# Geospatial Data System Architecture for Disaster Risk Management

Ever Enrique Castillo Osorio[1], Bashir Hayat[2], Kyong Hoon Kim[2] and Ki-Il Kim[3*]

*[1]Department of Urban Engineering, Gyeongsang National University, Jinju, Korea*
*[2]Department of Informatics, Gyeongsang National University, Jinju, Korea*
*[3*]Department of Computer Science and Engineering, Chungnam National University, Daejeon, Korea*
*[*]kikim@cnu.ac.kr*

## *Abstract*

*Currently, raster information taken for satellite sensors are highly necessary in activities related to disaster risk management. However, the images that come from satellites of polar orbit still present a low temporal resolution and the images that come from satellites of geostationary orbit have a low spatial resolution. Therefore, these limitations affect the processes of analysis of geographic information in the systems and search and rescue activities are not feasible to perform at the required time. Therefore, it is required to introduce new technology to improve disaster risk management by providing coarse and multi-view images. To achieve this, satellite images, aerial photographs captured from airplanes and UAVs, data collected by ground wireless sensor networks and cartographic information are integrated using a new-layered architecture. In this paper, we present a proposal methodology for processing raster information and their integration in an informatics architecture under the concept of data mart that allows the collection and storage for timely distribution of raster and vector geospatial data from different sources. Stakeholders can access to specific data and using this for activities of search and rescue actions of vulnerable population affected by natural disasters.*

***Keywords**: Spatial database, system architecture, geographic information system, satellite image processing*

## 1. Introduction

Current disaster management system such as detecting and monitoring flooding is largely dependent on raster image. But, in many situations of search and rescue, institutions cannot make use of information from different sources because have not been integrated under computer and geospatial criteria in a spatial data base that allows selecting what is necessary at the right time. Therefore, fast search task is not feasible.

However, the current technology can be taken advantage in such a manner that data is obtained using variety of devices in real time to the places affected by natural disasters and analyze the situation of the vulnerable population. Satellite images, aerial photographs captured from UAVs, and data collected by ground wireless sensor networks, integrated with cartographic information can be used to search and rescue activities.

To achieve this, in this paper, we explain a proposal methodology for processing raster information and their integration in a database architecture under the concept of data mart that allows the collection and storage for timely distribution of raster and vector

geospatial data from different sources, so it can be used in activities of disaster risk management quickly and efficiently.

Having the right information at the right time, specially in the prevention phase, will be decisive for adequate disaster risk management, thus being able to reduce the potential damages that internal and external geodynamic events could cause to the human being and to the activities undertaken in interaction with the planet.

The rest of this paper is organized as follows. First, we explain the research background and define the challenges. In the Section 2, an overview and related work of the research topics is briefly described. The new methodology for processing of raster image data are presented in the Section 3. In Section 4 is presented the geospatial data system architecture for disaster risk management. And then, we make a conclusion in the Section 5.

## 2. Overview

In this section, we introduce about geographic information system, raster and vector models, spatial databases and data mart, which are related to this work. In addition, we present the lines of research through the various analyzes in the development of geospatial data systems architectures.

### 2.1. Geographic Information System

Geographic Information System o GIS, is a technology of geographic information management, and is formed by properly programmed computer equipment, software that allows managing spatial data and performing complex analysis (functions and methods) with these, following the criteria determined by a team of experts and according to user needs. Therefore, there are five elements that constitute a system of these characteristics, the software, the hardware, the geographic database, the functionalities and the users. Although all of them have to fulfill their function to make the system effective, there are differences in their relative importance [1].

In addition, geographic information is the differentiating element of a geographic information system versus other types of information systems, so this type of information contains two different components: on the one hand there is the spatial component and on the other the thematic component of the data. For example, a river that has a corresponding geometric shape captured in a plane also has other associated data such as pollution levels, flow level, length, *etc*. An example of a map developed in geographic information systems is shown in Figure 1.

Therefore, the GIS has to work with both pieces of information: its location perfectly defined in the map and its associated thematic attributes. That is, this has to work with cartography and with databases at the same time, linking both parts and forming with them a single geographic database. This ability to associate thematic databases together with the precise spatial description of geographic objects and their relationships (topology) is what differentiates a GIS from other information management systems [2].

**Figure 1. Map Developed in Geographic Information Systems**

### 2.2. Spatial Data Set

A spatial data set manages in a matrix, the information of geographical places by columns (the spatial component) and the attributes of these places by rows (the thematic component). The great advantage of this matrix is that we can quantitatively process the information, for example to check the spatial variation of a variable we only have to analyze a row.

The spatial component refers to the geographic location of the objects, either geometrically or absolutely in relation to some external reference system and the topological qualitative relationships that it maintains with other spatial objects, for example: to the left of, under, inside, *etc*. This topological referencing facilitates the development of complex analysis and operations with spatial data and is what distinguishes a GIS from a digital mapping system that only uses absolute referencing.

The thematic component of the data represents patterns of variation that have certain regularities both in space and in time. We can identify two types of behavior of this data:

- The spatial autocorrelation, in which the thematic values are more similar between near objects in the space than between objects far from each other.

- The temporal autocorrelation, in which the thematic values also change in the time axis and the changes are gradual.

### 2.3. Data Models

The data model is the rules used to create a real-world model. The two main models of data used in GIS are the vector model that emphasizes the properties and the raster model that emphasizes location.

The vector model uses vectors defined by pairs of coordinates relative to some cartographic projection system for the description of geographic objects. With a pair of coordinates and their altitude, they generate a point, with two points they generate a line, and with a group of lines the polygons are constructed.

The raster model is based on an implicit conception of neighborhood relations between geographic objects. Divide the study area of the database into a regular mesh of small cells (pixels) and assign a numerical value to each cell as a representation of its thematic value. Since the mesh is regular (the pixel size is constant) and the coordinate position of the center of each cell is known, then all pixels are georeferenced.

### 2.4. Spatial Database

The spatial database is a collection of data referenced in space that act as a model of reality. A repository that contains spatial data within a database management system. It can contain vector data, raster data, tables, and other GIS objects. In addition, the term geodatabase is used to name a geographic database that is a relational database containing geographic information [3].

The spatial database model supports an object-relational model of vector data. In this model entities are represented as objects with properties, behavior and relationships. Supports a variety of different types of geographic objects built into the system. These object types include simple objects, geographic features (objects with localization), networks and topology (objects having spatial relationships with other characteristics), annotation features, and other more specialized types of features. It also saves raster data as a store of this information.

### 2.5. DataMart

It is a specific version of data warehouse. They are subsets of data with the purpose of helping in an adequate decision-making about a specific activity. The data within a data mart can be grouped, reviewed and disseminated in multiple ways so that different groups of users can use them in the most convenient way according to their needs.

One definition of data mart is "a flexible set of data, ideally based on the most atomic (granular) data possible to extract from an operational source, and presented in a symmetric (dimensional) model that is most resilient when faced with unexpected user queries"[4].

The Data mart is a query-oriented system, in which data loading (high) batch processes with low and known frequency are produced. It is consulted using OLAP (On Line Analytical Processing) tools that provide a multidimensional view of information. In these databases, it is possible to build EIS (Executive Information Systems) and DSS (Decision Support Systems).

### 2.6. Related Work

Some studies have been carried out to propose systems and architectures for disaster risk management. The author in [5] present an architecture of 3D of geographic information system along with its map service protocol for providing maps that can be used for aircraft simulation route during search and rescue operations, but is not integrated with information captured from satellites or UAVs. In [6] a real time satellite data auto-processing system has been developed and through the operation of the system a massive enhancement images in database have been set up, however, it is only applicable for satellite. The author in [7] present a system structure for building remote sensing images database and the architecture is based in web service technology. In [8] the researcher show an integrate Cloud-GIS and supporting technologies for the fast multi-sensor data fusion and describes the general target of the architecture, but is applied for risk assessment in power systems. The research explained in [9] was done about a new process and architecture for remote sensing image data mining based on data warehouse and a prototype system was design.

## 3. Processing Analysis of Raster Data

Raster information will be store in the database, but first it must be processed. According with that, the proposal methodology that is developed for the processing, describe and involves a set of activities, such as the use of algorithms to perform an appropriate classification of the information, including the geometric, spectral, spatial and radiometric characteristics, and finally must be converted to vector format, and stored in the spatial tables of the database. Figure 2 show the methodological schema.
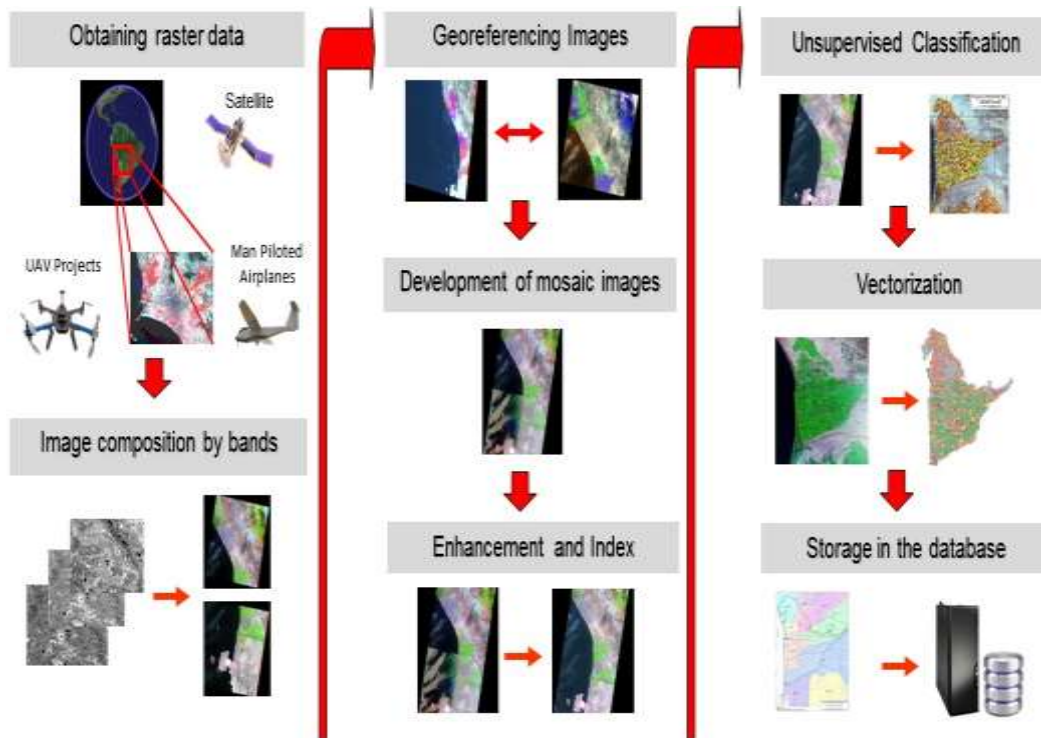


**Figure 2. Methodological Scheme for Processing of Raster Information**

The methodology allows performing numerous operations on a raster image and generating an answer to specific geographical questions. By manipulating imagery data values and positions, it is possible to get features that would not normally be visible. The process of the methodological scheme is described in next paragraphs.

### 3.1. Obtaining Raster Data

The images in raster format are obtained from the data capture equipment, such as satellites, airplanes or UAVs. According to the characteristics of equipment, such as its position relative to the ground, amplitude and coverage of the lens, data processing capacity, among other factors, the resolution of the information will also be determined.

Currently, raster data can also be obtained from official web sites of the space programs. In all cases, the metadata must be verified, because it is detailed the geometric and cartographic information of the image, and indicates the level of pre-processing that has the raster information downloaded. The web portals for downloading raster images used in the research are:

• Earth Resources Observation and Science Center (EROS) web page, of the U.S. Geological Survey (USGS): http://glovis.usgs.gov
  This web portal contains information about various space programs such as LANDSAT, ASTER, MODIS and more.

- The Earth Explorer web page: https://earthexplorer.usgs.gov/

It presents a very friendly system so that the user can search, select and download raster satellite images. This page is also provided by the U.S. Geological Survey (USGS).

- The Earth Science Data Interface of the Global Land Cover Facility web page: http://glcfapp.glcf.umd.edu:8080/esdi/

It presents a graphical search environment for LANDSAT, MODIS and AVHRR satellite imagery. In addition, data can be downloaded about the terrain model and raster information already processed.

### 3.2. Image Composition by Bands

After obtaining the raster images from the date of analysis, in case of satellite images, the combination is performed in a single file for all bands (each band contains information captured within a specific range of the electromagnetic spectrum, and previously they are separated in different files).

For example, in case of LANDSAT satellite image, it has seven bands, or ASTER satellite image has a total of fourteen bands. All these spectral bands are grouped into a single file, and after that from that file, a group of three bands can be selected according to the analysis that is wanted to perform. In case of common band combinations in RGB comparisons for Landsat 7 or Landsat 5 are used bands 7,4,2, but in case of Landsat 8 has the combination of 7,5,3 bands.

### 3.3. Georeferencing Raster Data

If the raster data has no defined the cartographic parameters such as the coordinate projection system and the reference ellipsoid, then should be used checkpoints previously identified in the study area in order that raster image is spatially located.

To achieve this, we work with the Polynomial transformation, in which it use a polynomial based on checkpoints and the algorithm of least squares fitting (LSF). The objective of the algorithm of least squares fitting is derived a general Equations (1) and (2) that can be applied to all points as showed in Figure 3, usually at the expense of a slight movement of the positions of the control points.

$$X' = AX + BY + C \tag{1}$$

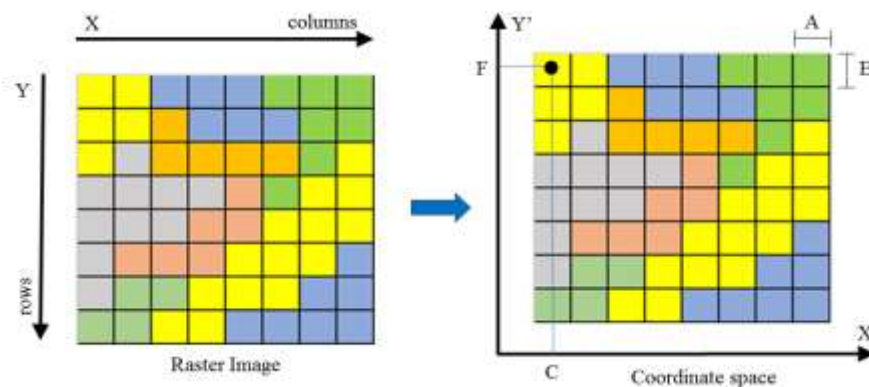$$Y' = DX + EY + F \tag{2}$$



**Figure 3. Application of the General Equations According to the LSF Algorithm**

Where:

- X is column count in raster image.

- Y is row count in raster image.

- X' is horizontal value in coordinate space.

- Y' is vertical value in coordinate space.

- A is width of cell in map units.

- B is a rotation term.

- C is the x' value of the center of the upper left cell.

- D is a rotation term.

- E is negative height of cell in map units.
- F is the y' value of the center of the upper left cell.

### 3.4. Development of Mosaic Images

This process allows matching two or more raster images containing areas of land that are spatially continuous. The method used in this process is Principal Component Analysis (PCA), which is a statistical procedure concerned with elucidating the covariance structure of a set of variables. In particular, it allows us to identify the principal directions in which the data varies [10].

The process in two dimensions of the PCA method is showed in Picture 4, and consists in using all the data points we find the mean values of the variables ($\mu_{z1}$, $\mu_{z2}$) and the covariance matrix $\Sigma$ which is a 2 x 2 matrix in this case. If we calculate the eigenvectors of the co-variance matrix we get the direction vectors indicated by $\Phi = [\phi_1, \phi_2]$ we create a transformation matrix which takes our data points from the [$x_1$, $x_2$] axis system to the axis [$\phi_1$, $\phi_2$] system with the Equation (3).

$$P\phi = (Px - \mu x). \Phi \tag{3}$$

Where:

- Px is any point in the [$x_1$, x2] axis system

- $\mu x = (\mu_{z1}, \mu_{z2})$ is the data mean

- $P\phi$ is the coordinate of the point in the [$\phi_1$, $\phi_2$] axis system.
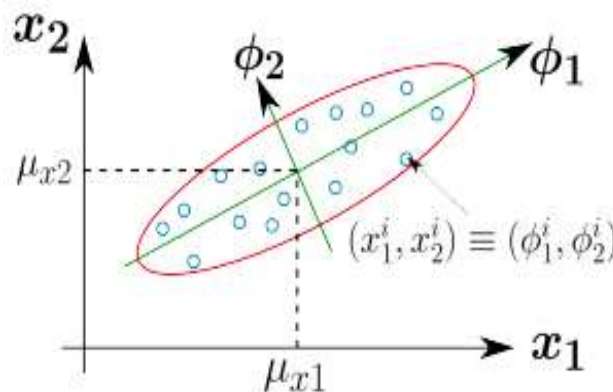


**Figure 4. Process in Two Dimensions of the PCA Method**

## 3.5. Enhancement and Index

This process generates a combined result of bands where spatial attributes related to the physical and a chemical characteristic of the land is highlighted, allowing us to identify the common geographical features of the study area. A particular features to identify are the areas of vegetation, geological structure and composition of the soil.

One of the most important cases is to determine areas with vegetation, because knowing their location, then it is possible to evaluate the areas of the terrain that have greater resistance to the effects generated by mass movements by the action of internal and external geodynamic phenomena. Therefore, their helps in actions of disaster risk management.

For this subject we propose to use the Normalized Difference Vegetation Index (NDVI). In case of satellite images, the raster data acquired in visible and near infrared spectral bands are adequate to determine the large differences in the reflectance of plants and to group them according to their similarity and determine their spatial distribution in these satellite images The NDVI is calculated from these individual measurements with the Equation (4).

$$NDVI = \frac{(NIR - VIS)}{(NIR + VIS)} \tag{4}$$

Where, NIR is the data captured in infrared band, and VIS is the data captured in visible band. NDVI is calculated from the visible and near-infrared light reflected by vegetation. Healthy vegetation absorbs most of the visible light that hits it, and reflects a large portion of the near-infrared light. Unhealthy or sparse vegetation reflects more visible light and less near-infrared light.

Besides, in a multitemporal evaluation, the areas where vegetation has increased or decreased can be determined. In Figure 5 is showed that in the course of 3 different years the vegetation in this area has been decreasing. With data taken in the field is possible to evaluate the factors involved in the fact.
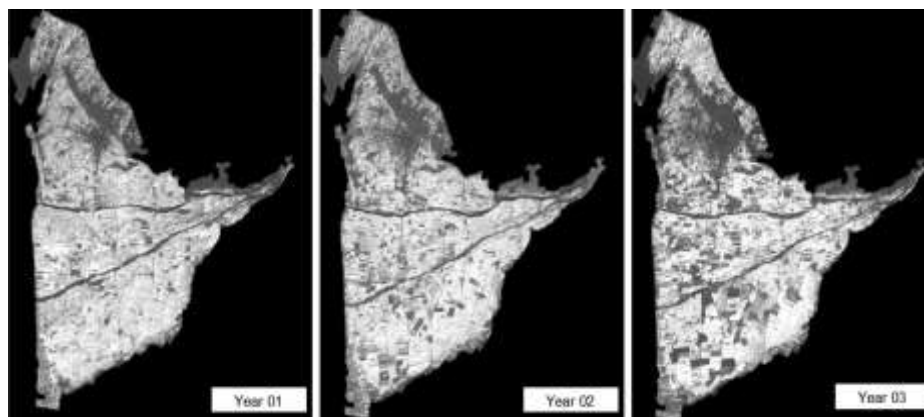


**Figure 5. Results of Multitemporal Images Processed Using the NDVI Method**

## 3.6. Unsupervised Classification

This method allows grouping of pixels in ranges based on reflectance data that they store. Each range is assigned a color for better differentiation, and then monitoring sample points are identified. These must be located in various color ranges to obtain samples for each group.

The algorithm has some further refinements by splitting and merging of the clusters. The objective function (which is to be minimized) is the sums of squares distances

(errors) between each pixel and its assigned cluster center. This is showed in the Equation (5).

$$SS_{distances} = \sum_{\forall x} [x - C(x)]^2 \tag{5}$$

Where C(x) is the mean of the cluster that pixel x is assigned.

Minimizing the $SS_{distances}$ is equivalent to minimizing the Mean Squared Error (MSE). The MSE is a measure of the within cluster variability, and showed in the Equation (6).

$$MSE = \frac{\sum_{\forall x} [x - C(x)]^2}{(N - c)^b} = \frac{SS_{distances}}{(N - c)^b} \tag{6}$$

Where N is the number of pixels, c indicates the number of clusters, and b is the number of spectral bands.

### 3.7. Vectorization

In this process, the raster information (consisting of an array of pixel data) will be converted into a vector format (based on geometric elements such as points, lines and polygons). The pixels that have been classified under the same group in the previous step will be consolidated to make the vector elements. The advantage of format conversion is that each geometric element created contains the information of the element itself, and in this case, the information to be stored is the reflectance range that was classified.

### 3.8. Storage in the Database

This is the final process for storing the raster information. Once this information has been converted to vector elements, these elements are stored under the concept of spatial data tables. In this concept, each vector element is linked to a record with data from this feature, and stores geometric data such as position (coordinate pairs) and location based on the projection system and reference ellipsoid in which it has been spatially georeferenced.

## 4. System Architecture for Disaster Risk Management

Raster information captured from different sources (satellite, aircraft, UAVs), must be integrated into a flexible system that allows the user to access this information immediately to search and rescue actions in case of natural disasters. The proposed system architecture work in four layers, links and integrates the raster information with cartographic information and other ground data in a database that is able to provide processed information, using protocols for design, storage and publishing based on open software. The methodological schema is showed in the Figure 6.
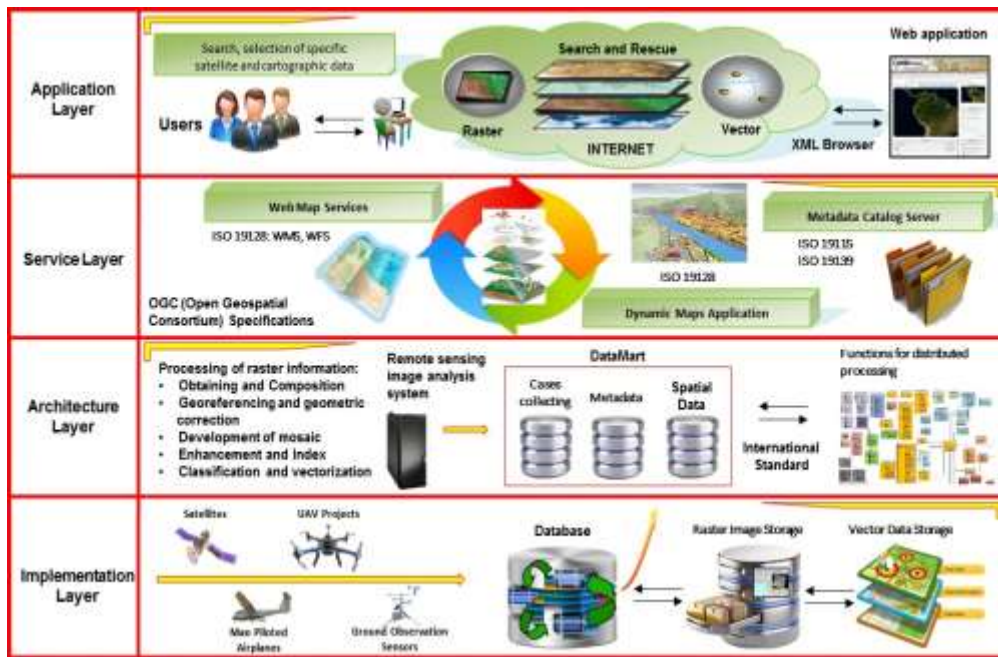
**Figure 6. Methodological Scheme for System Architecture**

### 4.1. Implementation Layer

In this layer, it is required to defined the sources of information (satellites, main piloted airplanes, UAV projects and ground observation sensors), and the hardware architecture, operating systems, networking environment, storage networks and other infrastructure which is necessary for work. In addition, it has been defined to work with vector and raster cartographic information in different storages but in an integrated way. This has been defined, because the raster information, will be converted to vector information, so should be avoided duplication in the same database.

### 4.2. Architecture Layer

We define the mechanisms and structures for managing massive raster data and geographic information for the system as a whole with security integration. Is planning also in this layer to apply the pre-processing described in the section 3 of the document, in order to convert the raster data to vector data.

In addition, here is develop tools and internal process to manage the data under the concept of data mart. The proposal of data mart for this layer will have three main topics:

### 4.2.1. The Cases Collecting

This is important to process specific cases according with user's requirements and obtain good analytical data.

Requests should be based on the users' need to find answers about the occurrence of an external or internal geodynamic event and will work under the following steps:

- The user consults the register of metadata catalogs, to review the available information previously processed.

- The user selects the period and place of analysis. It uses a coordinated pair, a reference of administrative or hydrographic boundaries, or a previously delimited area.

- The user sends the request to the server.

- The server processes the request, overlapping the study area with the available cartographic information of the place.

- The server performs the comparison of the data in the period of analysis required by the user.

- The server processes the raster information and extracts the data from the vector elements to be represented as records.

- The server returns the mapping result and data records through web map services.

- The user can consult or download the information obtained.

### 4.2.2 The Metadata

Metadata is the information about the data that will allow its interpretation. It can be used by users, so they know about what information exists in the database and they can get more adequately the results of their requirements.

According to the information needs of the data, and to be correctly interpreted, the metadata can be classified and structured into the following groups:

- Description.

- Inventory.

- Cataloging.

- Publication.

- Diffusion.

- Re-use.

- Search.

- Query.

- Access.

In the scheme presented in Figure 6, the link from the metadata catalog server to the spatial database is considered, where the metadata tabs will be stored and queried according to the ISO 19115 standards

### 4.2.3. The Spatial Data

Is composed by the vector information obtained from the pre-processing of raster information. This information also is linked with the basic cartographic information available.

For proper management of the spatial data, the database must have the capacity to contain the following elements:

- Object class. They are the tables of the database in which the properties of the vectors are associated. Rows in a table work like instances of the object that have special characteristics in the GIS. An example of an object class will be the different soil types feature obtained from the data grid of the raster file.

- Feature class. It is a collection of elements of the same type. An element will be an object that has a location stored in a row as one of its geometric properties. An example of feature class will be the types of soils, types of vegetation and other characteristics obtained from the raster file.

- Feature attributes. They are the properties stored as fields in a feature class table. Attributes will define standard and custom properties of elements and will be numeric, text, image, or descriptive identifiers.

- Spatial reference. It is the land-based coordinate system in which the data set is represented. The spatial reference will include properties such as the cartographic projection of the map and the datum.

- Subtypes. It is a set of classes for the elements of a feature class. A feature class may have elements with the same properties but with a different role or meaning in the data model.

- Feature dataset. It is a collection of feature classes with the same spatial reference. The feature classes will be organized into networks or planar topologies in a feature dataset.

- Relationships. It is the association between two objects. A relationship will allow to work with features and related rows in tables as well as relationships between rows.

- Geometric networks. It is a collection of feature classes that are part of a network of edges, connections and corners. An example will be the water network of the study area.

- Planar topologies. It is a group of geometric relations that share the elements with each other. For example, feature classes, such as terrain, vegetation, and water types, can share common boundaries in their polygons.

### 4.3. Service Layer

The protocols and communication interfaces that will be used to provide services are defined. This service has to provide evaluation function about hazards for search and rescue activities.

For this proposal, OGC protocols for management web services and dynamic maps are used, for example ISO 19128, that defines the web map services available to send data from a remote server using Internet (Web Mapping Service – WMS and Web Feature Service – WFS).

- Web Map Service (WMS): It presents an easier type of connection of cartographic interconnection by Internet. It will return a graphic (gif format) that fits the destination map. It presents limited possibilities to consult the remote data and supports any format, although more appropriate in this case, because the layers of information have nothing to filter.

- Web Feature Service (WFS): This type of connection will allow to download vector information from remote points, and it has all the other functionalities described in the WMS.

Also for metadata service is used the ISO 19115 core, which defines the structure of metadata, and the ISO 19139, which defines how to publish metadata through the web, specially using the XML language.

### 4.4. Application Layer

We are planning to develop a pilot application generated with open source software for data dissemination that allows the user to obtain the necessary information in search and rescue activities. The software works in web environment.

## 5. Conclusion

In this paper, we proposed a new methodology for processing raster information from different sources to obtain vector data. For improve this methodology we employed effective method algorithms at each stage of the process. Furthermore, we

propose and explain a new system architecture for disaster risk management. This new architecture is based on data mart and was designed under layered architecture. In addition, it can include geospatial data according to geometric characteristics. Finally, we explain each layer of the proposal system architecture.

Related to this work, we will continue the research with the implementation of this system architecture and development of experimental results.

## Acknowledgments

## References

[1]  J. Bosque, "Sistemas de Información Geográfica", 2nd Edition, Rialp Editions, S.A., Madrid, Spain, **(1997)**.

[2]  A. Nabil and A. Gangopadhyay, "Database issues on geographic information systems", 1st Edition, Norwell, USA, **(1997)**.

[3]  A. MacDonald, "Building a Geodatabase", Environmental Systems Research Institute – ESRI, Inc, USA, **(2005)**.

[4]  R. Kimball and M. Ross, "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling", 3rd Edition, Wiley Publishing, USA, **(2013)**.

[5]  B. Krisbiantoro, H. Hindersah and T. Mardiono, "3D GIS System Architecture for the Aircraft Simulation Route of Search and Rescue Operation", In: 2012 International Conference on System Engineering and Technology, pp. 1 – 5. IEEE Press, Indonesia, **(2012)**.

[6]  W. Li and X. Shan, "Study of Real Time Processing for Geostationary Satellite Data Applied to Seismologic Monitoring", In: 2010 Second IITA International Conference on Geoscience and Remote Sensing, pp. 227 - 230. IEEE Press, China, **(2010)**.

[7]  J. Shi, Y. Chen and C. Liu, "Database System for Archiving and Managing Remote Sensing Images", In: 2009 First International Workshop on Database Technology and Applications, IEEE Press, China, **(2009)**, pp. 536 - 538.

[8]  L. Li, F. Cheng, W. Sun, Y. Xu and X. Wang, "Study on evaluation system of meteorological hazards for power grid based on Cloud GIS", In: 2014 International Conference on Power System Technology (POWERCON 2014), pp. 1848 - 1852. IEEE Press, China, **(2014)**.

[9]  P. Du and Y. He, "The Applied Research of Remote Sensing Image Data Mining Architecture", In: 2009 Asia-Pacific Conference on Information Processing, pp. 301 - 303. IEEE Press, China, **(2009)**.

[10] Imperial College London, http://www.imperial.ac.uk.