# A UNIFIED APPROACH FOR SPATIAL DATA QUERY

Mohammed Abdalla [1], Hoda M. O. Mokhtar [2], Mohamed Noureldin [3]

[1, 2, 3] Faculty of Computers and Information, Cairo University, Giza, Egypt

### ABSTRACT

*With the rapid development in Geographic Information Systems (GISs) and their applications, more and more geo-graphical databases have been developed by different vendors. However, data integration and accessing is still a big problem for the development of GIS applications as no interoperability exists among different spatial databases. In this paper we propose a unified approach for spatial data query. The paper describes a framework for integrating information from repositories containing different vector data sets formats and repositories containing raster datasets. The presented approach converts different vector data formats into a single unified format (File Geo-Database "GDB"). In addition, we employ "metadata" to support a wide range of users' queries to retrieve relevant geographic information from heterogeneous and distributed repositories. Such an employment enhances both query processing and performance.*

### KEYWORDS

*Spatial data interoperability; GIS; Geo-Spatial Metadata; Spatial Data Infrastructure; Geo-database.*

## 1. INTRODUCTION

The need to store and process large amounts of diverse data, which is often geographically distributed, is obvious is a wide range of application. Most GISs use specific data models and databases for this purpose. This implies that making new data available to the system requires the data to be transferred into the system's specific data format and structure. However, this is a very time consuming and tedious process. Data accessing, automatically or semi-automatically, often makes large-scale investment in technical infrastructure and/or manpower inevitable. These obstacles are some of the motivations behind the concept of information integration. With the increase of location based services and geographically inspired applications, the integration of raster and vector data becomes more and more important [24]. In general, a geo-database is a database that is in some way referenced to locations on Earth [27]. Coupled with this data is usually data known as attribute data. At-tribute data are generally defined as additional information, which can then be tied to spatial data. GIS data can be separated into two categories: spatially referenced data, which is represented by vector and raster forms (including imagery); and attribute tables, which are represented in tabular format. Within the spatial referenced data group, the GIS data can be further classified into two different types: vector and raster. Most GIS applications mainly focus on the usage and manipulation of vector geo-databases with added components to work with raster-based geo-databases. Basically, vector and raster models differ in how they conceptualize, store, and represent the spatial locations of objects. The choice of vector, raster, or combined forms for the spatial database is usually governed by the GIS system in use and its ability to manipulate certain types of data. Nevertheless, integrated raster and vector processing capabilities are most desirable and provide the greatest flexibility for data manipulation and analysis. Many research papers discussed raster-vector integration as presented in [24, 25, and 26]. In real world applications, the effective management and integration of information across agency boundaries results in information being used more efficiently and

DOI : 10.5121/ijdkp.2013.3604

effectively [14]. Hence, developing interoperable platforms is a must. Several research work have been directed towards establishing protocols and interface specifications offering support for the discovery and retrieval of information that meets the user's needs [3]. In [1], the authors refer to spatial interoperability as the ability to communicate, run programs, or transfer spatial data between diverse data without having prior knowledge about data sources characteristics. Motivated by the importance of designing interoperable environments spatial data infra-structures (SDI) were developed. A spatial data infrastructure (SDI) is a data infrastructure implementing a framework of geographic data, metadata, users, and tools that interact to use spatial data in an efficient way [3]. Another definition for SDI was presented in [7], in this paper the authors define an SDI as the technology, policies, standards, human resources, and related activities necessary to acquire process, distribute, use, maintain, and preserve spatial data. In general, SDI is required to discover and deliver spatial data from a data repository, via a spatial service provider, to a user. The authors in [2] defined the basic software components of an SDI as (1) a software client: to display, query, and analyze spatial data (this could be a browser or a Desktop GIS), (2) a catalogue service: to discover, browse, and query metadata or spatial services, spatial datasets, and other resources, (3) a spatial data service: to allow the delivery of the data via the Internet, (4) processing services: such as datum and projection transformations, (5) a (spatial) data repository: to store data, e.g. a spatial database, and (6) a GIS software (client or desktop):to create and update spatial data. Beside these software components, a range of (international) technical standards are necessary that enable the interaction between the different software components. Another vital component of an SDI is the "metadata" which can be viewed as a summarized document providing content, quality, type, creation, and spatial information about a data set [8]. The importance of metadata in spatial data accessing, integration and management of distributed GIS resources was explored in several works including [18, 19, 20, 21, 22]. Metadata can be stored in any format including text file, Extensible Markup Language (XML), or database record. The summarized view of the metadata enhances data sharing, availability, and reduces data duplication. Inspired by the importance of developing an interoperable framework for spatial queries, in this paper we present an interoperable architecture for spatial queries that utilizes metadata to enhance the query performance. The proposed approach provides usage of modern and open data access standards. It also helps to develop efficient ways to achieve inter-operability including consolidation of links between data interoperability extensions and geo-graphic metadata.

The main contributions of the paper are summarized as follows:

- Developing an interoperable framework that converts the basic five vector data formats (AutoCAD DWG, File Geo- database, Personal Geo-database, Shape file, Coverage, and Geography Markup Language) into a single unified "gdb" format.
- Presenting an easy to use tool for searching at the feature data level of spatial vector data using metadata criteria.
- Using XML-metadata style for expressing the feature metadata, such representation is thus not restricted to a particular standard or profile.
- Improving the quality and performance of spatial queries by filtering the number of candidate results based on the features expressed in the metadata.
- GIS users face an opportunity and a challenge in manipulating and accessing the huge volume of data available from various GIS systems. The proposed approach can help making it easier for them to find, access, and use other data sets. It also helps them to easily advertise, distribute, reuse, and combine their data with other data sets.
- The proposed approach provides effective and efficient data management for processing heterogonous data. The power of the proposed model comes from integrating sources and displaying to the human eye the proximity-based relationships between objects of interest. Proximity can't be "seen" in the data, but it can be seen on a map.

The rest of the paper is structured as follows: Section 2 presents an overview of related work. Section 3 defines the problem. Section 4, presents our proposed solution and architecture. In section 5 we discuss the proposed system and the results achieved. In section 6 we discuss the analysis and testing of our implemented system. Finally, section 7 concludes and presents directions for future work.

## 2. RELATED WORK

The need for geo-data from distributed GIS sources is seen in many applications including decision making, location based services, and navigation applications. Integration of different data models, types, and structures facilitates cross-data set analysis from both spatial and non-spatial perspectives. This needs motivated several prior work on spatial data interoperability. In [4], a fuzzy geospatial data modelling technique for generation of fuzzy application schema is introduced. This approach aims to formalize the fuzzy model using description logic. The formalization facilitates automated schema mapping required for the integration process. In [5], service-based methodology has been discussed for integrating distributed geospatial data repositories in adherence to OGC specified open standards. The paper also describes the central role of a geographic ontology in the development of an integrated information system which are interoperable semantically, and utilizing it for service description and subsequent discovery of services. In [6], an important initiative to achieve GIS interoperability is presented, this is the OpenGIS Consortium. OpenGIS Consortium is an association looking to define a set of requirements, standards, and specifications that will support GIS interoperability. An approach for designing an integrated interoperability model based on the definition of a common template that integrates seven interoperability levels is proposed in [7].In addition, several work targeted SDI and Geo-Graphic metadata. Spatial data infrastructures (SDIs) are used to support the discovery and retrieval of distributed geographic information (GI) services by providing catalogue services through interoperability standards. A methodology proposed for improving current GI service discovery is discussed in [8]. When searching spatial data, traditional queries are no longer sufficient, because of the intrinsic complexity of data. As a matter of fact, parameters such as filename and date allow users to pose queries which discriminate among data solely on the basis of their organizational properties. In [9], a methodology for searching geographic data is introduced which takes into account the various aspects previously discussed. In [10], an approach to analyze geographic metadata for information search is introduced. In [11], the shortcomings of conventional approaches to semantic data integration and of existing metadata frameworks are discussed. On the other hand, the problem of vector and raster data integration was also investigated. Traditional techniques for vector to raster conversion result in a loss of information, the entities shape must follow the shape of the pixels. Thus, the information about the position of the entities in the vector data structure is lost with the conversion. In [12], an algorithm was developed to reconstruct the boundaries of the vector geographical entities using the information stored in the raster Fuzzy Geographical Entities. The authors utilize the fact that the grades of membership represent partial membership of the pixels to the entities, this information is thus valuable to reconstruct the entities boundaries in the vector data structure, generating boundaries of the obtained vector entities that are as close as possible to their original position. In [15], a new data model named Triangular Pyramid framework for enhanced object relational dynamic vector data model is proposed for representing the complete information required for representing the data for GIS based application. A spatial data warehouse based technique for data exchange from the spatial data warehouse is proposed in [13]. However, data warehouse based approach has several disadvantages keeping in mind the huge volume of data required to be updated regularly. Many of the problems associated with raster-to-vector and vector-to-raster conversion are discussed in [27]. In [23], the authors examine the common methods for converting spatial data sets between vector and raster formats and present the results of extensive benchmark testing of the proposed procedures. Also, in [16], many of the problems associated with raster-to-vector and vector-to-raster conversion are discussed. Raster maps are considered an important source of information. Extracting vector data from raster maps usually requires significant user input to achieve accurate results. In [17], an accurate road vectorization technique that minimizes user input is discussed; it aims to extract accurate road vector data from raster maps.

In this work we continue to explore possible approaches for vector and raster data integration to develop an efficient spatial data query tool.

## 3. PROBLEM DEFINITION

The quality of any geo-spatial information system is the main feature that allows system clients to fine-tune their search according to their specific needs and criteria. Nevertheless, disparate data sets exist in different geo-spatial databases with different data formats and models. Ac-cessing and integrating this heterogeneous data remains a challenge to efficiently answer user queries. In addition, with the increase in the GIS applications that are based on geographic information developing a unified approach for spatial query is a crucial requirement. Today, several formats exist for vector data including: AutoCAD DWG, File Geo-database, Personal Geo-database, Shape file, Coverage, and Geography Markup Language. Such diversity in data formats generates a problem in communication and data transfer between different data sources. In addition, geographical information may be stored using the vector or the raster data structure. The use of either structure depends on the methods used to collect the data and on the application that will use the information [12]. Also, such diversity in data models generates a problem in integration and data access operations between different data repositories.

Example 1: Consider 3 different data sources (DS1, DS2, and DS3) where each source stores the vector data in different format as shown in Figure 1.

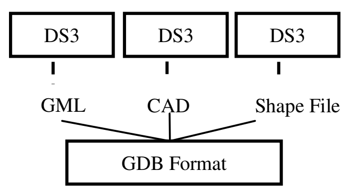


Figure 1. Querying different data sources

Assume a user query that requires data from all three sources. Such a query will require the user to physically pose three different queries to access the different formats. In addition, the user's query will eventually return different results in different formats. Motivated by the problem presented in Example 1, developing an interoperable platform is an optimal solution that unifies both the issued query and the query results. To achieve such operation, we need to convert the different spatial data formats (AutoCAD DWG, File Geo-database, Personal Geo-database, Shape file, Coverage, and Geography Markup Language, etc.) into a unified format. In this paper we select the File Geo-database format to be the final unified format.

Example 2: Consider two different data repositories with different data models (R1, R2). Assume that R1 has raster datasets and R2 has vector datasets as shown in Figure 2.

Assume a user query that requires data from both repositories regardless of data model representation.

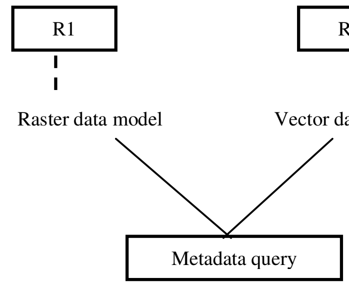


Figure 2. Querying different data Models

Using the same sources presented in Example 1, and issuing the same user query but assuming the existence of the required unified model, we then need to obtain a single unified query in "gdb" format. Again, motivated by the problem in Example 2, the query result still requires access to all repositories that have data in different models to retrieve all relevant data. Such access can be improved by understanding the query statement and filtering initial data to capture only relevant data. Such understanding and filtering process can be achieved using metadata.

## 4. PROPOSED SOLUTION

As discussed in Section 3, querying different spatial databases that store spatial data in various formats and models has a number of problems. In this paper we propose a new approach for spatial query processing and data accessing. The proposed architecture is composed of six main layers as shown in Figure 3.

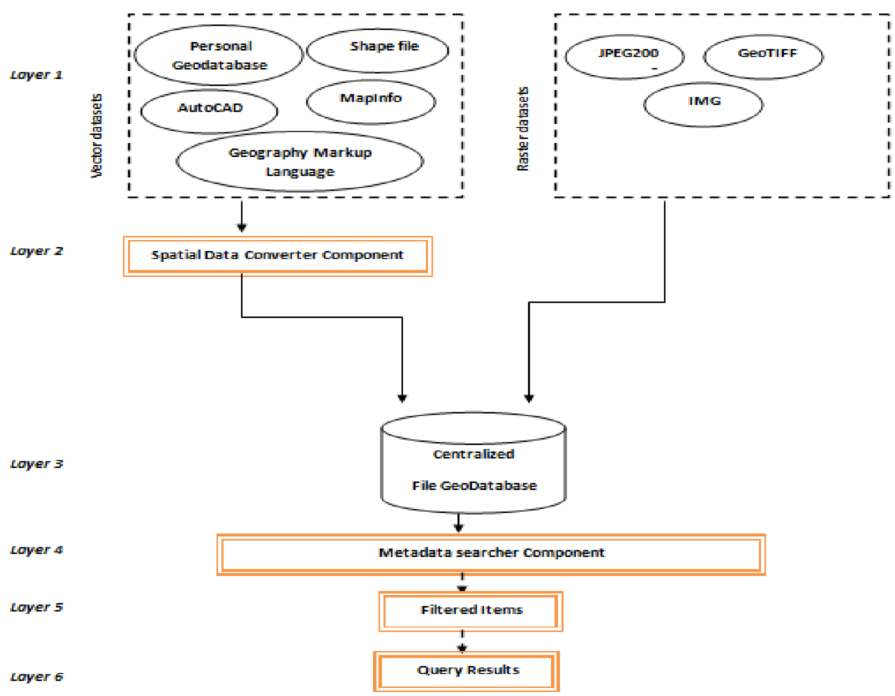


Figure 3. Proposed architecture

The first layer represents different data sources with different vector data formats (.shp, .mif, .cad, .gml, .mdb) and raster data formats. The second layer contains the spatial data converter component that is responsible for unifying the vector data formats. The third layer contains the resulting converted data in a single unified format. The fourth layer is the metadata searcher component that is responsible to find and access the most suitable datasets regardless of the initial data models and structures. The fifth layer contains the filtered items by the metadata component. And finally, the sixth layer contains the final user query results. The main characteristic of our proposed model is that we build a layer in our architecture that supports "interoperability" operations by developing a spatial data converter component that converts different spatial data formats (AutoCAD DWG, File Geo-database, Personal Geo-database, Shape file, Coverage, and Geography Markup Language) into a single format (File Geo-database "gdb" ).

Nevertheless, the top reasons for choosing the file Geo-database as our final unified format are:

- File geo-databases format is ideal for storing and managing geospatial data.
- File geo-databases format offers structural, performance, and data management advantages over personal geo- databases and shape files.
- Vector data can be stored in a file geo-database in a compressed, read-only format that reduces storage requirements.
- Storing raster in geo-database format manages raster data by subdividing it into small, manageable areas called tiles stored as large binary objects (BLOBs) in a database.
- File geo-databases format provides easy data migration.
- File geo-databases format is inclusive: one environment for feature classes, raster datasets, and tables
- File geo-databases format is powerful: enables modelling of spatial and attribute relationships.
- File geo-databases format is scalable: can sup-port organization-wide usage and workflows, and can be used with DBMS like Oracle, IBM DB2, and Microsoft SQL Server Express.

In addition our model has a layer that provides usage of modern and open data access standards, and helps to develop efficient ways to achieve inter-operability including consolidation of links between geo-graphic data interoperability extensions and geo-graphic metadata by developing a metadata searcher component that looks in repositories which have data in different spatial data models, structure, and formats and finds the most proper datasets. In the following discussion we present our proposed spatial data conversion algorithm.

**Algorithm 1: Spatial Data Converter**

**Input**: A different number of spatial databases with different vector data formats (GML, CAD, MIF, mdb, and shp).
**Output**: A different number of spatial databases with unified vector data format (File Geo-database).
**Begin**
  Get the path of the input file;
 Create an empty output file with the same name of the input file
  and replace extension with "gdb";
  Define a new GeoProcessor object;
**If**(data format "gml" or "cad" or "mif") **Then**{
  Define a quick import object;
  Set input file as input to QuickImport object;
  Set the created empty output file as output to
  quick import object;

```
   Pass QuickImport object to GeoProcessor ;
}
ElseIf(data format is “mdb”)   Then {
   Initialize a CopyTool;
   List all feature classes, data sets, and tables of the input file ;
Loop until no features, dataset, tables found
Begin
  Set the feature or dataset or table as an input to the CopyTool;
  Create the output path of the dataset or feature as the name of the
  Created output file and append to it the name of item;
  Set the item path as output to CopyTool;
  Pass CopyTool object to GeoProcessor;
End loop
{   ElseIf (data format is “.shp”)   Then}
   Define new Feature class object with the path of the shape file ;
   Define an Append object;
   Set input to Append object as feature class created from shape file;
   Set output to Append object the path of the created output gdb
   appended to it the name of feature class name;
}
EndIf
   Execute conversion using GeoProcessor;
End
```

By applying Algorithm 1 on the different data sources with different data formats in layer 1, we obtain in layer 3 a single unified data format and structure (File Geo-database “gdb”). The motivation behind choosing theses five formats for conversion is that these formats are very flexible in terms of the ability to mix all sorts of geometry types in a single dataset, openly documented, support geo-referenced coordinate systems, and are considered stable exchange formats. A successful conversion between (AutoCAD DWG, Map Info., Personal Geo-database, Shape file, Coverage, and Geography Markup Language) and File Geo-database format is done, considering the same shape size, origin and orientation, the same results are obtained. The areas occupied by entities inside the original file and the converted one are always the same. Then, in layer 4 motivated by the problem presented in Example 2, we developed a “Metadata Searcher” component as shown in Figure 4. The metadata searcher component defines some properties (for example: number of features, creation date, geographic form, feature name, and reference system), and searches in different data sources and Repositories for items that match those properties. The metadata feature selection component proceeds as follows.

**Algorithm 2: Metadata Feature Selection**

```
Input: A different number of spatial databases with unified vector data format (File Geo-
database “.gdb“)
Output: A collection of features that match metadata criteria.
Begin
  Define metadata search properties and values;
  Define the path that contains the converted data “GDB;“
  List all the converted gdb files
  Loop until no files found
  Begin
    Loop FOR EACH features and datasets in gdb file
    Begin
      If item matches defined metadata properties and values Then
```

```
        Add item to filtered item list
      End If
    End loop
  End loop
End
```

We apply Algorithm 2 in layer 4 in our proposed architecture on a different number of spatial databases with unified ".gdb" format and raster datasets. Then, for every data source the algorithm searches for the features and data elements that match the metadata search criteria, and save the selected items in the list of filtered items that eventually contribute towards the user query result.

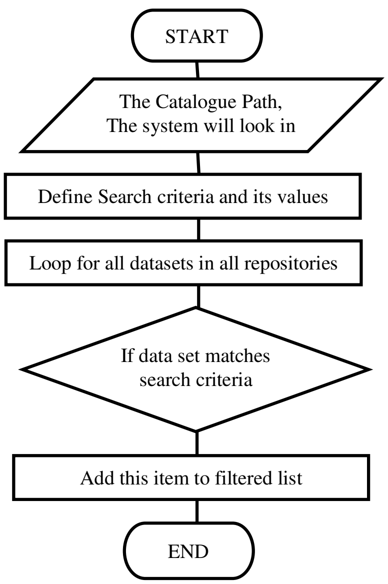


Figure 4. Metadata Searcher Component Flow Chart

**Algorithm 3: Raster Query**

```
Input: Raster dataset.
Output: Raster Result set.
Begin
  Create the RasterExtractionOp object.
  Declare the Raster input object.
  Declare a RasterDescriptor object
  Select the field used for extraction Using RasterDescriptor
  Set RasterDescriptor as an input to RasterExtractionOp object.
  Execute Query using RasterExtractionOp object.
  Save the Results in new Geodataset.
End
```

Next, layer 5 maintains the filtered items resulting from the different data sources that match the specific metadata properties and is ready to receive user query. The filtered raster dataset will be queried by applying Algorithm 3 and filtered vector datasets will be queried either by Sptial data qurey functions or attributte data statements. Finally, layer 6 contains the actual combined user query results that composed of raster and vector datasets against the filtered items that are then presented to user.

## 5. RESULTS AND DISCUSSIONS

In this paper we present a holistic approach to unify spatial data query schemas. Various data accessing and metadata management steps have been used and subsequently employed to contribute towards designing a framework for efficiently answering spatial data query. In our design we focused on the following features that the proposed system satisfies:

- Easy to access geospatial data repositories and retrieving data in transparent way. The file Geo-database "gdb" format was chosen in our model for reasons discussed before in section 4.

- Developing an interoperable framework that links both semantic interoperability and syntactic interoperability is a promising scenario for deriving data from multiple sources with different data formats and models.

- Metadata descriptions adopted in the proposed system are not reliant up on specific profile or standard. XML-based metadata was chosen to ensure flexibility for discovering resources and features.

Taking those constraints into consideration, we built an easy to use tool that unifies different vector formats into a single "gdb" format, accesses different spatial data models (Raster and Vector) repositories, and processes user queries using spatial metadata that helps to enhance the query performance. Figure 5 and Figure 6 show the initial input to the system where data is presented in different spatial formats and models. This initial format is then unified as shown in Figure 7.

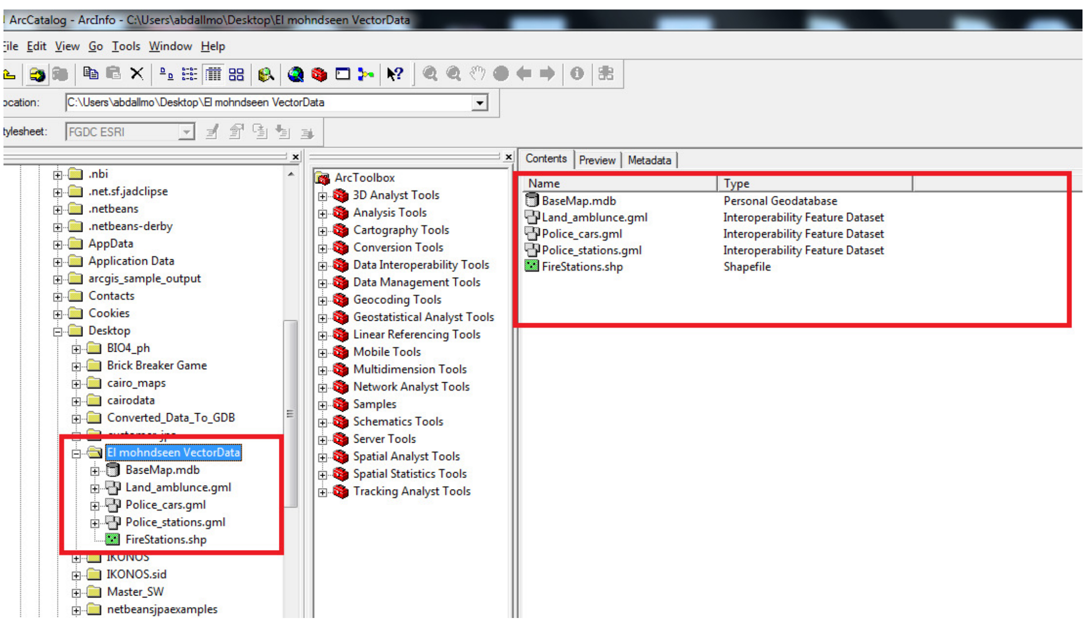


Figure 5.Vector Data before applying spatial data converter

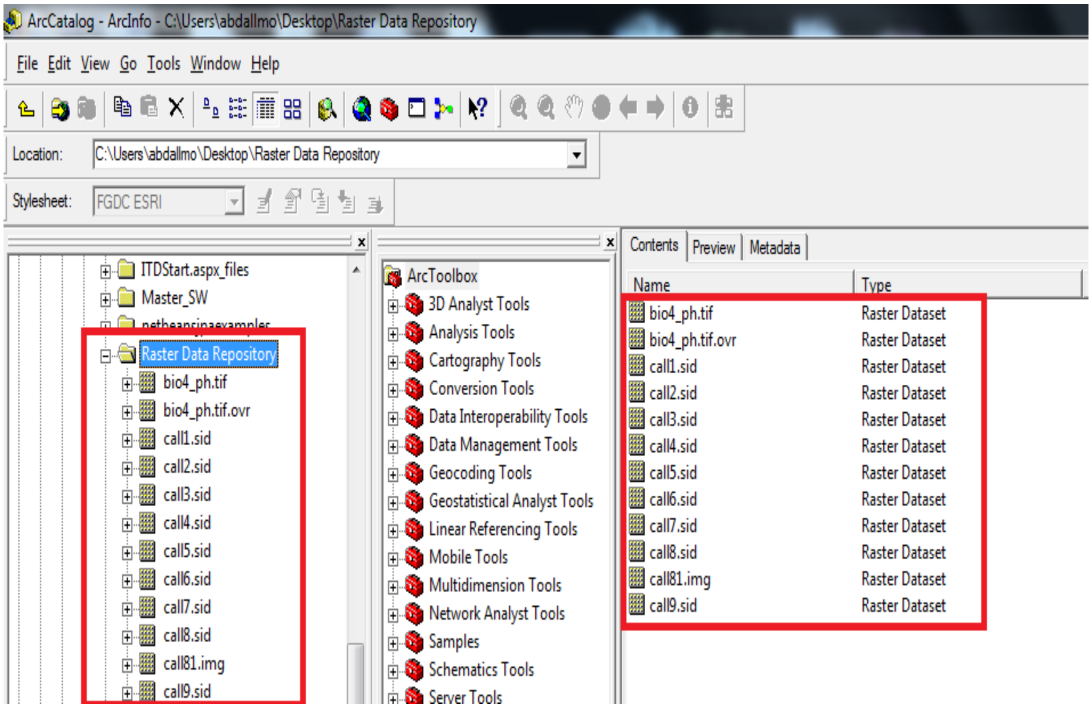

Figure 6. Raster Data Repository

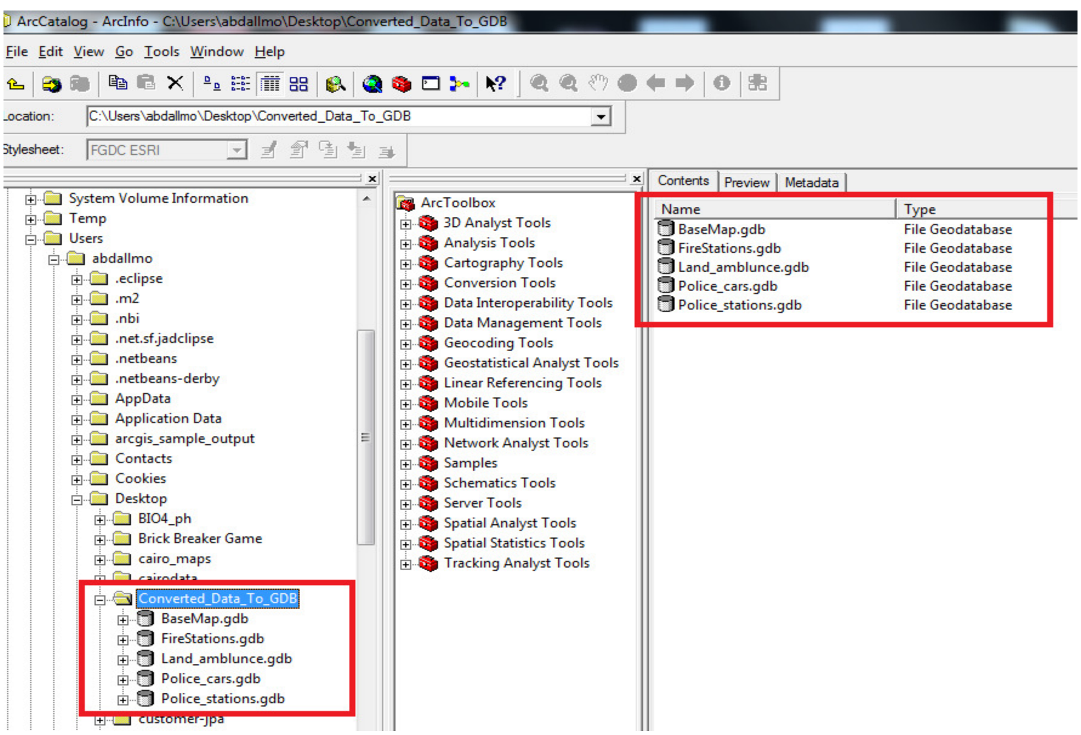


Figure 7. Unifed "GDB" Format

Once the data is unified, the system starts processing spatial queries. It accepts the criteria defined by the user that constrain the required output. Those constraints along with the metadata help to locate the candidate data in different files. For instance some users are interested in files that have specific number of features, specific creation date, or feature name that start with specific pattern, or contains specific pattern. Augmenting metadata in the system allows the user to select all the criteria he needs, and search in the catalogue path to locate matching data sets and feature classes.

Example 3: Consider a MQ1 (Metadata Query) with the following selection criteria as shown in Figure 8:

Data Representation *equals* vector digital data, Feature Name *contains* Streets, Feature Count *greater than* 180, East bounding coordinate *equals* 31.219267, Data Form Value *equals* File Geo-database Feature Class, Creation Date *equals* 20121118, and Reference System *equals* WGS_1984_UTM_Zone_36N

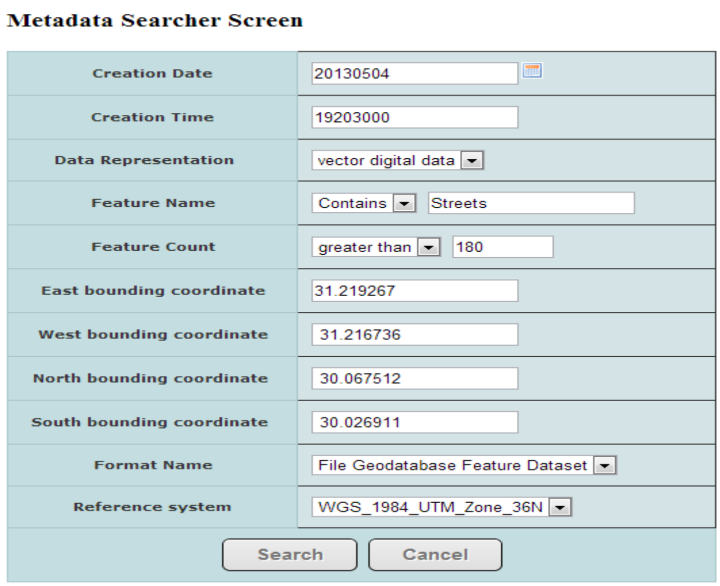


Figure 8. Metadata Searcher Screen.

After metadata data query results are retrieved, the user has the ability for selecting features from single or multiple vector feature classes and datasets retrieved. For single feature class the user poses vector attribute data query based on specific values and selected criteria (VQ1), for multiple feature classes and datasets the user poses spatial vector data query based on selected topological relation between features and values used in selected features buffering.

For VQ1 (vector attribute data query) the user can also specify the values associated to each feature as follows: "*Ename ≠ 'NULL', Width > 15, Shape_length >200 and METERS > 0*"

For VQ2 (vector spatial data query) the user can also specify the topological relation and values associated to buffer features as follows: "*Select features from "Fuel_Stations" are within a distance of "Buildings" with a buffer to features in buildings of 190.000000 Meters*".

Using the sample dataset shown in Figure 7, the system will retrieve three feature classes that match the user specified criteria and values as shown in Figure 8.

Attributes of street

| OBJECTID | FNODE_ | TNODE_ | LPOLY_ | RPOLY_ | LENGTH | GAMAL_ST_ | GAMAL_ST_I | ENAME | WIDTH | METERS | State | Shape_Length |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 891 | 5974 | 4762 | -1 | -1 | 1633.1182 | 4750 | 1593 | El-Sodan St. | 82 | 0 | 2.064423e+257 | 1633.11826 |
| 109 | 2976 | 2912 | -1 | -1 | 1452.7500 | 2861 | 426 | Ahmed Oraby St. | 73 | 0 | 4.746832e+250 | 1452.750053 |
| 1378 | 2825 | 3521 | 0 | 0 | 1389.2641 | 11375 | 785 | Teeret Al-Zomor | 69 | 0 | 5.437559e+256 | 1389.264104 |
| 205 | 3504 | 2828 | -1 | -1 | 1346.8636 | 3123 | 3873 | El-Sodan St. | 67 | 0 | 8.281353e+269 | 1346.863662 |
| 2566 | 2838 | 13549 | 0 | 0 | 1132.0905 | 17590 | 1 | Gamal Abd El-Naser -Kornish El-Nile- st. | 57 | 0 | 1.473003e+271 | 1132.090511 |
| 2604 | 13662 | 5979 | 0 | 0 | 1084.7631 | 17924 | 570 | Al-Sodan st. | 54 | 0 | 6.418636e+264 | 1084.763183 |
| 1570 | 4809 | 4483 | 0 | 0 | 974.2568 | 16356 | 506 | Al-Gabalaiah St. | 49 | 0 | 1.402003e+263 | 974.256803 |
| 237 | 3737 | 3173 | -1 | -1 | 951.32181 | 3223 | 1 | Gamal Abd El-Naser -Kornish El-Nile- st. | 48 | 0 | 2.769939e+263 | 951.321814 |
| 2603 | 5986 | 6581 | -1 | -1 | 858.72792 | 17922 | 1684 | Al-Sodan st. | 43 | 0 | 3.550931e+262 | 858.727916 |
| 565 | 4149 | 4844 | -1 | -1 | 780.62579 | 3889 | 1 | Al-Nile st. | 39 | 0 | 6.214419e+260 | 780.625788 |
| 2218 | 3521 | 13458 | 0 | 0 | 773.21987 | 17004 | 4068 | Teeret Al-Zomor | 39 | 0 | 8.508911e+257 | 773.219868 |
| 1366 | 5929 | 5417 | 0 | 0 | 766.94344 | 9713 | 1399 | Abdel Aziz Aal Sooud St. | 38 | 0 | 4.979650e+276 | 766.943444 |
| 2217 | 13457 | 3508 | 0 | 0 | 769.78622 | 17003 | 1 | El-Sodan St. | 38 | 0 | 2.332587e+258 | 769.786221 |
| 56 | 2769 | 2600 | -1 | -1 | 694.72152 | 2729 | 34 | El-Matar St. | 35 | 0 | 5.309502e+241 | 694.72152 |
| 874 | 5774 | 5890 | -1 | -1 | 686.2657 | 4694 | 1417 | Tharwat St. | 34 | 0 | 6.245687e+264 | 686.265695 |
| 350 | 4175 | 4219 | -1 | -1 | 678.34309 | 3468 | 1 | Al-Borg st. | 34 | 0 | 3.878907e+275 | 678.34309 |
| 1384 | 4186 | 3688 | 0 | 0 | 666.38592 | 11381 | 4067 | Teeret Al-Zomor | 33 | 0 | 1.282313e+249 | 666.385923 |
| 2195 | 5412 | 5873 | 0 | 0 | 627.78509 | 16981 | 1631 | Teerat El-Zomor st. | 31 | 0 | 1.869965e+263 | 627.785091 |

Record: 12 Show: All Selected Records (18 out of 2613 Selected) Options

Figure 9. Vector Attribute table


Figure 10. Vector Query Result

Those matching classes are retrieved based on the metadata used in the query. The final results are then displayed or presented to the user as shown in Figure 9 and Figure 10.Motivated by Example 3, assume that user interested to find all datasets in all repositories regardless of data representation model that have the following criteria "East bounding coordinate equals 31.219267" the User to AND the Query appear in Example 3 with the another one in Example 4 to find and access all required datasets.

Example 4: Consider a MQ2 (Metadata Query) a user change query selection criteria to be:

"Data Representation *equals* raster digital data; Feature Name *contains* call, east bounding coordinate *equals* 31.219267, Data Form Value *equals* Raster Dataset, Creation Date *greater than* 20121220, and Reference System *equals* "IMAGINE GeoTIFF ERDAS, Inc. Al"

After metadata data query results are retrieved the user has the ability to query Raster data using the cell value. To query a grid, the user has to use a logical expression such as RQ1: [Count] >700 AND [Temp_C]>=40.34. It is also possible to query multiple grids by cell value.

| OID | VALUE | COUNT |
|---|---|---|
| 0 | 189 | 742 |
| 1 | 225 | 847 |
| 2 | 229 | 873 |
| 3 | 236 | 818 |
| 4 | 243 | 802 |
| 5 | 253 | 782 |
| 6 | 266 | 727 |
| 7 | 272 | 709 |
| 8 | 278 | 790 |
| 9 | 281 | 737 |
| 10 | 284 | 707 |
| 11 | 290 | 725 |
| 12 | 298 | 827 |
| 13 | 301 | 764 |
| 14 | 304 | 825 |
| 15 | 309 | 1048 |
| 16 | 312 | 943 |

Figure 11. Raster Attribute table

Figure 12.Raster Query result

Figure 13.Final Query results Integrated Map.

According to Example 4, the sample dataset is shown in Figure 11and Figure 12, the system will retrieve one dataset that matches the user specified criteria and values. This matched dataset is retrieved based on setting values used in Example 4 query. The final results are then displayed or presented to the user as shown in Figure 13.

## 6. QUANTIFIABLE ANALYSIS AND TESTING

To clarify our justifications of using centralized file geo-database as a back end geospatial data store, and for linking geographic metadata with data interoperability extensions, we proposed a platform connecting different data sources and formats for implementing a unified approach for spatial data query. A framework example was also implemented and tested. In this section we investigate the design and features of the implemented system. Based on our previous discussion, in this framework we develop two main components namely, a spatial data converter, and a metadata searcher. In addition, we also developed the basic operations performed by those two components as discussed earlier. The main characteristic of those developed operations is that they hide implementation details from the user providing him with a transparent communication with the system.

Following the architecture proposed in [1], our proposed system architecture is composed of four layers; presentation layer, business logic layer, data access layer, and data management tier. The function of each layer is as defined in [1]. Flyweight and façade design patterns were used for implementing the four layers mentioned above [28] [29]. Were the system starts with the user inputting a physical location path for the spatial dataset. Then, the spatial data irrespective to its original format is converted using the spatial data converter into the unified GDB format. Once the unified data is ready, the user is requested to input the metadata search criteria and parameters. Finally, based on user requests, the metadata searcher component retrieves the results from the unified geo-database and returns the results to the user.

**Performance Test:** The proposed framework was also tested using random features of sizes: 5000, 10000, 50000. Those features were first inserted and integrated into the centralized geo-database along with their associated geo-graphic view and attribute tables. Then, to evaluate the performance two queries were designed and posed against the system. We used the test queries to test our proposed framework. The first query (Q1) aims to retrieve raster datasets and performs “raster query by attribute” against result set. The other one (Q2) aims to retrieve vector feature classes and then perform “Vector attribute Query” against result set.

For both queries, we measured the average run time and used it as a metric for evaluating the performance. Tables 1 and 2 present the results obtained from both queries.

Table1. Performance test for retrieving features (Q1)

| **Number of Features** | 5000 | 10000 | 50000 |
|---|---|---|---|
| **Number of features Retrieved** | 178 | 231 | 343 |
| **Time(Milliseconds)Retrieving and manipulating data with implemented system** | 195 ms | 210 ms | 350 ms |
| **Time(Milliseconds)Retrieving and manipulating data without implemented system** | 230 ms | 360 ms | 500 ms |

Table2. Performance test for retrieving features (Q2)

| **Number of Features** | 5000 | 10000 | 50000 |
|---|---|---|---|
| **Number of features Retrieved** | 103 | 189 | 243 |
| **Time(Milliseconds)Retrieving and manipulating data with implemented system** | 60 ms | 198 ms | 220 ms |
| **Time(Milliseconds)Retrieving and manipulating data without implemented system** | 105 ms | 230 ms | 380 ms |

The results displayed above show that the proposed solution is an efficient solution for retrieving and manipulation of spatial data.

## 7. CONCLUSIONS AND FUTURE WORK

Efficiency of the planning system needs accessible, affordable, adequate, accurate and timely spatial and Non-spatial information. Information integration and sharing in turn needs an efficient route that can give possible access to the needy. The potential route can be achieved and accessed through the implementation of a well structured interoperable approach towards good information management. This paper introduces the issues of data interoperability, advantages of Geo-Graphic metadata, and its mechanism for data interoperability. In this paper we proposed an interoperable framework for spatial data query. Developing spatial data converter component which enables the proposed framework to accept vector data in various formats and unifies them into a single "gdb" format, which can be integrated with different raster datasets. GDB format can give users the capability to easily and dynamically publish and exchange data in an open, non-proprietary industry-standard format, thus maximizing the re-use of geospatial data, eliminating time-consuming data conversion and reducing associated costs. The resulting files are then input to a metadata selection component that uses the spatial features metadata to answer the user queries more efficiently. For future work we plan to extend our work to consider raster data in order to present a complete interoperable platform for spatial data. We also think that testing the system on various queries can strengthen our work. based on the search results we still need to develop a "ranking component" based on data mining techniques that is able to integrate with our proposed model, to sort results based on the importance of information value to the user is must. Finally, the current proposed approach still cannot solve the problem of semantic interoperability, investigating this point can be a good point for future work.

## REFERENCES


[1] Amirian, Pouria, and Ali A. Alesheikh (2008). "Implementation of a Geospatial Web service Using Web Services Technologies and Native XML Databases". Middle-East Journal of Scientific Research Vol. 3, No. 1, pp. 36-48.

[2] D. D. Nebert, Developing Spatial Data Infrastructures: The SDI Cookbook. GSDI: Global Spatial Data Infrastructure Association, 2004.

[3] P. B. Shah and N. J. Thakkar.Geo-spatial metadata services - ISRO's initiative, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.Vol. XXXVII. Part B4. Beijing 2008.

[4] Mukherjee, Soumya K. Ghosh, Formalizing Fuzzy Spatial Data Model for Integrating Heterogeneous Spatial Data, 2nd International ACM Conference on Computing for Geospatial Research & Applications (COM.GEO 2011), ACM, 25:1--25:6, Washington, DC, 23-25 May, 2011.

[5] Manoj Paul, S. K. Ghosh, A Framework for Semantic Interoperability for Distributed Geospatial Repositories, Journal of Computer and Informatics, Special Issue on Semantic e-Science, Vol. 26, 73-92, 2008.

[6] Open Geospatial Consortium. http://www.opengis.org, accessed on May 2013.

[7] Manso, M. A.; Wachowicz, M.; Bernabé, M. A.: "Towards an Integrated Model of Interoperability for Spatial Data Infrastructures". Transactions in GIS, Vol. 13, No.1. 2009, pp. 43-67.

[8] Luca Paolino, Monica Sebillo, Genoveffa Tortora, Giuliana Vitiello: Searching geographic resources through metadata-based queries for expert user communities. GIR 2007: 83-88

[9] Michael Lutz, Ontology-based service discovery in spatial data infrastructures. Geographic Information Retrieval conference, 2005.

[10] R Albertoni, A Bertone, M De Martino, Visualization and semantic analysis of geographic metadata. Proceedings of the 2005 workshop on Geographic information retrieval, 9-16.

[11] Nadine Schuurman, and Agnieszka Leszczynski.,Ontology-Based Metadata ,T. GIS 10(5):709-726 (2006)

[12] Caetano, M. and Painho, M. (eds). Conversion between the vector and raster data structures using Fuzzy Geographical Entities. Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 5 – 7 July 2006, Lisboa, Instituto Geográfico Português.

[13] Roth, M.T, Wrapper Architecture for Legacy Data Sources. Proceedings of the International Conference on Very Large Databases (23rd VLDB), 1997.

[14] Abbas Rajabifard, Data Integration and Interoperability of Systems and Data.2nd Preparatory Meeting of the Proposed UN Committee on Global Geographic Information Management, 2010.

[15] Ms. Barkha Bahl1, Dr. Navin Rajpal and Dr. Vandana Sharma, Triangular Pyramid Framework For Enhanced Object Relational Dynamic Data Model for GIS, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011.

[16] Russell G. Congalton , Exploring and Evaluating the Consequences of Vector-to-Raster and Raster-to-Vector Conversion, Photogrammetric Engineering & Remote Sensing, Vol. 63, No. 4, April 1997, pp. 425-434.

[17] Chiang, Y.; and Knoblock, C. A. Extracting Road Vector Data from Raster Maps. In Graphics Recognition: Achievements, Challenges, and Evolution, Selected Papers of the 8th International Workshop on Graphics Recognition (GREC), Lecture Notes in Computer Science, 6020, pp. 93-105. Springer, New York, 2009.

[18] In-Hak Joo; Tae-Hyun Hwang; Kyung-Ho Choi, "Generation of video metadata supporting video-GIS integration," ICIP '04. 2004 International Conference on Image Processing, 2004, vol.3, pp.1695-1698, 24-27 Oct. 2004.

[19] Tae-Hyun Hwang; Kyoung-Ho Choi; In-Hak Joo; Jong-Hyun Lee, "MPEG-7 metadata for video-based GIS applications," Proceedings of IEEE International Geoscience and Remote Sensing Symposium, IGARSS '03, pp.3641-3643, vol.6, 21-25 July 2003.

[20] Goncalves Soares Elias, V.; Salgado, A.C., "A metadata-based approach to define a standard to visual queries in GIS," Proceedings of the 11th International Workshop on Database and Expert Systems Applications, pp.693-697, 2000.

[21] Yingwei Luo; Xiaolin Wang; Zhuoqun Xu, "Extension of spatial metadata for navigating distributed spatial data," Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS '03, vol.6, pp.3721- 3723, 21-25 July 2003.

[22] Spery, L.; Claramunt, C.; Libourel, T., "A lineage metadata model for the temporal management of a cadastre application," Proceedings of the 10th International Workshop on Database and Expert Systems Applications, pp. 466- 474, 1999.

[23] Joseph M. Piwowar, Ellsworth F. Ledrew, Douglas J. Dudycha. Integration of spatial data in vector and raster formats in a geographic information system environment. International Journal of Geographical Information Science 01/1990; 4, pp. 429-444.

[24] Feng Lin; Chaozhen Guo, "Raster-vector integration based on SVG on mobile GIS platform," 2011 6th International Conference on Pervasive Computing and Applications (ICPCA), pp. 378-383, 26-28 Oct. 2011.

[25] Changxiong Wang; Shanjun Mao; Mei Li; Huiyi Bao; Ying Zhang, "Integration of vector and raster-based coal mines surface and underground contrast map," Second International Workshop on Earth Observation and Remote Sensing Applications (EORSA), pp. 309-312, 8-11 June 2012.

[26] Xuefeng Cao; Gang Wan; Feng Li, "Notice of Violation of IEEE Publication Principles 3D Vector-Raster Data Integration Model Based on View Dependent Quadtree and GPU Friendly Rendering Algorithm," International Joint Conference on Computational Sciences and Optimization, CSO 2009, Vol. 2, pp.244-247, 24-26 April 2009.

[27] Russell G.Congalto, Exploring and Evaluating the Consequences of Vector-to-Raster and Raster-to-Vector Conversion, Photogrammetric Engineering & Remote Sensing, Vol.63, No.4, April 1997, pp.425-434.

[28] Horner, M., 2006. Pro.NET 2.0 Code and Design Standards in C#. California, USA, APress Publishing.

[29] O'Docherty, M., 2005. Object-oriented analysis and design: Understanding system development with UML 2.0., New Jersey, USA, John Wiley and Sons, Inc.

## AUTHORS

**Mohammed Abdalla** is a Software Engineer of Software Development at Hewlett-Packard Enterprise services, Inc. Mr. Mohammed have +5 years experience in software development and analysis, with previously experiences in developing ERP, E-commerce, mobile payment and e-payment applications, In June 2008 earned a Bachelor of computer Science degree from the Cairo university with grade very good and excellent grade in graduation project.

**Dr. Hoda M. O. Mokhtar** is currently an associate professor in the Information Systems Dept., Faculty of Computers and Information, Cairo University. Dr. Hoda Mokhtar received her PhD in Computer science in 2005 from University of California Santa Barbara. She received her MSc. and BSc. in 2000 and 1997 resp. from the Computer Engineering Dept., Faculty of Engineering - Cairo University. Her research interests are database systems, moving object databases, data warehousing, and data mining.