Application of Business Intelligence in studies management of Hazard, Vulnerability and Risk in Cuba

To cite this article: C Milanés-Batista et al 2020 IOP Conf. Ser.: Mater. Sci. Eng. 844 012033

View the article online for updates and enhancements.
Application of Business Intelligence in studies management of Hazard, Vulnerability and Risk in Cuba

C Milanés-Batista¹ᵃ, H Tamayo-Yero², D de Oliveira³, J R Nuñez-Alvarez⁴ᵃ

¹Civil and Environmental Department. Universidad de la Costa, Colombia, ORCID: 0000-0003-2560-8859.
²Computer Engineering Department, Universidad de Oriente, Cuba.
³Departamento de Electricidad, Universidade Federal de Santa Catarina, Brasil.
⁴Department of Energy, Universidad de la Costa, Colombia, ORCID: 0000-0002-6607-7305.

ᵃCorresponding author: jnunez22@cuc.edu.co; cmilanes1@cuc.edu.co

Abstract. In Cuba, the state invests considerable resources in the establishment of preparation plans to mitigate and to minimize the negative impacts of natural threats. As a sample of it, since the year 2010, the country carried out the studies of hazard, vulnerability, and risk (HVR), however, the form in which the results of these studies are analyzed present serious limitations for the excessive quantity of data that are dispersed and not very understandable for users belonging to the Centers of Risks’ Management make decisions in an agile way. The present work, exposes the management’s pattern of the studies of Hazard, Vulnerability and Risk in Cuba alongside the occurrence of hydrometeorological extreme events in vulnerable territories; and the structuring process of the computer proposal that modifies and computerizes the current analysis procedure of these studies in the country. As contributions of the research and by means of the use of techniques and tools of Business Intelligence, is designed and implemented, for the case of Santiago de Cuba province, the data warehouse that centralizes the results of the studies for hydrometeorological extreme events in 2011 and 2016. A decision support system (DSS-HVR) is built, and integrated at the developed data warehouse in this work, it allows the analysis of the studies in a holistic manner and from several perspectives, to obtain outstanding and better-represented information through interactive and dynamic reports, both tabulate and graph representation, all as a support for a fast and effective decision making.

Keywords. Business Intelligence, decision support system, data warehouse, hazard, vulnerability and risk.

1. Introduction

The current increase in the frequency and the destructive force of extreme hydrometeorological events in Cuba, and its main manifestations, such as: coastal flooding due to heavy rains, sea penetrations, and the
effects of high winds [1] have contributed to the fact that in Cuba perfects the political, economic, social, scientific-technological and environmental approach to risk management through the implementation of a set of strategies, which are the result of multidisciplinary projects and research, aimed at preventing, reducing and controlling risk factors of disasters, and thus reduce vulnerabilities to the negative impact of these hazards [2].

The Santiago de Cuba province is located south of the Island of Cuba, in direct contact with the Caribbean Sea [3], a position that increases its fragility in the face of extreme hydrometeorological events, which already occur with a high frequency and greater destructive power [4].

Coupled with this situation, the progressive economic development and population growth that exists in towns near the Santiago coast stand out, which is multiplying the risk of affecting economic and social infrastructure and the loss of the most precious of society, “life of man”, so it is necessary to establish more efficient preparedness and risk management plans, in order to mitigate and minimize the conditions of vulnerability caused by the main damages.

In 2005, the first Directive No. 1 of the Vice President of the National Defense Council for the organization, planning and preparation of the country in the event of disasters [5] emerges, which is perfected in 2010 [6]. Since then, the Ministry of Science, Technology and Environment (CITMA) was responsible for conducting studies of hazard, vulnerability and risk (HVR) in different types of situations. At the same time, CITMA delegated to the Environmental Agency (AMA) the responsibility for compliance with this directive, with the help of the participation of the country's scientific potential, [1, 7]. These studies allow to identify, measure, quantify, analyze and understand the risks associated with extreme natural phenomena, establish an action model that places great emphasis on preventive aspects, since it is not only about responding to the phenomena, but about anticipating by identifying hazards and their risks, how to manage them, transform them and modify them with the objective of reducing susceptibilities [8]. The studies also offer an overview of the coastal and marine environment and the socio-economic conditions to act proactively in the face of natural disaster risk [1] to, from their results, contribute to the development of disaster reduction plans, management territorial and the design of sustainable economic and social development policies. However, despite the advantages obtained in the implementation of the HVR studies for the province of Santiago de Cuba, the following difficulties can be seen as fundamental:

I. Significant amount of data that is scattered or not accessible. Not all decision makers can access the information resulting from HVR studies in the province. The results were stored in abundant maps, spreadsheets and hard-to-understand text documents for quick and timely decision making;

II. For different actors to access the required information, it must be duplicated, which does not guarantee the reliability and integrity of the data [9]. This may result in documents with conflicting information and does not ensure access by specific levels of users;

III. There is no homogenous format that allows presenting the data properly, nor the ability to analyze them according to the different perspectives and types of danger. Neither does the way in which the information is presented today allows us to analyze the effects that the interactions between different frontal systems would have;

IV. There is no analysis or reporting system that generates useful information for planning, neither does it consider the new information needs of risk managers and decision makers;

V. The information requires systematic updating (through censuses, inventories, frontal systems that have occurred over time, etc.), however, there are currently no information systems in the country that are considered true monitoring systems for the management of HVR studies. The follow-up processes that are carried out are often manual.
The information above justifies the following question: How to contribute to improve the decision-making process in reducing risk to hydrometeorological events from the analysis of the current results of the Danger, Vulnerability and Risk studies?

The current research describes the design and implementation of a decision support system (DS) with support in a data warehouse, which optimizes the processing of the results of HVR studies in the province of Santiago de Cuba before extreme hydro meteorological events, to be integrated for analysis and obtain relevant information with the desired level of detail for a quick and effective decision making in disaster risk reduction.

2. Methodology
Through the review of the methodologies that support the studies of HVR in the face of the exposed hazards and the results collected from the spreadsheets, it was possible to identify which parameters were taken into account to calculate the vulnerability, to the different categories of Hurricane (I, III, V), in each Popular Council of the nine municipalities that make up the province. For high winds and heavy rains, all municipalities were analyzed. In the case of the danger of the sea penetration, only the two coastal municipalities that the study region has are valued, which correspond to the municipalities of Guamá and Santiago de Cuba [3].

Vulnerability is classified according to its type as: structural, non-structural, functional, economic, social and ecological, according to different criteria and expressed from the mathematical point of view as a value limited between zero and one. It takes value \(1\), when the damages are total and, zero \(0\) when these are null [1]. Table 1 shows intervals for assessment different level of susceptibility used in this research. Table 2 reflects the parameters that define the calculation of vulnerabilities to the hazards of coastal flooding due to sea penetration (1), heavy rains (2) and effects due to high winds (3) respectively.

In the case of coastal flooding due to heavy rains, physical vulnerability is also calculated as the sum of structural, non-structural and functional vulnerabilities [1]. From the results, the total vulnerability \(V\) is obtained as a sum of the previous ones and the estimated level of susceptibility for each popular council is evaluated by intervals as shown in Table 1.

Considering the value of the danger of a potentially damaging event with a certain intensity, which is estimated through topographic, geomorphological, permeability criteria, among others, the total risk \(R\) for each hydrometeorological hazard was calculated as:

\[
R = P \times \sum V
\]  

And is valued as shown in Table 3.

To estimate the risk range of the coastal flood hazard due to sea penetration, a comparison of this value is established between the popular councils studied in each municipality or province, the maximum risk value obtained within them is considered for the same intensity of danger and the highest obtained is divided by three to establish the range and breakdown at high, medium and low risk [1]. This procedure is performed for each hazard intensity analyzed.

2.1. Business Intelligence in the management of HVR studies in Cuba.
In the investigation the appropriate techniques and tools proposed by the Business Intelligence (BI) process were used to exploit and analyze structured information on a specific area [10], regularly contained in data stores and which have as Aim to develop perspectives that lead to better and better decision making [11]. Among the features provided by these components are:
I. Accessibility to information: Data is the main source of this concept. It is a priority to guarantee users' access to them regardless of their origin [12];

II. Increase efficiency: With BI all the information can be centralized and visualized on the same platform and converted into useful and organized information, saving time and making decision-making more efficient;

III. Support in decision-making: It seeks to go further in the presentation of information, so that users have access to analysis tools that allow them to select and manipulate only those data that interest them [13].

Table 1. Intervals for the assessment of the level of susceptibility. Source: Modified from AMA 2014

<table>
<thead>
<tr>
<th>Hazard</th>
<th>Vulnerability Level</th>
<th>Low</th>
<th>Average</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal flooding due to heavy rains</td>
<td></td>
<td>0-0.33</td>
<td>0.34-0.66</td>
<td>0.67-1</td>
</tr>
<tr>
<td>Coastal flooding by sea penetration</td>
<td></td>
<td>0-0.33</td>
<td>0.34-0.66</td>
<td>0.67-1</td>
</tr>
<tr>
<td>Affectations due to high winds</td>
<td></td>
<td>0-0.3</td>
<td>0.4-0.6</td>
<td>0.7-1</td>
</tr>
</tbody>
</table>

Table 2. Definition of the parameters for the calculation of Vulnerability to the different hazards. Source: Modified from AMA 2014.

(a) | Vulnerability Type | Parameters (1)          | Parameters (2)          | Parameters (3)               |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural</td>
<td>Permeability</td>
<td>Permeability</td>
<td>Construction damage</td>
</tr>
<tr>
<td></td>
<td>Pending</td>
<td>Pending</td>
<td>Housing affectation index</td>
</tr>
<tr>
<td></td>
<td>Ground level</td>
<td>Ground level</td>
<td>Housing quality index</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>Location</td>
<td>Construction height index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wooded index and affected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>buildings</td>
</tr>
<tr>
<td>Non-structural</td>
<td>Affectation of transport</td>
<td>Affected roads</td>
<td>Access roads blocked</td>
</tr>
<tr>
<td></td>
<td>routes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sewage system</td>
<td>Drainage network</td>
<td>Aerial power networks and</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>high voltage towers affected</td>
</tr>
<tr>
<td></td>
<td>Aqueduct system</td>
<td>Damaged electrical</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>networks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other vital lines</td>
<td>Gas lines affected</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communication lines</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>affected</td>
<td></td>
</tr>
<tr>
<td>Functional</td>
<td>Availability of generator</td>
<td>Availability of generator</td>
<td>Availability of generator sets</td>
</tr>
<tr>
<td></td>
<td>sets</td>
<td>sets</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Health system preparation</td>
<td>Health system</td>
<td>Health system preparation</td>
</tr>
<tr>
<td></td>
<td>Accommodation capacity</td>
<td>Accommodation capacity</td>
<td>Accommodation capacity</td>
</tr>
<tr>
<td></td>
<td>Access to isolated areas</td>
<td>Access to isolated areas</td>
<td>Reserve supplies</td>
</tr>
<tr>
<td></td>
<td>Reserve supplies</td>
<td>Reserve supplies</td>
<td>-</td>
</tr>
<tr>
<td>Social</td>
<td>Population Affectation</td>
<td>Population Affectation</td>
<td>Population Affectation</td>
</tr>
<tr>
<td></td>
<td>Risk perception</td>
<td>Risk perception</td>
<td>Risk perception</td>
</tr>
</tbody>
</table>

4
Table 3. Intervals for risk level assessment. Source: Modified from AMA 2014

<table>
<thead>
<tr>
<th>Hazard</th>
<th>Risk Level</th>
<th>Low</th>
<th>Average</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal flooding due to heavy rains</td>
<td>0-0.33</td>
<td>0.34-0.66</td>
<td>0.67-1</td>
<td></td>
</tr>
<tr>
<td>Affectations due to high winds</td>
<td>0-0.11</td>
<td>0.12-0.43</td>
<td>0.44-1</td>
<td></td>
</tr>
</tbody>
</table>

2.2. Use of multidimensional modeling technique for data storage.
As part of the design of the computer proposal, the multidimensional model was used to store the data obtained from the HVR studies carried out. This model is a technique of designing database structures that takes into consideration essential elements of the standard conceptual model “entity-relationship”, and which as advantages it offers, simplicity of navigation, survey and allows the rapid analysis of information based on of several indicators to support decision making [14]. It consists of two types of tables: dimension tables, each formed by tuples of attributes of the dimension and a fact table, composed of tuples, one for each recorded fact. This fact contains some variable or variables measured or observed [15]. The model proposes three multidimensional schemes, and because of this research the two employees are exposed: 1) The star scheme, which consists of a central fact table and several dimension tables related to it through their respective keys. This model must be totally denormalized, that is, it cannot be presented in the third normal form (3rd FN) [13] and the 2) constellation scheme: a model composed of a series of star schemes, formed by a main fact table and by one or more tables of auxiliary facts, which may be summaries of the principal. These tables lie in the center of the model and are related to their respective dimension tables [16].

2.3. Data Warehouse usage to support the DS-HVR.
The variety of formats, types of storage and the permanent difficulty in presenting reports derived from the analysis of HVR studies in the province turns necessary the structural transformation that modifies the current process that manages the results. As an option to change, analytical tools and techniques provided
by business intelligence were selected, which among its solutions is evidenced the creation of Data Warehouses: data repositories related to the activities of a company and recorded in a database designed specifically to obtain strategic information and support the decision-making process [17]. These repositories also enable the extraction of data from operational systems and external sources, to then be integrated, homogenized and subsequently analyzed by the indicated users, allowing timely, reliable and integrated information [16]. The proposal for the implementation of a data warehouse, made it possible to develop a design focused on responding efficiently to what are the elements, taken from all external sources, which most affect the levels of susceptibility and the likelihood of being affected by exposure to a certain hazard and how to be analyzed from different approaches and in the period determined. The above responds to the fundamental characteristics expressed by [18] about data stores: theme-oriented, integrated, time-variable and non-volatile.

2.4. Data warehousing architecture
In the proposal for the selection of a data warehouse as a solution to the problems raised, the need to build repositories that individually house data related to the parameters and indicators that define the state and behavior of the province in the face of confrontation with each of the mentioned hazards, to later compose the general warehouse on which its subsequent analysis can be carried out in an integrated way. To reach it the Data Mart resource is used, which is no more than the implementation of a Data Warehouse with scope restricted to a functional area, problem, issue or group of needs. According to the operations that are desired or required to be developed, the data mart is expressed through the following architectures: Top-Down: first the warehouse is defined and then the data mart is developed, constructed and loaded from it and Bottom-Up: where the warehouse is built centrally and incrementally over time from the data mart independently developed [19, 20, 21].

The construction of the data warehouse was conducted by a process called Data Warehousing: responsible for extracting, transforming, consolidating, integrating and centralizing the information obtained from operational sources and defining the set of tools necessary to consult, analyze and present the information [16] so that the managing staff does not require advanced computer skills for access.

The stages that are carried out during data warehousing, adapted to the Bottom-Up architecture defined above, are used in the present investigation and consisted primarily of examining and selecting the external sources from which the data resulting from the studies will be obtained. In this case, spreadsheets referring to the two stages 2011 and 2016, to then be extracted, transformed to solve problems of inconsistency between them and guarantee their optimum quality, and once purified they are loaded to the warehouse.

The data warehouse is managed in terms of fact tables: precisely containing fact, numerical values or result of mathematical expressions of high utility. In this context are those referring to the parameters and indicators defined for each hazard; and the tables of dimensions: perspectives for estimating the facts, as they are, by popular councils, municipality, among others; and they are organized around a multidimensional database or cube. It is important to note that these procedures were performed for each hazard and that once completed they are visualized and explored in an integrated manner. Finally, through the Online Analytical Processing (OLAP) operation, the DS-HVR is formed, which allows, through an intuitive graphical interface, that decision makers can generate multidimensional queries to the data and obtain desired reports in real time and in a familiar format that support the decision-making process to minimize the negative impact of these events.
2.5. Methodology and tools used during data warehousing.

The Pentaho open source suite is defined as a BI platform and provides the set of utilities for the construction of data warehouses. Pentaho Data Integration (KETTLE) V5.3 was used, which includes a set of tools, which, in a graphic way, speeds up the process of extracting, transforming and loading (ETL) the data contained in the spreadsheets towards the multidimensional database. Data sources were accessed through visual components and prepared for analysis, relevant transformations were defined by interconnecting blocks and with the adjustment of various functions they were programmed for execution.

Still, Mondrian Schema Workbench V3.8, is the mechanism that from a visual interface provided connection to the defined schemas, corresponding to the behavior of the data for each hazard and that make up the warehouse, to make the multidimensional scheme or cube in terms of dimensions, hierarchies and facts, and subsequently perform the OLAP operation and finally through the BI-server V6.0. The cubes are published, accessible through the browser, and with the use of various add-ons such as Saiku Analytics and Pentaho Audit, among others, the processes of exploration, analysis and information security are improved and pertinent reports are created dynamically in support of the role of risk managers in the event of the different hazards.

The methodology used for the construction of the Data Warehouse was the HEFESTO methodology. It was selected for its many benefits which correspond to the following characteristics: a) The objectives and expected results in each phase are easily distinguished and are simple to understand; b) is based on the user's requirements, so its structure is able to adapt easily and quickly to changes in business and; c) punctually specify the steps to follow in each phase, unlike other methodologies that mention the processes but do not explain how to explain them [16]. Figure 1 shows the steps that guide this methodology and were applied to develop the results that this work presents.

![Figure 1](image1.png)

**Figure 1.** Steps oriented in the Hephaestus methodology. Source [17].

3. Discussion and Results

The spreadsheets from which the results of the HVR studies are obtained have a design that can hardly be interpreted by the exposed analysis tools, so it became necessary to create intermediate sheets in the document itself that respond to the format requirements, making links to the data without duplication.

According to the provisions of the Hephaestus methodology, these documents were examined and the correspondence with the indicators and analysis perspectives that define the scope of the Data Mart that make up the Data Warehouse was determined (Figure 2).
Figure 2. Establishment of correspondences for the control of values by vulnerability, for each type of hazard.

For the correspondence between the data sources and the determination of the indicators and analysis perspectives, the need to form a new secondary fact table for the estimation of the risk associated with the corresponding hazard is established, since it is independent of the type of vulnerability, and therefore the Constellation scheme, described in (2.1), is incorporated into the design of the multidimensional database [22, 23].

To design the data mart that make up the data warehouse, the Star and Constellation schemes were used. The database design is shaped from the logical models defined in Figures 3 and 4.
Figure 3. Logical model of the Data Mart for each type of danger.

Figure 4. Logical model of the Data Mart for each type of hazard.
Once the multidimensional database was created, the dimension tables and the fact table of the data mart defined for each hazard were loaded through the ETL process (Extraction, Transformation and Load) as part of the premises that BI incorporates to eliminate inconsistency and heterogeneity in the data contained in the operational sources.

Once the database is loaded, the multidimensional cubes are built. The structure of the cubes is carried out in correspondence with the data mart defined, with the aim of representing the flat data found in the rows and columns of the database in a matrix of N dimensions [17], in which for each data mart, the values of the parameters and indicators that act as facts and that are determined from predefined or customized aggregation functions are included for subsequent evaluation criteria, and the analysis perspectives represented by the dimension tables.

Finally, the DS-HVR is formed, accessible through the browser, and in which the created schemes are selected, corresponding to the distribution of the developed data warehouse, and visualized the dimensions, measures or facts that compose them and with only to carry out drag and drop operations of these components within visual containers, the query or report is constructed in a tabular or graphic format dynamically, with improved representation and export options to different formats such as PDF, Excel and PNG.

Within the system, different lists and option buttons are displayed that provide navigability to the interactive and dynamic analysis of the data contained in the warehouse.

Figure 5 shows the multidimensional consultation carried out by the user in tabular format, corresponding to the result of the multi-vulnerability and multi-risk analysis of the popular councils belonging to the coastal municipalities Santiago de Cuba and Guaná, before the hydrometeorological hazards, simultaneously. The query is constructed by dragging and dropping operations, giving the manager the possibility to modify and display their results dynamically as it is structured. Figure 6 shows the behavior in vulnerabilities of the popular councils of the coastal municipalities of the province for the year 2016, specifying the differences before the manifestation of the three dangers in unison. In this way, the changes occurred in the territories before exposure to events and alert about the level of attention they should receive to reduce fragility conditions are verified.

**Figure 5.** DS-HVR. Tabular Report obtained from multi-vulnerability and multi-risk analysis in coastal municipalities of the Santiago de Cuba province (2011 and 2016).
3.1 Information security management from DS-HVR.
Access to the information contained in multidimensional cubes can be restricted based on the different permissions granted to each role defined in the system. So, users belonging to a specific role will have specific access to content defined only for their level.

From DS-HVR it is possible to configure this functionality at the same time as it is enhanced with the inclusion of the Audit V2.0 add-on, to manage the entire record of activities occurred within the platform. It means that the user in charge of managing the contents, may Verify all the historical actions carried out by all the users that have accessed the system, guaranteeing the correct operation and security in the management of the stored information, given its vital importance. The present research takes into account the analysis of threat, vulnerabilities and risks and data bases to consider in the study area [3,24,25].

Figure 6. DS-HVR. Graphic Report on Bars of multi-vulnerability analysis in coastal municipalities of the province (2016).
4. Conclusions
The Hazard, Vulnerability and Risk (HVR) studies before hydrometeorological events, are among the first of their kind elaborated and implemented in Cuba. Because of their importance, they constitute instruments for disaster risk management and timely decision making for governments at their different levels of action.

In this article we present a large-scale model that optimizes and computerizes the management of HVR studies. With the result achieved, all the hydrometeorological hazards information is centralized in a single database, which was designed in a multidimensional way when evaluating historical trends in a comparative way. The decision support system (DS-HVR) is formed, which, integrated into this database, allows studies to be analyzed in an integrated way and from various perspectives, to obtain relevant and better represented information through dynamic reports, which guarantees fast and effective decision making.

Our work also includes the following contributions:
- It allows obtaining information through interactive consultations with the desired level of detail, performing simple operations without this implying to have advanced programming or database knowledge;
- It provides data representation in 11 types of graphics that provide dynamism, flexibility and high power of intuition, with style options and real-time filtering, adapting easily to the information needs of decision makers;
- It is immediately accessible through the network, which allows the information to be available to several users at the same time;
- It guarantees information security, through access restricted by specific roles where only the actors with defined permissions can access the information.

In future work we intend to:
- Extend the data set integrating other hazards of natural origin such as seismic threats, severe drought, forest fires, technological threats, among others;
- Develop appropriate probabilistic and statistical models, through the use of specialized computer tools such as Weka, CAPRA and CRISIS, which allow estimating, quantifying or predicting disaster risk in objective terms, given the relevance of these aspects as necessary components to be incorporated into territorial planning processes to face the impact of these hazards;
- Build a geospatial module integrated to the platform that allows representing the indicators highlighted by type of threat, spatially distributed through different shades of color, where the intensity gradation expresses the intervals of the indicator in territorial units.

Acknowledgments
The authors like to thank the Master's program in Integrated Coastal Zone Management coordinated by CEMZOC at the Universidad de Oriente, Cuba, as well as the project Project Index No. INV.1106-01-007-12 named “Resilient cities: minimization of vulnerabilities to phenomena extreme weather and climatic changes in coastal communities”.

References


