality data in the SICO database in near real time (every 10 minutes). Some of the most advanced systems in the world have a delay of at least a day until the information on the death certificates is available for analysis and research. However, despite its unique characteristics, the initial SICO architecture presented a number of problems which made eVM a less-than-ideal tool for its intended use. Firstly, while the death certificates data stored in SICO and shown in eVM are property of the
DGS, all the information technology, servers and programmers supporting them are headquartered at the SPMS. This created unnecessary bureaucracies that hindered its improvement. DGS had no way to access the SICO data directly and in an “easy-to-read” mode, having instead to rely on SPMS every time they needed to perform a different analysis, not available through the eVM user interface. Also, between e-mail exchanges, authorizations, and validation, a simple change on a graph or correction of small mistakes in data took on average a week.

The application was also very limited in the analysis and visual displays made it only presented a small number of graphics counting the number of deaths in different conditions for every day. While useful, its potential was far wider. Most users of the application accessed its performance as unacceptable, as it took around one minute to load a single webpage. There was a need to not only facilitate the access and treatment of national mortality data, but also to motivate their use by national public health authorities both in Portugal and around the world. This would allow for better management and development of policies at regional and national levels.

This project was proposed to overcome some of the above identified limitations of the previous eVM application, eVM 1.0, and development environment. The objectives were to:

1. Reformulate the system architecture to support separate development and production environments;

2. Design and develop a new version of the eVM, eVM 2.0, with more analytical abilities and upgraded interface;

Previously (on the eVM 1.0), the SPMS was responsible for both the development and the production environments of the eVM. It was an objective of this work to bring the functions of development and update of the application to the DGS. By making this change, the eVM could be developed directly according to this authority's desires and needs for the application, reducing developing time and avoiding problems of miscommunication. The change would also substantially reduce other bureaucratic obstacles, hastening the process of any modifications or additions. The new eVM also needed to have more throughout and complete analytical abilities: in this way to help fulfill not only the DGSs but also other health authorities and public users research needs. A better interface would complement these abilities, with more content and better organization than the one existing previously.

At the end of the project, the planned changes in the development environment were successfully accomplished. The DGS is now the responsible entity for the development of the eVM application, while the SPMS is still responsible of providing its infrastructure.

A better interface, with more content and different ways to present the data from the eVM, was developed. Most of the information is updated every 10 minutes on a request. The result is that information can now be accessed more easily by the public, which may use it for varied reasons as research or journalism, and internal users of the DGS and other health authorities, which can use it for the development of new policies for health improvement.

2. Mortality Surveillance Systems

Public health surveillance is defined as the continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice [1]. According to the WHO [2], health surveillance systems can prove useful by serving as an early warning system for impending public health emergencies, documenting the impact of an intervention, or track progress towards specified goals and monitoring and clarifying the epidemiology of health problems, to allow priorities to be set and to inform public health policy and strategies.

The collection of quality data is an essential step in the definition of any plan, and planning for public health is no exception to this rule. It is of upmost importance that the information used is correct and up to date, or the behavior of diseases may be misunderstood, health programs may not accomplish their goals, and resources may be incorrectly allocated [3]. Without quality public health data, interventions may be misguided and wasteful.

By evaluating public health surveillance systems, stakeholders ensure that public health problems are being monitored efficiently and effectively. There are some key factors for evaluating a surveillance system, which refer to important features [4]. Firstly, the Usefulness of the system is evaluated. A surveillance system is useful if it contributes to the prevention and control of adverse health-related events, including an improved understanding of the public health implications of such events. Secondly, the evaluator studies the different system attributes: Simplicity, Flexibility, Data quality, Acceptability, Sensitivity, Predicted Value Positive, Representativeness, Timeliness and Stability.

Quality of health in general can nowadays be measured by various global health indicators, and each indicator can be measured individually.
WHO’s *World Health Statistics* yearly report has detailed information for the study of these indicators [5]. Mortality, the main focus of this dissertation, is one of such key indicators.

There are various methods for the surveillance of mortality. Most European countries collect individual mortality data in order to annually monitor the impact of chronic diseases, plan and evaluate public health interventions.

In Europe, the EuroMOMO project (European Monitoring of Excess Mortality), implemented in 2008, provides weekly updates of all-cause mortality levels in multiple European countries/regions and provides useful insight into the state of mortality data in Europe. It was created to operationalize real-time surveillance, as sharing international data is sometimes a prolonged process. EuroMOMO processes the mortality data it receives and presents the general state of mortality surveillance in Europe (see http://www.euromomo.eu).

In 2008 the EuroMOMO team published a survey on existing and planned mortality surveillance systems within 32 countries [6], in order to be aware of the existing resources and try to understand whether the existing systems could have been integrated into the Europe-wide mortality monitoring system. The survey concluded at the time that:

- Seven had functioning mortality systems → Belgium, France, Germany, Italy, Portugal, Spain and Switzerland
- Six countries had mortality systems in pilot phase → Denmark, Germany, Hungary, Ireland, the Netherlands and Scotland
- Three were in planning phase → Greece, Sweden and the United Kingdom

Table 1 presents the activation year of the mortality surveillance systems in operation in Europe, along with their earliest year of available data.

<table>
<thead>
<tr>
<th>System</th>
<th>Activation year</th>
<th>Earliest data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>2005</td>
<td>1985</td>
</tr>
<tr>
<td>France-I</td>
<td>2004</td>
<td>Not available</td>
</tr>
<tr>
<td>France-II</td>
<td>2008</td>
<td>Not available</td>
</tr>
<tr>
<td>Germany</td>
<td>2007</td>
<td>2006</td>
</tr>
<tr>
<td>Italy-I</td>
<td>2004</td>
<td>1995</td>
</tr>
<tr>
<td>Italy-II</td>
<td>2005</td>
<td>2003</td>
</tr>
<tr>
<td>Spain</td>
<td>2004</td>
<td>1981</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2006</td>
<td>1969</td>
</tr>
<tr>
<td>Portugal (VDM)</td>
<td>2004</td>
<td>1955</td>
</tr>
<tr>
<td>Portugal (eVM)</td>
<td>2014</td>
<td>2009</td>
</tr>
</tbody>
</table>

Table 1: Activation year and oldest available data of European Mortality Surveillance Systems

Diseases (ICD) is the most widely used in medical care and classification of causes of death [7]. The ICD is a statistical classification of diseases and health related problems. ICD makes it easier to compare the health conditions of the regions across the globe, as it is used by all WHO member states.

The ICD is used in many types of health and vital records, including health records and death certificates. In addition to enabling the storage and retrieval of diagnostic information for clinical, epidemiological and quality purposes, these records also provide the basis for the compilation of national mortality and morbidity statistics by WHO member states, as well as being used in reimbursement systems and in automated decision support in health care [8]. It also excels at being able to promote international comparison of statistics.

Portugal, as one of the member states, also adopted this tool. The DGS is the officially designated institution that represents Portugal in the mortality reference group of the WHO. All SICO death certificates are coded at the DGS. As such, by analyzing this information in death certificates, the eVM can easily scout death certificates by causes of death.

### 3. eVM 2.0

As a result of the state of the art review and evaluation of the first version of eVM, I have identified, together with the DGS personnel, the requirements for the new eVM, both non-functional and functional.

The most important non-functional requirement was that the environment available for development, operation and maintenance of eVM should be reorganised. Instead of giving the SPMS full re-
sponsibility for both the development and operation of the eVM, the development should move to DGSs side. The transfer would give the DGS more control of the software process, enabling a reduction of bureaucracy and implementation time. It would also allow updating and implementing new features in the eVM more in accordance to the needs of the DGS, as directions would start to be directly given from DGS staff to the development team.

The functional requirements refer mostly to fulfilling information needs which eVM 1.0 did not provide, or to present information in a more complete way. Firstly, the new version would have to display the same information as version 1.0 did, but through a completely revamped interface, adding dynamism and interactivity. This was of importance to show eVMs unique characteristic of data completeness with near real-time updates, motivating the use of the application by all stakeholders in mortality data.

Secondly, eVM needed much comprehensive capability to perform analyses by region, to increase the understanding of how certain health events can be related to a certain area. This would allow for more distinctive and objective public health action in certain cases, including analyses of the number of deceased in smaller areas, such as municipalities and health centers, and interactive maps.

Thirdly, the DGS needed a way to monitor the deaths by influenza (and related diseases). No system existed in Portugal to automatically receive data and display information on the incidence of these diseases. The application can automatically identify these cases using the ICD codification in death certificates and make this monitoring.

The DGS also wanted to model confidence intervals for the expected number of deaths in the near future. This is a feature found in many mortality surveillance systems around the world, with high usefulness: it is a very simple way to have an understanding on whether the rate of death is higher than expected, and if action is necessary.

Some information present in eVM 1.0 was in error. This was found in face of data present in detailed DGS internal reports on mortality, which have the highest accuracy of mortality data in Portugal – some specific cases were wrongly processed by the eVM 1.0 algorithms. It was of importance to fix these errors in the new version of the software to increase the accuracy of the data, and as such all generated information was cross-validated with these internal reports.

These above requirements were the ones deemed of highest priority for the new eVM.

Figure 1 presents the structure of environments of eVM 2.0. Unlike in the 1.0 version, the development environment is now located on the DGS side. Programmers at the DGS have a better understanding of the medical context of the eVM, which makes them more suitable for the development of new features to be included in the application. Furthermore, they are in direct contact with the authorities who decide which upgrades or changes the eVM needs. This reduction of bureaucracy and better understanding of needs accelerates the development process and brings more utility to new updates of the application.

After coding and functional testing of new updates, a revised version of the application is submitted to the SPMS. Tests are performed in a local quality assurance environment and, if the application is working properly, it is then passed on to production.

Unlike in the previous version of the eVM, in this version the development environment accesses a synchronized copy of the SICO database, instead of a version with simulated data. This proved to be a considerable improvement in the development process, as the database used for eVM 1.0 testing did not properly mimic the SICO database.
change allowed me to correct data processing errors that would not be detected with the previous database.

Obtaining the data on the SICO database was previously much harder. It required SPMS to extract, process, and send the requested data in .xls format to the DGS every time it was required. This change is considerable and brings flexibility to DGSs work, even beyond the scope of eVM.

3.1. System Architecture

The eVM architecture was completely remade in the new version. Not only new features were added, but they were also implemented in a way to make the most out of the eVM’s potential.

eVM follows a three-tier architecture scheme. This is a client-server software architecture pattern in which the user interface (presentation), functional process (logic), and data access are developed and maintained independently, most often on separate platforms. In eVMs case, the data tier refers to the SICO database, the logic tier to the eVM application, and the presentation tier to the user interface displayed in the user browser.

The logic tier is implemented as instances of R servers. R is a language created specifically for statistical computing which the DGS personnel has experience with, which made it an excellent choice for processing data from SICO. It is composed by two different data processing stages (see Figure 2). The first stage receives the requests for data, which is obtained by querying the SICO database, and saves the resulting data in memory in the format of R objects called data frames. The second stage receives the data frames and converts them into different formats (user interactive widgets) for final visualization.

The presentation tier – the eVM user interface – is a composition of various interactive widgets generated by the logic tier. These widgets can also be generated by an external client request or integrated into a dashboard for end-user access. Each of these components is referred by its own associated URL, which brings two main advantages:

- Standalone eVM components can be embedded into external webpages in a more flexible way. It avoids the necessity of simultaneously generating all objects of the eVM in a response to every client request;
- Each one of the components can be loaded asynchronously in the eVM webpage. This allows the progressive presentation of information to the end users, instead of making them wait for all components to load for visualization.

In this new version of eVM, more data elements are retrieved from SICOs database than in the previous eVM version.

The SICO tables hold a large quantity of data, which will only increase with time. It would not be ideal to obtain all SICO data every time an update to the eVM data is requested. An internal assessment at DGS shows that 99.5% of death certificates are emitted within 2 days of the corresponding death. With this in mind, to find a compromise between data accuracy in eVM and suitable computational time, in its real-time updates, the eVM will only extract the death certificates from SICO from the two previous days, and update the information in the R servers memory with these data. For completeness, the application runs a full update every night.

3.2. User Interface and Functional Features

The eVM webpage has a main body where the information is displayed, with a sidebar which allows the user to navigate to pages with different information (see Figure 3). The following information is available in the eVM:

Mortality Data Today: This is the entry page of the eVM (see Figure 3). It’s a simple page presents some interesting and easy-to-read data, showcasing the near-real-time abilities of SICO.

Daily Mortality Surveillance Graphs: These graphs allow for visualization of daily mortality by different characteristics. The user can see total mortality, mortality by age group, mortality by manner of death and by external manner of death.
The graphs present a number of features which makes information visualization more intuitive, such as the ability to choose which information lines are displayed or to select a time frame to observe. The user can visualize different years of data by selecting the correspondent tab in the page.

### Influenza-like illness mortality surveillance:

When registering a death certificate on the online SICO platform, the doctor types in the conditions or pathological states which ultimately took to death. These descriptions are later used by DGS staff to code the cause of death using ICD-10 [8], at the DGS, and are accessible to eVM. The eVM analyzes these descriptions and codes to make surveillance on mortality by influenza-like illness.

If a death certificate has already been coded, the eVM looks for specific ICD codes to discover whether the death was caused by an influenza-like illness. If it has not been yet coded (a death certificate is usually only coded months after it is registered), the eVM extracts from the written text provided by the doctors patterns to infer the code using `agrep`, an approximate string matching algorithm. This allows the identification of candidate death certificates for further processing (see Table 3 for the codes and patterns (keywords) the eVM uses in this analysis). A second match is then made to exclude unwanted ones.

The terms used for matching and also the terms excluded were defined by DGS experts based on experience by using a sample set of the death certificates of 2014, in which the ICD-10 coding had been complete.

<table>
<thead>
<tr>
<th>ICD codes</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>J09, J10</td>
<td>Pneumonia</td>
</tr>
<tr>
<td>J11, J12</td>
<td>Gripe</td>
</tr>
<tr>
<td>J13, J14</td>
<td>Infecção Respiratória</td>
</tr>
<tr>
<td>J15, J16</td>
<td>Influenza</td>
</tr>
<tr>
<td>J17, J18</td>
<td>Bronquite aguda</td>
</tr>
<tr>
<td>J19, J20</td>
<td>Infecção bronco</td>
</tr>
<tr>
<td>J21, J22</td>
<td>Infecção das vias aéreas</td>
</tr>
<tr>
<td>J998, J960</td>
<td></td>
</tr>
<tr>
<td>J969, J069</td>
<td></td>
</tr>
<tr>
<td>J440</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: ICD Codes and Keywords for eVM 2.0 Influenza-like Illness Analysis

Like in the daily mortality graphs, it is possible to visualize the different years by selecting different tabs. An analysis of the number of deaths by age groups is also available here, for access to more detailed information.

### One week mortality predictions:

A key requirement for the new eVM was adding the prediction of short-term mortality. An R library implementation of the Arima model [9, 10] uses the data since 2009 to predict the mortality in the next seven days, both total and by influenza-like illness.

Figure 4 presents the eVM graph for all-cause mortality prediction. The DGS decided for a prediction limited to seven days, as the prediction error margin significantly increases with time. Information for more days would have an error margin too high to be reliable.
Information Tables: Information tables enable the presentation of the precise quantitative data in the graphs and maps. See in Figure 5 an example of a table generated by eVM. The tables have a number of capabilities which make them interactive. Namely, the user can:

- Sort the tables by any desired column;
- Filter each column by desired values;
- Use a search box to search for a certain value or text in all content of the data table;
- Select the amount of rows displayed for screen page.

There are tables to present the yearly total mortality, mortality by age group, premature mortality and mortality by manner of death. The generated pages show the absolute number of deaths for each characteristic, for every day of the year.

Maps – Regional Analysis: The eVM presents the geographic distribution of Portuguese mortality data in maps, organized both by district and health region. The maps display the monthly number of deaths per 100,000 inhabitants in each represented area of the country. Different years of data are shown in separate tabs. Figure 6 shows one of these sets.

Geographical areas are shown with color densities according to their mortality ratio. The correspondence between ratios and colors in the map is also displayed. When the user passes the cursor over the map, the name and ratio for that area are shown. The user can also select which month to view using a slider. By pressing the play button, the displayed graph automatically switches between different months, giving an animation with the evolution of mortality over time.
4. Evaluation

4.1. User Satisfaction Survey

To study the satisfaction of the target users of the application, I applied a survey, offering to a small set of users a parameterized questionnaire, which is an adaptation from the Questionnaire for User Interaction Satisfaction (QUIS) [11]. The QUIS aims to access the end users subjective satisfaction regarding some aspects of computer interface and usability.

The QUIS consists of two sections, one that measures the overall satisfaction with a system, and another that measures the satisfaction in specific user interface aspects of the system. Evaluation of satisfaction with the interface included measures for five elements: screen components, terminology and system feedback, learning factors, system capabilities and on-screen help. QUIS ratings measure users objective satisfaction through a 9-point scale.

The DGS personnel in general demonstrated high overall satisfaction with the software. The evaluation results for overall satisfaction are summarized in Figure 7, and the results for specific interface factors in Figure 8.

4.2. System Performance

To evaluate the performance of the eVM, I resorted to JMeter, an open-source tool to test functional behavior [12]. It can be used to simulate a heavy load on a server in order to test its strength. I simulated HTTP GET requests to the eVM webpages with a variable number of users, in order to observe how well the systems respond to different stress levels. The main performance indicator taken into consideration was the response time, i.e. the time the server takes to return the eVM webpage and its components.

The results for both versions of the eVM are presented in Figure 9. It is important to mention that eVM 1.0 requests, tested in the production environment, are spread among five servers, while eVM 2.0, running in the quality environment, uses one server only. For eVM 1.0, we can see that the average response time of the system for 5 to 30 users is in the order of 1 minute, an elevated value. In comparison, eVM 2.0 presents much lower response times for a small number of users. Response times scale linearly with the number of users, as expected.

Figure 9 shows also the percentage of successful assertions for each number of simultaneous users. It is possible to observe that it presents solid results until 50 users. When 100 users access simultaneously, it starts presenting some instability. However, near 4 in 5 requests are successful with a single server, which is a very considerable upgrade from eVM 1.0.

5. Conclusions

The new eVM application has been deployed in the SPMSs infrastructure and is providing the services it was designed for. The development of updates and new versions will become, in the future, a responsibility of the DGS. Users of the application can visualize data which they could not beforehand in the eVM 1.0. Furthermore, special attention was given to the user interface of the application for intuitive display of information. As a public application, it is important that the general public can easily find and analyze the information they need.

The information which was chosen to be presented in the application was decided by staff at the DGS. It was selected based on their experience, to be of highest relevance and priority both for public users and policy-making health professionals in Por-
After conducting a survey was positive.

The eVM is unique in the world in the sense that it allows for an in-depth analysis on the mortality of the population of an entire country in almost real-time. It aggregates this information by several useful dimensions of analysis.

Although the eVM was a previously existing application, the work developed in this project introduces new capabilities and much more to offer both to professionals in health and the general public. It allows to easily identify trends in mortality, both general and specified. This ability to display mortality in near-real-time can go past health and extend to other fields. For instance, individuals connect to law can study homicide rates in near-real-time, understanding trends of potential criminalization.

The fact that this system is unique and innovative can also motivate other countries to start using the same “in-real time” model as the eVM + SICO use, since it is a powerful tool. The tool has already been presented publicly by the media and to international health entities such as the WHO [13].

5.1. Future Work

The new eVM is a considerable upgrade from the previous version. There are, however, some ways in which I believe it can be improved in the future.

There is some information which can be added in future versions which might prove useful, as the necessities of the users become more specific. The eVM is prepared to easily accommodate new information. An example would be to separate, within each geographical division, deaths by different characteristics, facilitating study for health-center policy makers.

The integration of analysis and/or algorithms used by other applications or entities is something which is also planned.

Another point to work on in the future is the data visualization interface/eVM webpage. In the long-term future, the DGS could consider re-integrating the data originated from R into a different end-user application, for instance developed by a user experience specialist in developing interactive information visualization applications.

References


Figure 9: Response Times (in ms) and Successful Requests (in %) vs. Different Amount of Simultaneous Users Accessing eVM 2.0

<table>
<thead>
<tr>
<th># simultaneous users</th>
<th>Response times (ms) eVM 1.0</th>
<th>Response times (ms) eVM 2.0</th>
<th>Successful requests (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>60273</td>
<td>6448</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>62388</td>
<td>12853</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>60297</td>
<td>22878</td>
<td>98</td>
</tr>
<tr>
<td>30</td>
<td>60905</td>
<td>25754</td>
<td>98</td>
</tr>
<tr>
<td>50</td>
<td>Not available</td>
<td>57401</td>
<td>94</td>
</tr>
<tr>
<td>100</td>
<td>Not available</td>
<td>130371</td>
<td>78</td>
</tr>
</tbody>
</table>


