

Review

A Benchmarking Analysis of Open-Source Business Intelligence Tools in Healthcare Environments

Andreia Brandão¹, Eliana Pereira¹, Marisa Esteves¹, Filipe Portela^{2,3,*}, Manuel Filipe Santos³, António Abelha¹ and José Machado¹

¹ Algoritmi Research Centre, Department of Informatics, University of Minho, Braga 4710-057, Portugal; a60196@alunos.uminho.pt (A.B.); a58549@alunos.uminho.pt (E.P.); b7284@algoritmi.uminho.pt (M.E.); abelha@di.uminho.pt (A.A.); jmac@di.uminho.pt (J.M.)

² Instituto Politécnico do Porto, Escola Superior de Media Artes e Design, Vila do Conde 4480-771, Portugal

³ Algoritmi Research Center, Department of Information Systems, University of Minho, Campus de Azurém, Guimarães 4800-058, Portugal; mfs@dsi.uminho.pt

* Correspondence: cfp@dsi.uminho.pt; Tel.: +351-967-562-540

Academic Editor: Willy Susilo

Received: 21 July 2016; Accepted: 10 October 2016; Published: 13 October 2016

Abstract: In recent years, a wide range of Business Intelligence (BI) technologies have been applied to different areas in order to support the decision-making process. BI enables the extraction of knowledge from the data stored. The healthcare industry is no exception, and so BI applications have been under investigation across multiple units of different institutions. Thus, in this article, we intend to analyze some open-source/free BI tools on the market and their applicability in the clinical sphere, taking into consideration the general characteristics of the clinical environment. For this purpose, six BI tools were selected, analyzed, and tested in a practical environment. Then, a comparison metric and a ranking were defined for the tested applications in order to choose the one that best applies to the extraction of useful knowledge and clinical data in a healthcare environment. Finally, a pervasive BI platform was developed using a real case in order to prove the tool viability.

Keywords: Business Intelligence; open-source; healthcare; benchmarking

1. Introduction

Increasingly, the healthcare sector is looking for computer applications to support the daily practice of health professionals. Open-source software is particularly desirable in this sector as, besides being free, it has a source code that is fully available to users for viewing, reading, modification, and redistribution, without the restrictions of ownership of the product (unlike free software, which only allows its use without charge).

Open-source software differs fundamentally from an ownership model in terms of the development process and the product licenses. All open-source applications are licensed by an open-source license, which gives the user the right to use the software, access and modify the source-code, and redistribute the software for free. This type of software is very popular due to its many advantages, since it promises to accelerate the diffusion of Information Technologies (IT) solutions in healthcare. Thereby, it can contribute to reduced development costs [1,2].

On the other hand, the choice of a free license software in health and medical informatics is important because it determines the user's rights and can influence the developers' willingness to participate in a project, the quality of the product, and the willingness of users to adopt certain applications [1,2].

In terms of costs, organizations can save on licensing fees and reduce expenditures on specific computer hardware. However, organizations need to welcome and train specialized collaborators

in the adoption of open-source solutions. This type of situation has the hidden costs of highly skilled employees, implementation, maintenance, and a support process, which may lead to adopting proprietary solutions [1,2].

The acquisition of a Business Intelligence (BI) tool and its implementation are quite advantageous for the health organization. Extending their use can have a positive global impact. However, the healthcare environment has some particularities that a BI solution should be prepared to answer. For example, the BI system could lead to resource optimization in various departments; it will improve the clinical condition of the patient through efficient diagnosis and the identification and application of the best practice protocols for treatment, among others [3].

In order to help decision-makers make the best choices, a benchmarking study of BI tools focused on the healthcare environment was performed. After a thorough review of the literature, it was determined which tools were being used in this study. These tools were selected based on their good performance in several areas, such as management, healthcare, or retail [4–6]. Thus, the tools selected for this study were *QlikView* [7], *Palo BI Suite* [8], *Jaspersoft BI* [9], *Tableau Public* [10], *Spago BI* [11], and *Pentaho BI Suite* [12].

The analysis of this type of software emerged during a project that included the development of a BI platform in Maternity Care to visualize clinician and management indicators, as well as integrating Data Mining (DM) predictive models. All the tools were tested using real data provided by the *Centro Hospital do Porto* (CHP) and *Centro Materno Infantil do Norte* (CMIN). With this study, it is intended to select the BI tool that best suits the healthcare sector by using a practical case.

Besides the introduction, this article includes eight sections. The second section provides an approach to the background and related work, in which the concept of Business Intelligence is addressed and an introduction to the case study is presented. The third section is concerned with the application of BI tools in healthcare environments, followed by the requirements considered essential in these tools, in Section 4. Thereafter, follows Section 5, in which are addressed in more detail the tools selected; and Section 6, in which we discuss our approach to the case study. Finally, the last sections correspond to the results, discussion, and conclusion.

2. Background and Related Work

2.1. Business Intelligence

Business Intelligence (BI) is the transformation of information stored in knowledge, enabling us to provide adequate information to a particular user at the appropriate time in order to support the decision-making process in real time. Thus, BI integrates a set of tools and technologies that enables the collection, integration, analysis, and visualization of data.

For the implementation of a BI platform, it is necessary to perform some intermediate steps that are common in the development of this type of software tool, such as the construction of a Data Warehouse (DW). In this case, the Kimball methodology was chosen in order to design as well as develop and deploy the DW [13]. Thus, some of these steps are considered crucial for the successful implementation of a BI system, such as: tasks planning and expected results, defining the architecture that aims to follow the BI system, the selection and installation of the most appropriate BI tool, building the Data Warehouse dimensional model, the Extraction, Transformation and Loading (ETL) process, and, finally, the development of the BI application [13].

Thus, in order to develop a BI application, initially, it is necessary to choose the software that is most appropriate for achieving the desired outcomes. Thereby, it is necessary to undertake an analysis of most of the software available and choose the one type that provides the necessary and desired resources [3,14,15].

2.2. Examples of Application of BI Tools in Healthcare Environments

With the urgency of acquiring medical informatics applications, open-source software is receiving increased attention from the healthcare industry. For example, the open-source project *Care2X*, composed of the hospital information system, practice management, a central data server, and a health exchange protocol, is under development in Europe. *Care2X* was developed in order to overcome integration problems in a multiple incompatible network programs. It is possible to integrate almost any type of service, system, department, process, data, or communication of a hospital. *Care2X* supports the clinical workflow, incorporating diagnosis-related groups (DRG), as well as scheduling and electronic prescribing modules [16].

OpenEHR corresponds to another case, which is sponsored by the *openEHR Foundation*. It promotes the “development of an open, interoperable health computing platform, of which the major component is clinically effective and interoperable electronic health records (EHRs)” [17].

Canada Health Infoway, established by federal and provincial grants, began an open-source initiative in 2005 to develop software that hospitals and developers could use to ensure the secure exchange of medical records of patients between various entities [6]. Thereby, these initiatives suggest that open-source is a viable way of developing applications in healthcare [18,19].

In addition, it is noteworthy that currently there are already a few applications developed based on open-source tools, implemented in healthcare organizations, such as:

- *Turin ASL 3*, which is a system developed in conjunction with the open-source *Spago BI* that allows the assignment of permissions for use according to the different types of users. It provides analytical documents, enables the data visualization, and also allows the use of the Online Analytical Processing (OLAP) technology. This solution is implemented in local healthcare institutions and the Italian National Health Service. *Turin ASL 3* was born in 1995 and is implemented in two hospitals in the city of Turin (*Amadeo di Savoia* and *Maria Vittoria*), Italy [20];
- *St Antonius* is an application developed by *Pentaho BI Suite*, built in order to analyze the waiting times of patients. The main advantages of this application are the improvement of operational efficiency, elimination of costs resulting from the creation of manual reporting, behavioral analysis, and the identification of patterns and risk analysis. This application has been implemented in the St. Antonius Hospital, located in Nieuwegein, Netherlands [21].

2.3. Case Study—Context

The analysis of Business Intelligence open-source software was one of the steps of a process conducted in order to develop a BI platform to support decision-making in maternity care in *Centro Materno Infantil do Norte (CMIN)*. CMIN is one of the constituents of the *Centro Hospitalar do Porto (CHP)*, along with the *Hospital Santo António (HSA)* and *Hospital Joaquim Urbano (HJU)*.

The BI platform is directed to the modules of Gynecology and Obstetrics (GO) and Voluntary Interruption of Pregnancy (VIP), since these modules are lacking decision support systems. This platform aims to visualize the knowledge extracted from the data stored in information systems in CMIN, through their representation in tables, charts, and tables, among others, but also by integrating DM predictive models.

Some of the Voluntary Interruption of Pregnancy Key Performance Indicators (KPIs) that health professionals have interest in are the following:

- Characterization of the patient group by number of pregnancies and by date;
- Characterization of the patient group by number of children and by date;
- Characterization of the patient group by number of previous VIP experiences and by date;
- Characterization of the patient group concerning the revision consultation for date;
- Characterization of the patient group based on contraception early in the process by date.

All the information recorded in the CHP is stored in different systems, taking as an example the Nursing Support System (SAPE), where a portion of the data recorded in the VIP module is stored, and the Electronic Health Record (EHR), where all the patient data are stored.

The Medical Information Integration, Dissemination and Storing Agency (AIDA) guarantees the interoperability of these systems. AIDA is a system of intelligent agents that allows communication between different information systems in CHP [22,23].

3. Applications of Open-Source BI Tools in Healthcare

With the computerization of clinical processes in healthcare organizations, the storage of clinical data in databases has been increasing exponentially. However, a lack of technology to gather, analyze, and distribute the most relevant information makes these organizations rich in data but extremely poor in information [3]. Nonetheless, forward-thinking healthcare organizations are aware that the data and its treatment through the Business Intelligence (BI) technology are essential for an informed and accurate decision-making process, as well as necessary to improve services and ensure the future of these organizations.

Cases of adoption of open-source BI tools by healthcare organizations are scarce. However, the great benefits arising from its implementation in other areas have led to the introduction of BI technology in healthcare environments.

Healthcare organizations typically store how their processes should be performed, especially those that represent complex routine jobs involving multiple people and organizational units. In the context of BI, medical processes are focused on activities and work practices in necessary health services provided (medical and nursing treatments) for the proper operation of a healthcare organization.

Thus, intelligent technologies can be considered facilitators of the management, storage, analysis, and visualization, but they also ensure access to large amounts of data in the context of BI. For such, a wide variety of technologies such as expert systems, online analytical processing, Data Mining (DM) and knowledge extraction are used in the development of a BI system in the healthcare sector. These technologies are required to provide an integrated view of internal and external data (Data Warehouse), which is considered to be the foundation of a BI system. On the other hand, BI systems also include software that provides tools for improving processes in companies, platforms for creating reports, graphical display, dimensional analysis, and DM models [1,2,19].

4. Tool Requirements in Healthcare Environments

The requirements of healthcare organizations for the implementation of a BI system are mainly for providing information to aid the decision-making process at a strategic level, with certain implications at the operational level. The BI technology uses historical and current data in order to visualize them through reports, graphs, and Key Performance Indicators (KPIs), using analytical processing tools.

Thus, and keeping in mind the terms of clinical data, open-source applications must present a set of requirements to meet the needs of healthcare professionals. Some of these features are as follows [4,24–28]:

- *Performance*: assesses whether the tool has good performance in query processing with a high volume of data. In the healthcare sector, it is important that the performance is good, since it is an area where decisions have a major impact on the lives of human beings and sometimes need to be made in a very short period of time.
- *Online Analytical Processing ad hoc queries*: evaluate if the tool allows the user to have the freedom to define queries, which he considers appropriate in a given context. OLAP allows the users to perform ad hoc analyses on the data, considering multiple dimensions and providing the necessary information for an even more efficient decision-making process. This technique allows the analysis of the document's history, and the use of operations such as the roll-up, drill-down, slice and dice, and pivot. In the healthcare sector, the process of analyzing historical clinical data

is very important, since it allows the visualization of the patient and service evolution over time. Thus, this technique is essential in a BI system.

- *Architecture*: assesses whether or not the tool implements a Data Warehouse (DW) and OLAP architectures with high scalability, i.e., capable of processing information evenly, even if the load is increasing.
- *Display of Key Performance Indicators (KPIs)*: assesses whether or not the tool provides visualization of the KPIs of the organization. These indicators can be clinical or management ones.
- *Plug-ins*: assess whether or not the tool allows the development and use of plug-ins that add functionality to it.
- *Interactive visualization of data*: assesses whether or not the tool allows interactivity between the user and dashboards, reports, and graphs. This is a very important characteristic since interactivity is appealing to the user, and also facilitates the understanding of the information demonstrated.
- *Documentation*: assesses the quality of the documentation given by the tool. This feature is very important for the programmer who develops the application since the installation process is sometimes a complicated procedure that requires documentation.
- *Dashboards*: assess whether or not the tool supports the development of dashboards, enabling the integration of graphics, tables, and other analyses such as OLAP and DM.
- *Navigation Features*: assess whether or not the tool enables the creation of reports, using roll-up, drill-down, slice and dice, and pivot operations.
- *Extract, Transform, and Load (ETL)*: steps of the BI process responsible for the extraction, transformation, and loading of data by creating procedures incorporated into the tool.
- *Connection to the database*: it is very important that the BI tool enables a connection to be established to different databases so that it is possible to integrate information from different data sources. There are tools where the only possibility of connection is inherent to data visualization, others in which the connection can be made via the ETL process and via data display, and, lastly, those in which the connection is made only via the ETL process. In a hospital, this is also a key feature because, normally, these organizations have interoperable systems, and it is common to have different databases with clinical information. Thus, in order to facilitate the construction of the data warehouse (DW), we must choose a specific tool for the construction of the DW and, subsequently, a tool of BI is used to create OLAP cubes in order to visualize performance indicators.
- *Integration of dimensional model*: evaluates whether or not the BI tool allows the integration of a DW dimensional model.
- *Open-source*: assesses whether or not the tool presents a development model for which, besides being free, the source code is completely available for users to visualize, modify, and redistribute without restrictions placed by the owner of the product.
- *Export*: assesses whether or not the tool allows export to other formats such as PDF, HTML, spreadsheets, and others.
- *Pervasive*: assesses whether or not the free version of the tool provides a server that allows the development of a web application, which can be opened in a browser, or a mobile application that can be installed on different mobile devices and send alerts, and other pervasive data or healthcare characteristics [27]. If the tool has this feature, it is not necessary to install the application on all the computers in the organization, but only on a server, with all computers connected to the organization network able to access the web application. In a hospital, this feature is very important because, besides the reduction of costs in the application installation process, it also allows a reduction in the time spent, which is very important in the healthcare sector. Specifically, in the *Centro Hospitalar do Porto (CHP)*, this aspect is also very important for the development of BI applications, as once the BI system is integrated into the AIDA it enables interoperability in all the constituents of the organization. Thus, pervasiveness is implemented in this context, so that

the information is distributed to all the users of the organization and is not just focused at the top of the organizational pyramid.

- *Online Help*: assesses whether or not the tool provides the online help resources.
- *Support for mobile devices*: assesses whether or not the tool supports the use of mobile devices, which can be quite useful in a healthcare organization, in that health professionals are then able to access information by other means than a computer.
- *Data Mining*: evaluates whether or not the tool provides the ability to use data to predict chosen outcomes as clinical situations or behavior patterns.
- *Ease of Use*: assesses the ease with which a non-experienced user is able to identify and to find the tools' features, as well as how easy it is to perform them.
- *Attractiveness*: assesses the degree of a tool's interface attractiveness.
- *Customization of the interface*: identifies if the tool allows customization of the interface by the administrator.
- *User Profile*: verifies whether or not the tool allows the administrator to set hierarchies by assigning different permissions to different system users.
- *Real-time*: it is an approach to data analysis that allows users to access the application and the information in real time. This is an important feature in healthcare organizations because it is crucial that health professionals are able to access current data to support the decision-making process. It includes real-time data processing (ETL) and dashboard updates.

5. Business Intelligence Tools—Benchmarking

The study was based only on the analysis of open-source solutions. To achieve this goal, some works were analyzed using scientific databases. During this process the most used open-source solutions were identified. Then the top six solutions (*QlikView*, *Palo*, *Jaspersoft*, *Tableau*, *Spago*, and *Pentaho*) were selected to be explored.

In this study, benchmarking was used in order to compare tools and define evaluation metrics, focusing on the healthcare industry. In the following subsections, an overview of the Business Intelligence (BI) tools tested during this study are presented, based on the experiments made, as well as an extended analysis, i.e., literature review, of each [2,4,20,25,26,29–34]. All the BI tools were selected based on web-based studies, i.e., through a review of the literature it was determined which tools were to be used in this study.

5.1. QlikView

QlikView (QV) is a BI software developed by the Swedish company *QlikTech*. Although this software is a proprietary product, the company provides a full development version of the software for free. Nonetheless, the company offers various licensing options by limiting the use of the software in accordance with the license acquired by the user.

A key feature of this software is that it is fairly simple to extract data from different sources by allowing connections through Open Database Connectivity (ODBC) and Object Linking and Embedding Database (OLEDB), which are communication interfaces between the operating system and the various databases. OLAP also allows several operations by facilitating the user navigation between different dimensions through ad hoc queries. Moreover, it also enables the creation of a friendly and flexible interface with charts, pivot tables, and statistical analyses.

Another important feature of the software is that it does not use a System Management Database (DBMS) as a storage tool. The system connects itself to the base of the Transaction Processing in Real Time (OLTP), used only upon completion of the data loading process. Thus, the data are submitted to ETL processes and compressed in a file with *extension.qvw*, understood by QV. This *QlikView* file is a file that contains all the details required for data analysis, including the data itself, necessary to update the QV script file with new data from the data source; the layout information (folders, lists,

charts, etc.); alerts, bookmarks, documents and reports; and information about access restrictions and the module macros.

Moreover, the distribution of information is facilitated and the analysis can be performed regardless of the original data location or network conditions. Thus, the possibility arises of the end user viewing the application in many ways; once the *file.qvw* is generated it may be viewed on any machine with the data and reports produced. In addition, access can be from a browser, through an application server, according to the safety rules of the customer. However, this feature is only possible in the paid version of QV. In the free version, the final document can only be used on the machine where it was initially developed.

In conclusion, the advantages of QV are the ease of reporting by end users, since the ETL processes are being held, and the fact that reports can be produced with the inclusion of graphics and basic knowledge of aggregate functions (e.g., sums and counts).

5.2. Palo BI Suite

The *Jedox* products are completely open-source, ready to be installed and used. This software aims to minimize the amount of overhead, time, cost, and involvement of computer technologies in the creation and maintenance of BI solutions. Thus, users assume a leadership role through the creation and management of data in *Palo BI Suite*. The *Palo BI Suite in Community* version consists of the following components:

- *Palo OLAP Server* is a multidimensional OLAP in-memory (Molap). The data are stored and then organized into cubes, dimensions, elements, and element attributes. Compared with Rolap (Relational OLAP), the Molap has the potential to be 100 times faster. A simple Palo cube can contain data from multiple data sources, which simplifies the analysis of data from different data sources, for example, the comparison of actual and predicted data collection. Users can import data that rises incrementally, from any data source that Palo get access to, and stores them in the structure defined by them. The way that the tables are listed in the database is not relevant, since this architecture is defined in the software itself.
- *Palo Excel Add-in* allows a connection to Excel via a centralized and multidimensional database such as the *Palo OLAP server*, making it a highly sophisticated BI tool.
- *Palo Worksheet Server* provides a complete reporting and analysis Web-based system that can be maintained by the users. The reports can be created and published to the web, and a flexible and secure content management is provided. Since it is quite similar to Excel, users can easily adapt to this tool by creating reports in the same way as in Excel. However, the resulting calculation sheets are automatically updated when the databases are changed, e.g. when new data is added to a dimension.
- *Palo ETL Server* supports the ETL process. Palo is equipped for the extraction of large quantities of data from a wide range of data sources. The ETL server is not limited to load the *Palo OLAP Server*, but is also adapted to the specific needs of the import and export of data to and from Palo models.

5.3. Jaspersoft BI

The *Jaspersoft BI* is a tool, developed in 2001, in Java and Perl languages. This is an open-source tool available in two versions, the Community version and the Enterprise version. The version of *Jaspersoft BI Community* consists of six individual components:

- *Jaspersoft iReport Designer* is the report designer for *JasperReports* and *JasperReports Server*. It enables the creation of sophisticated layouts with graphics, images, and tables. It also allows access to data via JDBC, TableModels, JavaBeans, XML, Hibernate, CSV, and custom sources. On the other hand, reports can be exported in PDF, RTF, XML, XLS, CSV, HTML, DOCX, or OpenOffice formats.
- *Jaspersoft Studio* has the same functionality as the *Jaspersoft iReport Designer*, differing only in that it is based on Eclipse.

- *JasperReports Library* is an open-source reporting engine, fully written in Java, that can use data from any data source and produce pixel-perfect documents that can be viewed, printed, or exported in a variety of document formats.
- *JasperSoftReports Server* is a reporting server allowing access to reports and analysis that can be incorporated into a web page or a mobile application, and providing real-time information, which is scheduled for a browser, mobile device, printer, or e-mail inbox in a great variety of file formats. It is optimized to share, protect, and centrally manage their *Jaspersoft* reports and analytical views.
- *Jaspersoft OLAP* is a powerful environment for data analysis that can be accessed through an intuitive user interface, designed for the analysis of large volumes of datasets and to perform complex analytical queries. Consisting of an OLAP engine, this tool provides an interactive environment for users to perform slice and dice, pivot, and filter operations and to summarize data in real time through a web-based interface or MS Excel.
- *Jaspersoft ETL* is a simple deployment and running of the ETL process in many external systems. It is used to extract data from the transactional system, to create a data warehouse or data mart, and for creating reports or analysis.

5.4. Tableau Public

Tableau Public is a web-based software and the open-source version of *Tableau Desktop*. This tool allows the creation of interactive visualizations and the option to embed them in a website, as well as publishing them on *Tableau Public Gallery* or sharing them in the *Community Tableau Public*. This open-source version differs from the paid version mainly in that the views cannot be saved locally. Thus, it is only possible to generate a code that can be embedded in any web page. However, this free version also offers several types of representation of data, such as graphs, tables, and maps.

On the other hand, *Tableau Public* also allows the use of various file formats, such as text files, spreadsheets, databases, and comma-separated files. When establishing a connection to a data source, this software identifies the role of each field, including whether the field contains dimensions and facts, i.e., it recognizes the dimensional model of the structure where the data is stored. Thus, when the data are selected to open the *Tableau Public*, the titles, columns, and rows of the structured data are all recognized, and only a simple drag and drop of these values is needed in order to create charts.

5.5. Spago BI

The *Spago BI* tool is a full open-source software, and there is only a single version, i.e., the community edition, a completely free version. It is a tool developed by *Spago World* and supported by an open-source community and consists of several modules:

- *Spago BI server* corresponds to the main module of this software. It offers all the core and analytical capabilities of the application.
- *Spago BI studio* is a development environment based on Eclipse. It allows the user to design and modify all the analysis documents such as reports, OLAP, dashboards, and DM. The interaction between this module and the *Spago BI server* is possible due to the *Spago BI SDK* module.
- *Spago BI Meta* is a module oriented towards the management of metadata and search. The platform manages the metadata, allowing the user to edit and import from external tools such as ETL. This module enriches the knowledge base of metadata from *Spago BI server*, so that they can be easily queried through available tools, such as OLAP.
- *Spago BI SDK* is the specific tool used to integrate services provided by the server. Specifically, it is used by *Spago BI Studio* so that the users can download/upload the analysis of documents to/from the server. This module allows the integration of documents, due to a wide range of services available through a Web service, and the publishing of documents *Spago BI* on an external portal.

- *Spago BI Applications* is a collection of analytical models developed using *Spago BI*. These models are developed taking into consideration the sector of the market and the purpose or the end product, such as analytical specific component.

5.6. Pentaho BI Suite

Pentaho BI Suite software was developed by *Pentaho Corporation* in 2001, in the Java language, being the first BI platform to be released as an open-source alternative in the market. The *Pentaho* offers two types of licenses: the Community Edition (CE), which is the open-source version; and the Enterprise Edition (EE), based on a subscription model.

The *Pentaho BI Suite* project comprises a set of products: BI platform (server), reporting, OLAP analysis, data integration (ETL), dashboards, and Data Mining.

The platform that integrates the *Pentaho* is 100% Java 2 Platform Enterprise Edition (J2EE), thus ensuring its scalability, integration, and portability. Also, it allows the connection to databases of JDBC, IBM DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL, Firebird, and NCR Teradata types. *Pentaho* is structured into different modules, namely:

- *Pentaho BI Platform* provides various services to end users, such as subscriptions scheduling, reporting and integration tools, and incorporated centralized security.
- *Pentaho Reporting* allows the easy development of a report, enabling organizations to access, format, and distribute information. This module contains all the graphical features for the construction of reports as well as ad hoc queries.
- *Pentaho Analysis* provides an OLAP analysis, supporting the users in the decision-making process. The *Pentaho Analysis* facilitates the interactive exploration of information through the intersection of data, in addition to providing a complete integration with other services available in the *Pentaho BI Suite*, via plugins.
- *Pentaho Data Integration* is a powerful tool for ETL process using an innovative, metadata-driven approach.
- *Community Edition Dashboard* provides a graphical environment allowing users access to critical information essential to the understanding and optimization of organizational performance. Complete integration with *Pentaho Reporting* and *Pentaho Analysis* is possible.
- *Weka Pentaho Data Mining* enables a predictive analysis, providing information about hidden patterns and relationships between data, as well as performance indicators. This module provides a graphical interface for pre-processing of data, classification, regression, clustering, rules association, and visualization.

6. Business Intelligence Tools—Assessment Process

In this section, a relationship between the features referenced as the requirements of a Business Intelligence (BI) platform in a healthcare environment (Section 4) with the BI tools previously chosen and analyzed based on web-based studies, i.e., through a review of the literature, is presented (Section 5).

The assessment process was made in a group setting (10 people, five nurses and five IT specialists). A nurse was responsible for creating a team and providing us with the final observations. They performed two rounds of assessment and then provided us with the final assessment of each feature (group result). In order to maintain the anonymity of the results, we did not interfere in the assessment process. At the end, a meeting was held with the responsible nurse in order to understand the group's opinions.

Thus, in Table 1, a comparison of the tools and the requirements is given. For the comparative analysis of the selected tools, a classification criterion based on the following scale has been chosen:

Table 1. Comparative table of the BI open-source tools selected by features and group. (0 means unknown; 1 means absent; 2 means insufficient; 3 means sufficient; 4 means good and 5 means excellent.)

Features	BI Open-Source Tools						Group
	<i>Jaspersoft BI</i>	<i>Palo BI Suite</i>	<i>Pentaho BI Suite</i>	<i>QlikView</i>	<i>SpagoBI</i>	<i>Tableau Public</i>	
Performance	4	3	4	3	4	4	D
OLAP Ad hoc Queries	1	5	5	3	5	4	B
Architecture	4	4	5	4	5	4	D
Display of KPIs	1	1	5	4	4	4	A
Plug-ins	3	0	5	0	0	3	D
Interactive Visualization of Data	5	4	5	5	4	4	C
Documentation	4	4	2	2	2	3	F
Dashboards	1	1	4	4	5	4	B
Navigation Features	5	4	4	1	2	4	C
ETL	4	5	5	3	4	1	E
Connection to the Database	5	4	4	5	5	3	A
Integration of Dimensional Model	1	1	1	2	4	1	E
Open-source	5	5	5	1	5	5	D
Export	5	2	5	2	5	4	C
Pervasive	5	5	5	1	5	4	A
Online Help	4	2	3	4	3	4	F
Support for Mobile Devices	4	1	0	5	5	3	C
Data Mining	1	1	3	2	4	1	B
Ease of Use	4	4	4	4	4	5	F
Attractiveness	4	3	4	5	5	4	C
Customization of the Interface	4	0	5	5	5	5	F
User Profile	5	4	5	1	4	0	D
Real-time	5	4	5	1	5	1	A

Thus, for each of the characteristics, each tool was rated. This rating reveals the tool satisfaction and compliance level in accordance with the characteristic analyzed. The score was based on the experimentation and critical view of the users.

On the other hand, after analyzing each of the characteristics for each tool, the requirements were grouped by a degree of similarity among them. Six groups were considered, where all the requirements were distributed, and a percentage was attributed according to their importance in a BI platform in a healthcare context.

It must be noted that all the scores attributed and the weightings assigned to each group were defined based on the critical opinion of a multidisciplinary team of health professionals, i.e., physicians and nurses from *Centro Materno Infantil do Norte (CMIN)*, and IT professionals from *Centro Hospitalar do Porto (CHP)*, including the authors of this paper. Thus, the first part of the evaluation was based on collecting opinions from several professionals involved directly or indirectly with the use of BI tools and, then, the critical analysis and assessment of all the feedback collected from the professionals by the authors involved in this study.

Thus, initially, the group of indispensable features (*must have*), whose importance cannot be measured, i.e., they are strictly necessary, was identified.

Then, a group was defined in which the requirements were targeted for the benefit of the administrator responsible for the construction of the BI platform. This group was assigned a percentage of 5%, because they do not have great importance in the health organization. It may have some value only to the programmer who manages the tool.

Another set group was the group whose characteristics are advantageous for the end-user of the BI platform. This group was assigned a percentage of 25%. It has a higher percentage assigned

because it is important that the end-user is satisfied. Otherwise, this may cause the failure of the platform’s implementation.

On the other hand, the group with the highest percentage associated (30%) corresponds to the technologies that the tool incorporates, because the main focus of these tools are study and analysis functionalities. With the non-existence of features such as OLAP technologies or dashboards, a BI platform loses all its interest to the IT professionals and health professionals.

To some other important features, a percentage of 25% was allocated.

Finally, the last group created was associated with the processing of data and has a percentage of 15% associated. In this group, all the features included would be advantageous to incorporate in the tool; however, they do not have much influence on the success of the construction of the BI platform. For example, in the ETL process, if the tool could incorporate ETL procedures, it would have many benefits. However, the ETL may be performed using other tools.

The groups, associated features, and respective percentages are all shown in Table 2.

Table 2. The groups, features associated, and respective percentages.

Group		Characteristics	Percentage
A	Must Have	Pervasive Real-time Display of Key Performance Indicators Connection to the Database	-
B	Technologies	OLAP Ad hoc Queries Dashboards Data Mining	30%
C	End-User	Interactive Visualization of Data Navigation Features Export Support for Mobile Devices Attractiveness	25%
D	Other Important	Performance Architecture Plug-ins Open-source User Profile	25%
E	Data Processing	ETL Integration of Dimensional Model	15%
F	Administrator	Documentation Online Help Ease of Use Customization of the Interface	5%

7. Business Intelligence Tools—Assessment Results

After the construction of Tables 1 and 2, a final grade for each of the Business Intelligence (BI) tools was obtained. Thus, for each assigned classification by each requirement, the score was multiplied by the respective group percentages.

For example, in the case of the *Pentaho BI Suite*, the ratings for each group are as follows:

- Must Have: $5 + 5 + 5 + 4 = 19$;
- Administrator: $(2 + 3 + 4 + 5) \times 0.05 = 0.7$;
- End-user: $(5 + 4 + 5 + 0 + 4) \times 0.25 = 4.5$;
- Technologies: $(5 + 4 + 3) \times 0.3 = 3.6$;
- Other Important: $(4 + 5 + 5 + 5 + 5) \times 0.25 = 6$;
- Data Processing: $(5 + 1) \times 0.15 = 0.9$.

Therefore, it can be verified that the final evaluation of the *Pentaho BI Suite* is $19 + 0.7 + 4.5 + 3.6 + 6 + 0.9 = 34.7$. This procedure was repeated for each of the remaining tools.

In Table 3, the respective final ratings are shown.

Table 3. Final rating assigned to each of the BI tools evaluated.

BI Tool	Final Grade
<i>Spago BI</i>	34.85
<i>Pentaho BI Suite</i>	34.70
<i>Jaspersoft BI</i>	29.45
<i>Palo BI Suite</i>	25.00
<i>Tableau Public</i>	24.60
<i>QlikView</i>	21.95

Analyzing the values reported in Table 3, it could be concluded that, at first glance, the most appropriate tool would be *Spago BI*. Mathematically speaking, *Spago BI* is the better tool; however, the difference between it and *Pentaho Suite* is insignificant. Taking this point into consideration along with the fact that *Spago BI* installation is quite a complex process, there is little supporting documentation, and, in addition, this software takes up way more RAM, *Pentaho BI Suite* was the tool chosen to implement the case study.

To justify this choice, a final quick comparative analysis between *Pentaho BI Suite* and each of the other tools was made. Thus, comparing *Pentaho BI Suite* with *Spago BI*, it appears that both tools are very good, and have very close scores on different requirements. However, the main distinguishing feature between these two tools is that the *Spago BI* installation is more difficult.

Considering the *Jaspersoft BI*, it appears that the *Pentaho BI Suite* provides more capabilities and better interactivity in terms of dashboards. Moreover, *Pentaho BI Suite* has many more developed plug-ins for use, such as *CDE*, *Saiku*, and *OpenI*, all available in the marketplace, unlike *Jaspersoft BI*, which presents a very limited number of plug-ins to date.

Regarding *Palo BI Suite*, it does not allow the display of KPIs and, moreover, does not allow the integration of Data Mining (DM) processes, unlike *Pentaho BI Suite*.

On the other hand, the biggest disadvantage of *QlikView* regarding *Pentaho BI Suite* is the failure to allow access to dashboards through browsers. This is an important feature since it is essential that the application can be accessed anywhere through a web page and also by multiple users simultaneously. Besides that, *QlikView* is not a completely open-source software since this feature is only available in the server version. Analyzing other features, *QlikView* and *Pentaho BI Suite* are more or less very similar.

Finally, *Tableau Public* is considered the tool that has the lowest number of desired characteristics, and does not support some major features, including the connection to an Oracle database, which is where all the data used in this BI project are stored.

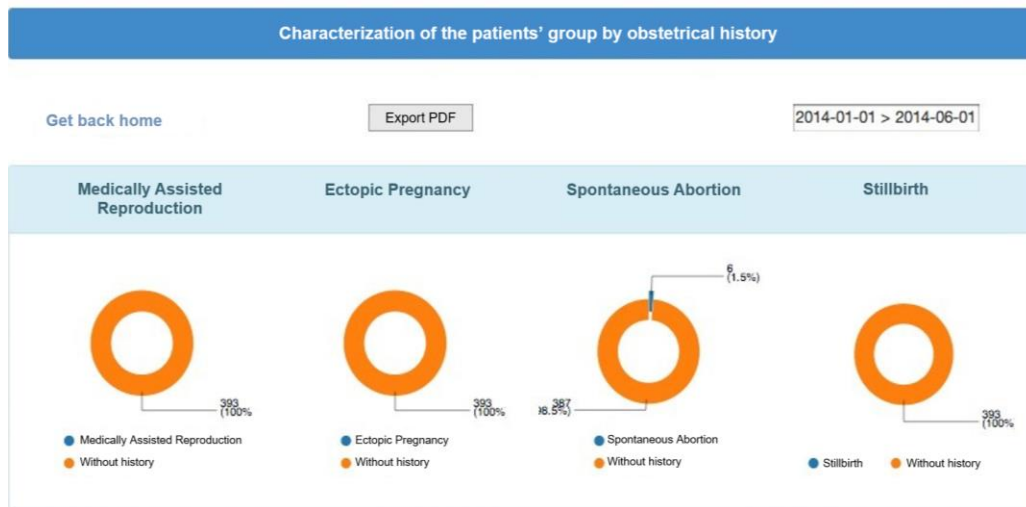
8. Case Study Application

All the requirements mentioned in Section 4 were selected based on the healthcare sector. So, considering these requirements, *Pentaho BI Suite* was the chosen tool to develop a BI application in *Centro Materno Infantil do Norte (CMIN)*, and to prove the efficiency of Business Intelligence (BI) tools in healthcare organizations.

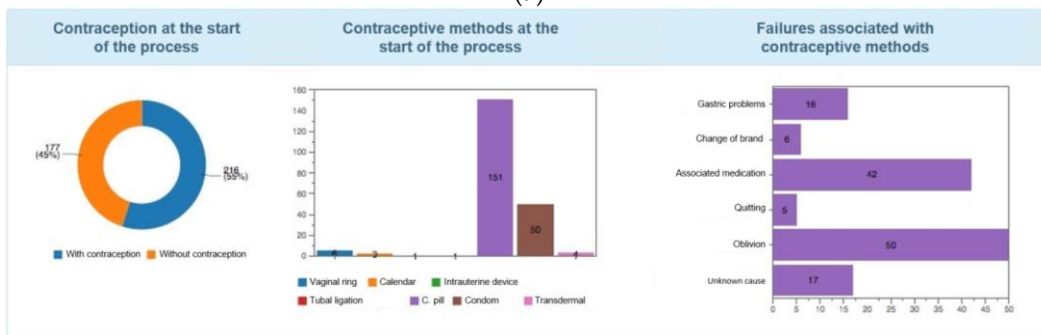
Pentaho BI Suite meets all the necessary requirements and has a larger number of advantages than the other BI tools. Moreover, it is a tool that has a nice interface, a variety of different forms of representation, and a long list of plug-ins that allows you to customize the application, such as *OpenI* and *Saiku*, used in this particular application.

In Figure 1, some pages of the BI application developed for CMIN are shown. Some of the indicators developed were the characterization of the patient groups by contraception used at the beginning and at the end of the Voluntary Interruption of Pregnancy (VIP) process, by location, and also by weeks of gestation. It is possible to observe a dashboard with a set of indicators presented through

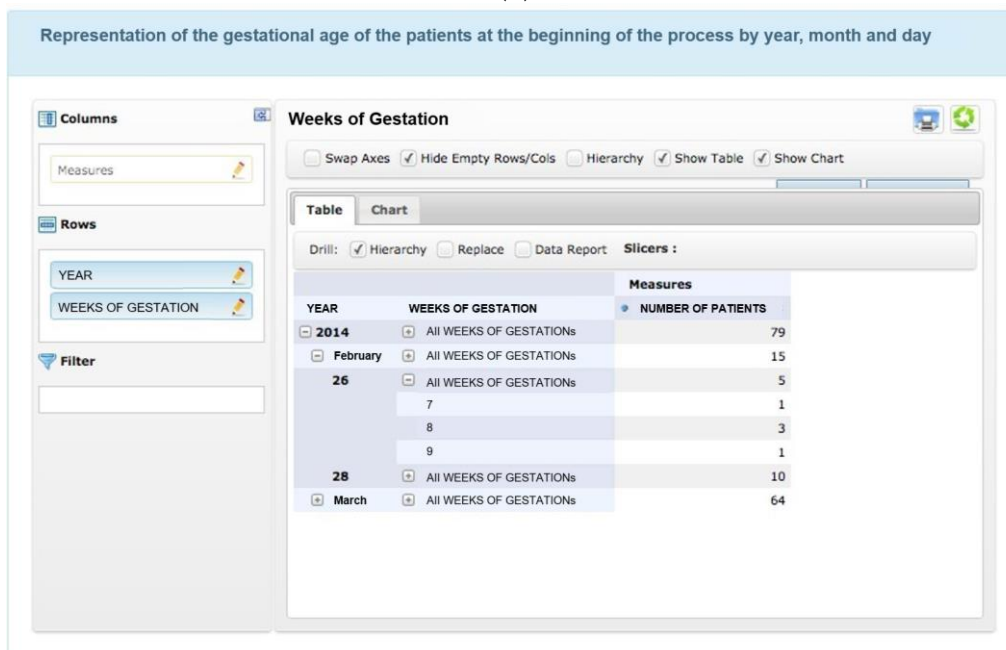
charts and OLAP technology (table). All the indicators are updated automatically and can be consulted anywhere and anytime, grouping the information by date (hour, day, month, and year).



(a)

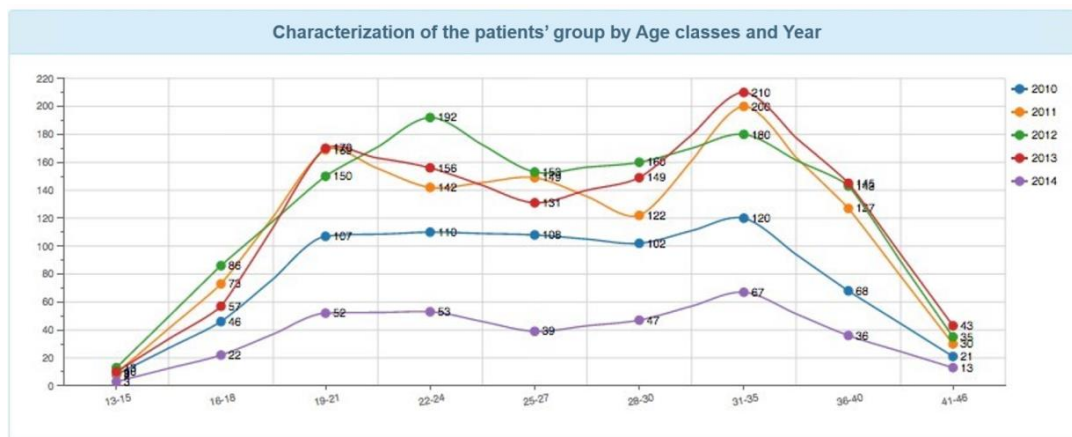


(b)



(c)

Figure 1. Cont.



(d)

Figure 1. CMIN’s BI application using *Pentaho BI Suite* software: (a) presents the obstetrical history of the patients; (b) shows the association between the women and contraception usage ; (c) presents the distribution of the gestational age of the patients at the begin of the abortion process; (d) makes an overview of the patients by age classes and years.

In this figure it is possible to see a characterization of patients by their obstetrical history (Medically Assisted Reproduction, Ectopic Pregnancy, Spontaneous Abortion, and Stillbirth). For example, in the case of spontaneous abortion only 1.5% of the patients are recurrent. In the next chart it is possible to observe the percentage of women who used contraception (45%) before the pregnancy and consequently had an abortion. In the third group of charts is the number of gestation weeks before the start of the abortion process. This analysis can be done by year, month, day, and number of weeks. For example, three women were admitted at eight weeks’ gestation on the 28th of February. Finally, the nurses can see an overview of the patients admitted by age classes in each year. In 2013, 146 of the admitted women were aged between 36 and 40 years.

Nowadays, a totally functional Pervasive Business Intelligence Platform developed using *Pentaho BI Suite* is implemented and used in CMIN by its health professionals in order to assist their decision-making process in maternity care services [35,36].

9. Conclusions

This paper is intended to facilitate the research work of developers who plan to deploy an open-source Business Intelligence (BI) platform in a healthcare environment and do not know which is the most appropriate tool to use. Healthcare institutions requiring similar features to the one presented in this paper may select the same software used in the case study based on the analysis made and benefits mentioned throughout this manuscript. It is also worth mentioning that this type of choice is crucial, since the applications developed in healthcare environments need to be more efficient and effective, as malfunction may have direct or indirect repercussions in terms of patients’ health.

Thus, with the analysis of all the BI tools discussed throughout this paper, it can be concluded that *Spago BI* is the most comprehensive open-source software; its Community version contains more features than most Enterprise versions of software analyzed. It can be concluded that *Pentaho BI Suite* is also a consistent and complete tool, similar to *Spago BI*, only differing in the fact that it does not incorporate a dimensional model structure. Therefore, these tools can be considered the two most suitable to healthcare environments, taking into account the characteristics required by this type of institution. Nonetheless, *Pentaho BI Suite* was chosen to develop the BI application in the *Centro Hospitalar do Porto* since its installation process is simpler. Additionally, the *Pentaho BI Suite* software is well developed regarding the integration of plug-ins, since it allows the customization of

the software according to the needs of the BI application. Now, we are designing a TAM evaluation in order to assess the quality of the solution developed.

However, it should be noted that the tools presented and analyzed throughout this paper are constantly evolving, and the features that they do not have now may be present in the near future.

Acknowledgments: This work has been supported by Compete POCI-01-0145-FEDER-007043 and FCT - Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/00319/2013.

Author Contributions: Andreia Brandão, Eliana Pereira and Filipe Portela performed the main study and wrote the paper. Andreia Brandão and Eliana Pereira developed the main platform. Marisa Esteves and Filipe Portela reviewed the paper and made the final study assessment. Manuel Filipe Santos, António Abelha, and José Machado were involved in the study and development process. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Janamanchi, B.; Katsamakos, E.; Raghupathi, W.; Gao, W. The State and Profile of Open Source Software Projects in health and medical informatics. *Int. J. Med. Inform.* **2009**, *78*, 457–472. [CrossRef] [PubMed]
2. Waring, T.; Maddocks, P. Open Source Software implementation in the UK public sector: Evidence from the field and implications for the future. *Int. J. Inf. Manag.* **2005**, *25*, 411–428. [CrossRef]
3. Ishola, M.D. Healthcare Business Intelligence: The Case of University's Health Center. Master's Thesis, Universiti Utara Malaysia, Kedah, Malaysia, 30 July 2011.
4. Bernardino, J. Open Source Business Intelligence Platforms for Engineering Education. In Proceedings of the 1st World Engineering Education Flash Week, Lisbon, Portugal, 27–30 September 2011.
5. Brandão, A.; Pereira, E.; Portela, F.; Santos, M.F.; Abelha, A.; Machado, J. Predicting risk pregnancy in maternity care using data mining. In Proceedings of the 7th International Conference on Agents and Artificial Intelligence, Lisbon, Portugal, 10–12 January 2015.
6. Canada Health Infoway. Available online: <https://www.infoway-inforoute.ca/en/> (accessed on 10 July 2016).
7. QlikView. Available online: <http://www.qlik.com/> (accessed on 12 October 2016).
8. Jedox. Available online: <http://www.jedox.com/> (accessed on 12 October 2016).
9. Jaspersoft. Available online: <https://www.jaspersoft.com/> (accessed on 12 October 2016).
10. Tableau Public. Available online: <http://www.tableausoftware.com/public/> (accessed on 12 October 2016).
11. Spagobi. Available online: <http://www.spagobi.org/> (accessed on 12 October 2016).
12. Pentaho. Available online: <http://www.pentaho.com/> (accessed on 12 October 2016).
13. Completo, J.; Cruz, R.S.; Coheur, L.; Delgado, M. Design and Implementation of a Data Warehouse for Benchmarking in Clinical Rehabilitation. *Procedia Technol.* **2012**, *5*, 885–894. [CrossRef]
14. Côrte-real, N.E.D.P. Avaliação da Maturidade da Business Intelligence nas Organizações. Ph.D. Thesis, Universidade Nova, Lisbon, Portugal, 18 February 2011. (In Portuguese).
15. Tereso, M.; Bernardino, J. Open source business intelligence tools for SMEs. In Proceedings of the 6th Iberian Conference on Information Systems and Technologies (CISTI 2011), Chaves, Portugal, 15–18 June 2011.
16. Care2X: The Open Source Hospital Information System. Available online: <http://www.care2x.org/> (accessed on 17 May 2016).
17. OpenEHR: An Open Domain-Driven Platform for Developing Flexible e-Health Systems. Available online: <http://www.openehr.org/> (accessed on 3 May 2016).
18. Welch, W.P.; Bazarko, D.; Ritten, K.; Burgess, Y.; Harmon, R.; Sandy, L.G. Electronic Health Records in Four Community Physician Practices: Impact on Quality and Cost of Care. *J. Am. Med. Inform. Assoc.* **2007**, *14*, 320–328. [CrossRef] [PubMed]
19. Open Source Software: A Primer for Health Care Leaders. Available online: <http://www.chcf.org/publications/2006/03/open-source-software-a-primer-for-health-care-leaders> (accessed on 11 October 2016).
20. SpagoBI for Areas. Available online: <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/ApplicationsAreas> (accessed on 18 May 2016).

21. Pentaho Healthcare Solutions Presentations. Available online: <http://pt.slideshare.net/Pentaho/pentaho-healthcare-solutions-presentation> (accessed on 10 July 2016).
22. Machado, J.; Alves, V.; Abelha, A.; Neves, J.M. Ambient intelligence via multiagent systems in the medical area. *Int. J. Eng. Intell. Syst. Electr. Eng. Commun.* **2007**, *15*, 151–158.
23. Peixoto, H.; Santos, M.; Abelha, A.; Machado, J. Intelligence in Interoperability with AIDA. In *Foundations of Intelligent Systems*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 264–273.
24. Filipe, P.; Manuel, F.S.; Marta, V.-B. A Pervasive Approach to a Real-Time Intelligent Decision Support System in Intensive Medicine. In *Knowledge Discovery, Knowledge Engineering and Knowledge Management*; Fred, A., Dietz, J.L.G., Liu, K., Filipe, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2013.
25. Lapa, J.; Bernardino, J.; Figueiredo, A. A comparative analysis of open source business intelligence platforms. In *Proceedings of the International Conference on Information Systems and Design of Communication*, Lisbon, Portugal, 16–17 May 2014; pp. 86–92.
26. Sá, T.; Mateus, A.; Iwanaga, M.; Ribeiro, A.; Escovedo, T.; Melo, R.N. Uma Análise Comparativa entre as Ferramentas OLAP como Apoio a Soluções de BI nas Empresas. Available online: http://www.infobrasil.inf.br/userfiles/15-S1-1-97127-ma%20An%C3%A1lise%20Comparativa____.pdf (accessed on 11 October 2016). (In Portuguese)
27. Orwat, C.; Graefe, A.; Faulwasser, T. Towards Pervasive Computing in Healthcare—A Literature Review. *BMC Med. Inform. Decis. Mak.* **2008**, *8*. [[CrossRef](#)] [[PubMed](#)]
28. Portela, F.; Santos, M.F.; Silva, Á.; Machado, J.; Abelha, A. Enabling a Pervasive Approach for Intelligent Decision Support in Critical Healthcare. In *Proceedings of International Conference (CENTERIS 2011)*, Vilamoura, Portugal, 5–7 October 2011.
29. Jedox Whitepaper. Available online: http://www.altavia.com/avcfiles/jedox_whitepaper.pdf (accessed on 9 July 2016).
30. Arriaga, A.; Marques, P. Web OLAP, Exploração de Sistemas OLAP via Web. Available online: https://www.researchgate.net/profile/Afonso_Arriaga/publication/264845178_Web_OLAP_Exploracao_de_Sistemas_OLAP_via_Web/links/56661e0d08ae192bbf92798a.pdf?origin=publication_detail (accessed on 12 October 2016).
31. Technologies, N. MicroStrategy 9 vs. QlikTech—A comparison white paper. Available online: <https://www.ndm.net/bi/pdf/MicroStrategy-9-vs-QlikTech-11.pdf> (accessed on 12 October 2016).
32. QlikView-vs-OLAP. Available online: <http://www.covaligroup.com/blog/bid/47883/QlikView-vs-OLAP> (accessed on 23 May 2016).
33. Rueter, A.M.; Solutions, S. Tableau for the Enterprise: An Overview for IT. Available online: <http://www.tableau.com/learn/whitepapers/tableau-enterprise> (accessed on 12 October 2016).
34. Qlik View. Available online: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Qlik+View#0> (accessed on 15 May 2016).
35. Pereira, E.; Brandão, A.; Portela, C.F.; Santos, M.F.; Machado, J.; Abelha, A. Business Intelligence in Maternity Care. In *Proceedings of the 18th International Database Engineering & Applications Symposium*, Porto, Portugal, 7–9 July 2014; pp. 352–355.
36. Brandão, A.; Pereira, E.; Portela, F.; Santos, M.; Abelha, A.; Machado, J. Real-time Business Intelligence Platform to Maternity Care. In *Proceedings of 2014 IEEE Conference on Biomedical Engineering and Sciences*, Kuala Lumpur, Malaysia, 8–10 December 2014.

