Biological Records Georeferencing and Digital Localities Gazetteer

Rubén Darío MATEUS SANABRIA and Sonia Mireya SUA TUNJANO, Colombia

Key words: Colombia, georeferencing, gazetteer, biological records, information standard.

SUMMARY

The GIS Unit of the Biological Resources Alexander von Humboldt Research Institute (IAvH) is in charge of all the spatial analysis research in the institution. Previous to the formation of this unit, the biological collections datasets of historical biological records contained insufficient information of location or gave a general description about the locality of the organism. With the technological advances in the handling of the information management, and the increasing use of the GIS, the investigators have improved the capacity of calculation and storage of this information. By the use of standards precise data of the description and geographic location of the biological collections have been obtained, allowing relate specimens to different ecosystems or environmental conditions.

The increased necessity to have georeferenced biological information, led us to the construction of a standard for georeference standard; a methodology for the geographic location and a digital of localities of biological registries, accessible to different users. The documentation of data of the locality or place where the registries are collected is indispensable because it provides origin data, and allows present/display and analyzes this information in space and time, generating capacity and knowledge for decision making as far as conservation and sustainable use of biodiversity.

We constructed a localities gazetteer considering it to be one of the most useful tools for the handling of geographic information. This allows collect scattered geographic data that and are of difficult documentation, cost reduction it can help us in with the purpose of allows the reduction of costs, duplicity of efforts and the interchange of data, thus the quality of the data becomes a decisive factor for its use (ICONTEC - technical committee 034).

Biological Records Georeferencing and Digital Localities Gazetteer

Rubén Darío Mateus SANABRIA and Sonia Mireya Sua TUNJANO, Colombia

1. INTRODUCTION

The knowledge of regional natural resources, considering the percentage of species not even discovered, requires the deepest diversity study of the quantity, and ecological distribution of the *biota*. This knowledge comes from basic investigation, which in fact, has conduced to the sign and ratification of several national and international agreements for the defense, protection and the natural resources management. ¨The agreement about the Biological Diversity, signed and ratified by Colombia through the law 165 of 1994, founds in the article number 7 the engagement of the implementation of systems for the data organization who let identify and follow the most important components of the biological biodiversity or the important ones for the preservation and sustainable use; and to the activities that have or could have prejudicial effects for the conservation and sustainable use of the biodiversity. According with this compromise, the Instituto de Investigación de Recursos Biológicos Alexander Von Humboldt (IAvH), has defined as one of his key investigation areas the inventory of the fauna and flora in the national territory.¨ (Rudas et al. 2002).

The big part of this knowledge, comes from different expeditions since the past centuries, has been collected and stored in different museums and herbariums nationally and abroad, and in the early days investigators have begun to catalogue and publish it.

With technological advances in telecommunications an informatics and the increasing utilization of Geographical Information systems (GIS), the investigators who collect the data in the field have improved the capacity to capture, estimate and improve the data. This processes helped to obtain more accuracy in the geographical location of the biological records and in particular of the biological collections. It`s necessary to implement and develop a methodology to capture and georeferenciate biological records, propose new georeferencing standards for localities of actual and historic biological records and to design a new application of digital gazetteer of localities that can help to capture, validation, storage and visualization of alphanumeric and geographical data

2. BACKGROUND

Colombia has a big variety of ecosystems as a result of a high climatic and topographic diversity, and their particular geographical situation is the shelter of one of the richest biotas of the world and a high level of endemism in several taxonomy (flora, herpetofauna, avifauna, mastofauna) (Mittermeier 1999) (Figure 1).

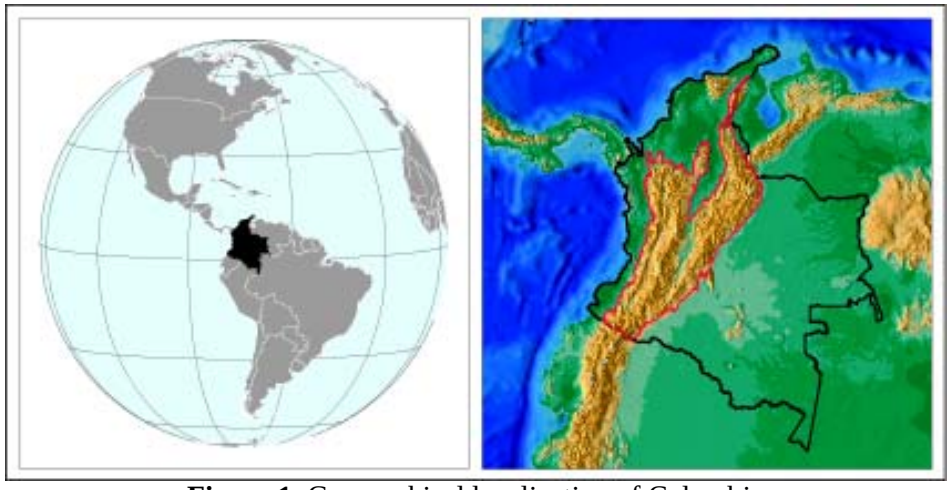

Figure 1. Geographical localization of Colombia

The institute through the Geographical Information Systems Unit (UNIGIS), had developed activities of acquisition, handling, processing, and analysis of cartographic information. In 1999 together with the program of Biodiversity Inventories starts the localities georeferencing process of collection of biological records, with the development of one easy handling application, lets to the investigators works with de historical and actual biological information standardized, and replicable through the time.

Before this work, the historic biological records collected, didn't have enough locality information or simply had a general description about the localization where the organisms were observed or collected. This, because investigators didn't give importance to the collection process, and didn't keep direct relationship between the species characterization versus the spatial location.

The cartography and technological advances let the investigators to make spatial analysis with the records of biological species collected with other spatial related information. Nevertheless, the researchers had to bring face to the fact that the observations didn't have the standardized locations or geographical coordinates, and without it they couldn't show the spatial information for advanced geographical analysis.

In this way, the researchers had to implement some methodologies for the capture and organization of spatial information in simple databases accessible for different kinds of users by developing protocols and standards for the geographical data capture with high trustworthiness level, for these reasons they must be submitted to a quality control in order to minimize the error in the pre-analysis processes.

With available information with the highest levels of precision the capacity to investigation could increase. This data can be shared and used for multiple purposes and optimize the knowledge and the use of information.

The localities gazetteers are one of the most useful tools for the handling of geographical information. They help in the collection of scattered geographical data with low levels of documentation in order to optimize their organization, documentation and standardization.

Using the localities gazetteers reduces costs, efforts duplicity risks and helps in data sharing, for that reason, the data quality becomes a decisive factor for its use.

3. METHODOLOGY FOR THE GEOREFERENCING OF LOCALITIES

In the process, inventory and publication to the scientist community, was built the methodology for georeferencing of localities of the biological records, in the next paragraphs the necessary steps will be detailed starting from the capture of the information, the methodology development, the storage and handling of the database (gazetteer) and the end user.

3.1 Basic Information

The first basic and essential activity for the georeferencing process is the gathering, processing, selection and acquisition of the basic cartography, in analog or digital format. This step, gives a reliable source for the spatial location of the biological records. This information comes from two different kinds of sources:

3.1.1 Primary sources

Localization data obtained with GPS navigators collected from the field sampling, giving primary information useful to get the detailed geographical position of the localities of the collection. Additionally they help to complement the description of some localities of historical records or can be used as new data sources.

3.1.2 Secondary sources

The principal secondary source of information used in the Humboldt institute, is the basic digital cartography, that has the toponymy attributes (i.e. urban centers, geographical names, political division), contour curves, rivers, and ways. This cartography is made by the ¨Agustín Codazzi¨ geographical institute (IGAC) scale 1:25.000 digitized by the National Statistics Department (DANE), entities responsible of their construction and distribution.

Likewise, the digital information of the National Imagery and Mapping Agency (NIMA), that has the information about contour curves, rivers, roads and 62.000 records of toponimical data including urban centers, geographical names, political division and rivers. This information is loaded in the geographical database of the gazetteer.

3.2 Localities Georeferencing

The methodology used for georeferencing the biological records in the different information sources who provide the data about the collection localities. The sources are:

- Existing biological records (historical) without georeferencing.
- Georeferencing records without standardized description.
- Actual records obtained using GPS technologies, cartography or other information sources.

For each case one locality compilation methodology was designed in order to obtain their spatial coordinates.

3.2.1 Historical biological records

This process starts with the acquisition of the geographical dictionaries of the country, localities gazetteer and analogue and digital cartography for the georeferencing process. The figure number 2, shows the principal activities developed in the georeferencing process and the capture of localities for the existent records.

3.2.1.1 Purifying and standardizing the localities names

The localities in the file or database of the records are identified; the minimal attributes and the localization geographical descriptors are depurated, complemented and standardized as the localities names. The localities standardization must depart from the global to the particular; that means, from the punctual description of where the sample is taken to the highest general hierarchical level.

3.2.1.2 Georeferencing the biological records

Once the locality names are depurated and standardized the researchers do the georeferencing based in the description of the place location and the definition of the precision level according with the following established parameters:

- Level 1. Coordinates that have been directly taken in the field with GPS, and when they are compared with the digital cartography, are located in the same place.

- Level 2. Coordinates taken from publications, museums, herbariums, and collection cards and when they are compared with the digital cartography, the localization is confirmed.

- Level 3. Coordinates estimated from the digital cartography where the researcher has the intersection point and its corresponding height or height ranges.

- Level 4. Coordinates taken from the digital cartography but without information about the height (over the sea level) of the place of collection.

- Level 5. Coordinates estimated from the digital cartography taken from the municipal or departmental urban zone as a sampling point, due to the low precision of the locality.

- Level 6. Coordinates estimated from the digital cartography, where the analyst doesn't have accuracy of the locality, in that case, they must be taken inside the administrative or political area.

- Level 7. Coordinates with localization information, where there is no description about the place of collection, but instead, they do the description of the region, river, department, country or municipality.

- Level 8. Localities without enough information for the georeferencing process.

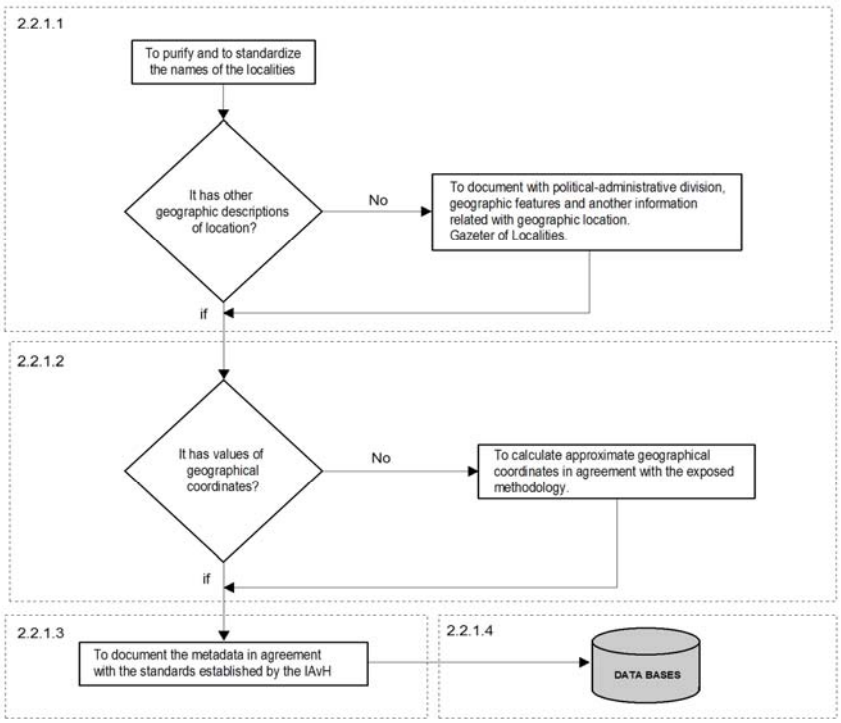

Figure 2. Procedure to georeference historical biological registries

3.2.1.3 Metadata documentation

It refers to the documentation of the metadata describing the source from the locality georeferencing was taken following the established standards in the "Protocol of documentation of the metadata related with biodiversity in the IAvH, (Rivera 2003).

3.2.1.4 Load into the data base of the localities digital gazetteer

Actualization in the data base of the biological records with the locality field that has the ID and the geographical information purified and standardized.

3.2.2 Georeferencing of biological records in field

The principal activities in the localities capture are showed in the figure 3, and are described as:

3.2.2.1 Identification and preparation of the cartography in the sampling zone

Specifically, is definition of the sampling zone and the identification of the zone where the investigators will work with. This process is started with the acquisition of the geographical dictionaries of the country, localities gazetteers, photos and analogue and digital cartography

as detailed as possible. A good cartographic level makes easy the identification of the principal characteristics in the study area like geomorphology, roads and villages.

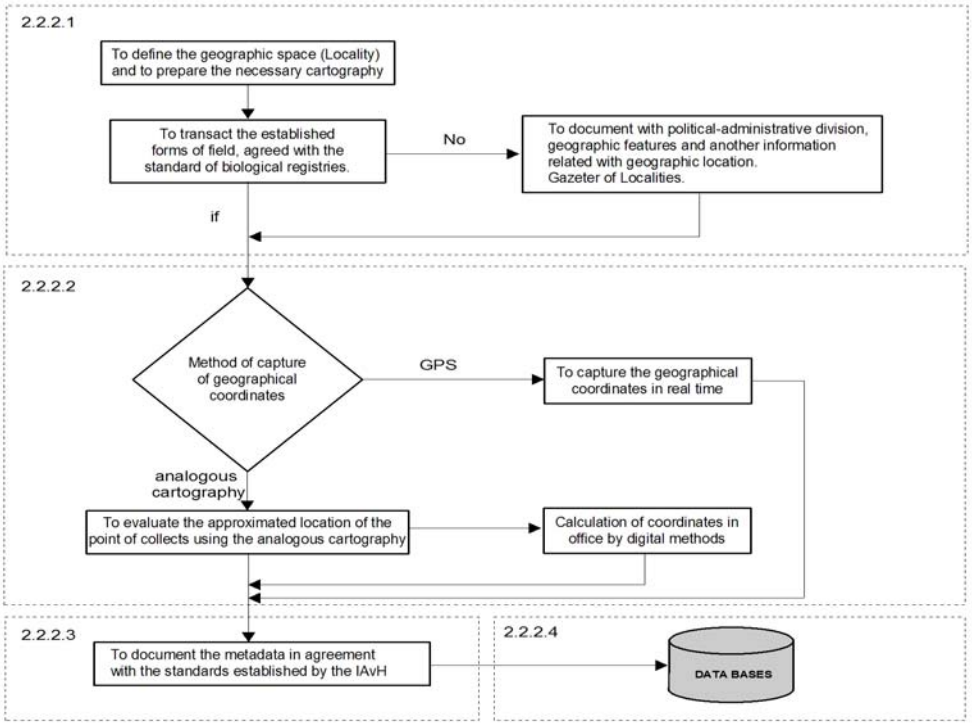

Figure 3. Procedure to georreference biological registries in field

3.2.2.2 Method for the capture of coordinates

In the field work, in the collection place the geographical information is captured. There are two ways to do this process:

- Coordinates capture trough the GPS.

- Coordinates capture through the analogue cartography: When the investigators doesn`t have available the GPS, they must work with analogue cartography.

3.2.2.3 Metadata documentation

In this steps the source where the georeferencing is taken and described following the standards used by the IAvH the “Data sets related with biodiversity metadata protocol document in the IAvH” (Rivera 2003).

3.2.2.4 Loading de localities digital gazetteer into the data base

Is the actualization process in the data base with the locality field in which is contained the ID and the geographical information purified and standardized.

3.2.3 Quality control of the georreferenciation

The collection localities georeferencing, must also follow the process of quality control (figure number 4) of the information stored in the data base, the objective of this step is to identify the possible inconsistencies in a data set in respect to their specifications. (ICONTEC NTC 5205 2003).

3.2.3.1 Sample selection

Is strongly suggested that the number of records to check equals the 15% of the total of georeferenced records, and could be selected using any convenient statistical method.

3.2.3.2 Locality revision

Once the analyst have the list with the selected records, a control format is filled in which the fields to review are showed. When the record is validated, is necessary to consider it under the standards established in the locality normalization. Additionally, is necessary to keep a record of the errors in the standardization of the localities and by this evaluate its quality. It is strongly suggested that if more than 5% of the data has significant differences, to review the georeferenced records.

3.2.3.3 Revision of Geographical coordinates

In order to review the geographical coordinates is necessary to dispose of the analogue or digital cartography. This step helps relate the fields with the localization coordinates. Those records that are not in concordance with the cartography must be reviewed in order to determinate the errors in the geographic coordinates or in the locality description. It is also convenient to keep a record of the corrections for statistical proposes.

3.2.3.4 Metadata documentation

In this steps the source where the georeferencing is taken and described following the standards used by the IAvH the “Data sets related with biodiversity metadata protocol document in the IAvH” (Rivera 2003).

3.2.3.5 Loading into the database

Actualization in the data base of the biological records with the locality field that has the ID and the geographical information purified and standardized.

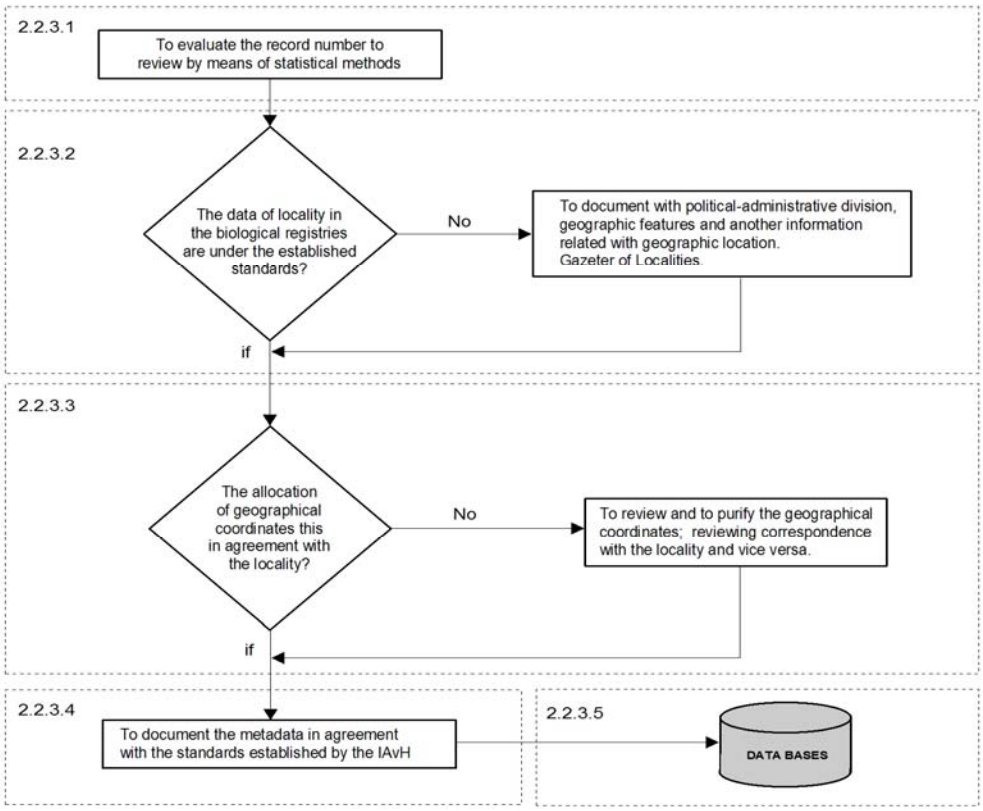

Figure 4. Processes for the control of quality of the localities and its coordinates

4. LOCALITIES GAZEETER APPLICATION

The meaning of the word "gazetteer" when is used in geographical sense, is defined as an "index or geographical dictionary". Then they are used in the atlas or map index, the gazetteers are simply aggregation of lists ordered alphabetically and related to localities or physical and cultural characteristics. (Columbia Gazetteer of the World Online, http://www.columbiagazetteer.org/). The functionality of an actualized gazetteer gives to GIS precise information about toponymy in a specific study zone and increases the reliability related with the input, quick validation and the data interpretation, and with the quick answering of localization questions.

In order to do the implementation of the methodology described above, the localities gazetteer application was developed as a tool for the capture the spatial information of the biological records under the normalized standards. With this application the researchers can do the described tasks and the user handling is almost "intuitive".

The application was built following the same design steps of a data base where the conceptual, logic an physic model are developed an are described as:

4.1 Functional And Conceptual Model

The building process of the schema used by the application is made through the conceptual model. The model representation is based in the Entity–Relationship model designed by Peter Chen and is showed in figure 5. In it, attributes, entities and relationships are identifies according with the perspectives that each user has about the data handling, its nature and its use through the different application areas.

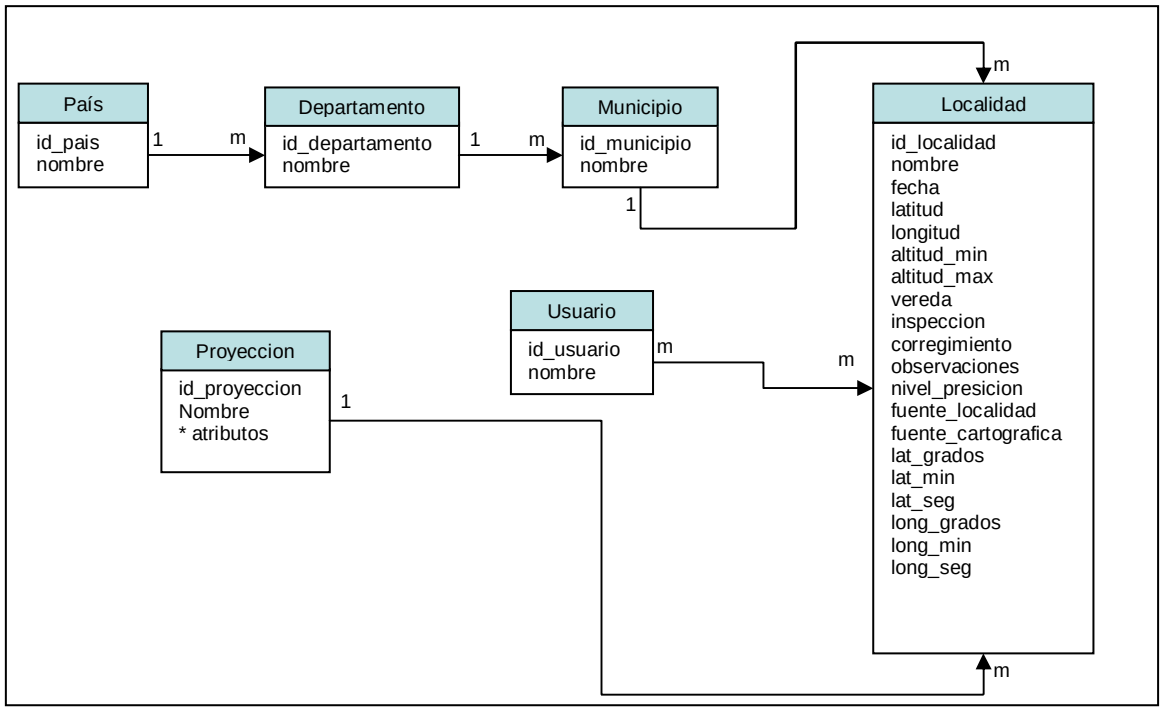

Figure 5. Model Entity - Relation

The tasks that will be done by the application, that is, how will the data transformation be done functions of correspondence, storage, actualization and modification are showed in the functional model (Figure 6) They can be summarized in: georeferencing or locality addition, actualization and localities query.

4.2 Logical Model

In this step, the information scheme used by the application is built. The conceptual model is transformed into the logical model, based in the data structure of the entity relationship model implemented in the data base.

This process is the source of information for the physical model design and plays an important role in the information maintenance.

4.3 Physical Model

Is the implementation of the application, how the user will interact with the stores data; living all the tools for accomplish the storage, speed and visualization requirements.

This application is conformed by three modules, the first one, is the users administration module for the handling, edition and querying of information. The second one allows the addition and modification tasks and the third one are for spatial and alphanumeric queries (figure 7).

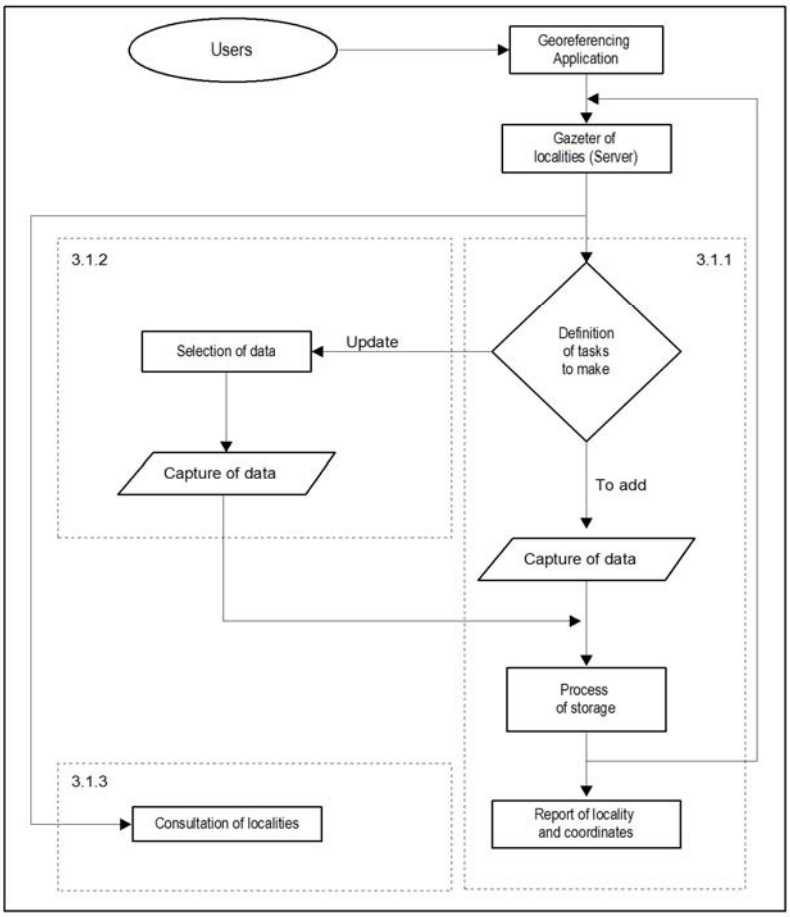

Figure 6. Functional model

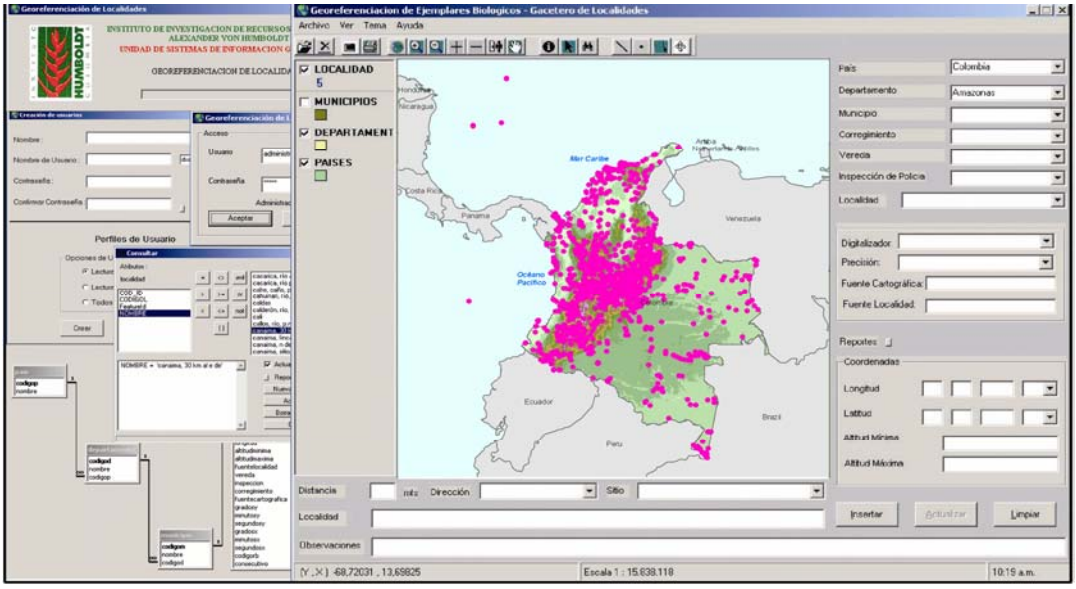

Figure 7. Physical model of the application

REFERENCES

Sua, S., Mateus, R.D. y Vargas, J.C. (2004) Georreferenciación de registros biológicos y gacetero digital de localidades. IAvH.

Mittermeier, R.A., Myers, N. y Mittermeier, C.G. (1999). Biodiversidad amenazada. Las ecorregiones terrestres prioritarias del mundo. CEMEX & Conservación Internacional. 430 p.

Gatrell,A.C., 1991. Concepts of space and geographical data en Maguire, D.J.; Goodchild, M.F. and Rhind, D.W. (Eds.) Geographical Information Systems: Principles and Applications.

ICONTEC. Instituto Colombiano de Normas Técnicas y Certificación. 2002. NTC 4611 Información geográfica. Metadato geográfico. Bogotá D.C.

ICONTEC. Instituto Colombiano de Normas Técnicas y Certificación. 2003. NTC 5043 Información geográfica. Conceptos básicos de calidad. Bogotá D.C.

ICONTEC. Instituto Colombiano de Normas Técnicas y Certificación. 2003. NTC 5205 Precisión de datos espaciales –pde-. Bogotá D.C.

IGAC. Instituto Geográfico Agustín Codazzi. 1995. Nombres originales de los territorios, sitios y accidentes geográficos de Colombia (Topónimos). Bogotá D.C

Rudas, G., D. Armenteras, S. M. Sua y N. Rodríguez. (2002). Indicadores de Seguimiento de la Política de Biodiversidad en la Amazonia Colombiana – 2001. Informe Final de Resultados, Proyecto Diseño e Implementación del Sistema Indicadores de Seguimiento de la Política de Biodiversidad en la Amazonia Colombiana. Instituto Humboldt, CDA, Corpoamazonia, Cormacarena, Instituto Sinchi, Unidad de Parques, Ministerio del Medio Ambiente (Crédito BID 774 OC/CO), Bogotá. (114 páginas + 6 documentos anexos).

CONIABIO. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, 2003. Sistema de Información Biótica. México. Recuperado en Enero 15 de 2004 de http://www.conabio.gob.mx/informacion/biotica_espanol/doctos/acerca_biotica.html.

Hijmans, Robert J., Luigi Guarino, Coen Bussink, Israel Barrantes, y Edwin Rojas, 2002. DIVA-GIS, versión 2. Sistema de Información Geográfica para el Análisis de Datos de Biodiversidad. Manual. International Potato Center, Lima, Peru. Recuperado en Diciembre 11 de 2002 de http://www.cipotato.org/diva.

INBIO. Instituto Nacional de Biodiversidad, 1989. Tecnología informática al servicio de la conservación de la biodiversidad. Atta. Costa Rica. Recuperado en Enero 15 de 2004 de http://atta.inbio.ac.cr/atta03.html.

Kenneth E., Foote and Donald J., Huebner, 2000. Error, Accuracy, and Precision. The Geographer's Craft Project, Department of Geography, The University of Colorado at Boulder. Recuperado en Junio 24 de 2003 de www.colorado.edu/geography/gcraft/notes/error/error_f.html

Rivera-Gutiérrez, H. F., Suárez-Mayorga, A. M., Varón-Londoño, A., 2003. Estándar para la documentación de registros biológicos, versión 4.1 (electrónica). Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia, 56 p. Recuperado en Octubre 3 de 2003 de http://www.humboldt.org.co/sib/pdfs/regBiologicos/Estandar_Registros_Biologicos.pdf

Rivera-Gutiérrez H. F., Suárez-Mayorga A. M., Varón-Londoño A., 2003. Estándar para la documentación de metadatos de conjuntos de datos relacionados con biodiversidad, versión 1.0 (electrónica). Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia, 61 p. Recuperado en Octubre 3 de 2003 de http://www.humboldt.org.co/sib/pdfs/metadatos/EstandarMetadatos.pdf

USGS. United States Geological Survey, 1981. The Geographic Names Information System (GNIS). Recuperado en Junio 30 de 2001de http://geonames.usgs.gov/

BIOGRAPHICAL NOTES

Engineer cadastral and geodesist, Distrital University “Francisco Jose de Caldas”, Bogotá, Colombia, specialist in GIS, specialist in Management of projects of Engineering, Senior Researcher in Geographical Information Systems (GIS), Database administrator, Humboldt Biological Resources Research Institute, Colombia, GIS Consultant and design of implementation of data bases. At The World Bank (Washington D.C.), To support the construction of the line bases and to make I diagnose of resources and the atmosphere in amazonía, Amazonico institute of scientific researches (SINCHI), Colombia. GIS Consultant and field coordinator, to establish a model of pursuit to the Program of Aspersion and activities directed to the eradication of illicit cultures. Natioal narcotic direction, Colombia, work in the elaboration of standards of geographic metadata, design and implemetación of standards for handling of space information, among others.

CONTACTS

Rubén Darío Mateus Sanabria
Biological Records Georeferencing and Digital Localities Gazetteer
Instituto de Investigación de Recursos Biológicos Alexander von Humboldt
Carrera 7 No. 35 - 20
Bogotá D.C.
COLOMBIA
Tel. + 57 1 6086900
Email: rdmateus@humboldt.org.co
Web site: www.humboldt.org.co