Database Design of a Geo-environmental Information System

George Roumelis[1], Thanasis Loukopoulos[2] and Michael Vassilakopoulos[2]
[1]Dept. of Informatics, Aristotle University, Thessaloniki, Greece
[2]Dept. of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece

Keywords: Geographical Information Systems, Environment, Database Modelling, Entity-Relationship Model, Database Queries, Database Indexing.

Abstract: Environmental protection from productive investments becomes a major task for enterprises and constitutes a critical competitiveness factor. The region of Central Greece presents many serious and particular environmental problems. An Environmental Geographic Information System is under development that will maintain necessary and available information, including existing environmental legislation, specific data rules, regulations, restrictions and actions of the primary sector, existing activities of the secondary and tertiary sectors and their influences. The system will provide information about the environmental status in each location with respect to water resources, soil and atmosphere, the existence of significant pollution sources, existing surveys, studies and measurements for high risk areas, the land use and legal status of locations and the infrastructure networks. In this paper, we present a Database Design that supports the above mentioned objectives and information provision. More specifically, we present examples of user queries that the system should be able to answer for extraction of useful information, the basic categorization of data that will be maintained by the system, a data model that is able to support such data maintenance and examine how existing indexing structures can be utilized for efficient processing of such queries.

1 INTRODUCTION

Environmental protection from productive investments becomes a major task for enterprises and constitutes a critical competitiveness factor. At the same time, the viability depends on the growth capability of all production sectors as well as on the guarantee that the environmental effects will not have a major social impact. The region of Central Greece, presents many serious, several and particular environmental problems. The main pollution sources and environmental degradation factors are:

- Many industrial, mining and energy units,
- Non-organized residential expansion,
- Absence of urban waste management,
- Absorption of aquatic resources by Attica basin.

As a result several problems arise such as the qualitative degradation of the primary sector, the limited exploitation of tourism resources and the potential ecological collapse of many sea and terrestrial areas. The lack of rational planning, spatial organization and operation control of the human activity effects is the root of the diverse and complex environmental problems.

In this context and especially for the environmental problems in Central Greece, the establishment of an Environmental Geographic Information System (EGIS) (Gandhi et al., 2009) is proposed which will include a multi-level software with all necessary and available information (Wainwright and Mulligan, 2013), including:

- Total existing legislation (Laws, specific restrictions, decisions, guidelines for national, regional and local planning, land use, NATURA areas, specific environmental studies, etc.) concerning the region environmental issues.
- The specific rules, regulations, restrictions and actions concerning the primary sector activities as well as the special conditions or requirements which are necessary for the operation of any infrastructure (national and local) in this geographical region.
- The existing activities of the secondary and tertiary sectors and their produced geographical influences. Towards this direction, it is

Roumelis G., Loukopoulos T. and Vassilakopoulos M..
Database Design of a Geo-environmental Information System .
DOI: 10.5220/0004952603750382
In *Proceedings of the 16th International Conference on Enterprise Information Systems* (ICEIS-2014), pages 375-382
ISBN: 978-989-758-028-4

necessary to take into account data from other works designed for specific sections of the region (e.g. Environmental Registry for the secondary sector, which has been prepared by certain prefectures).

The main objectives of the EGIS development are:

- The actual environmental impact assessment of spatial planning, sector policies and infrastructure, as well as the selection of appropriate solutions by researchers and inspecting mechanisms.
- The valid information about the existing restrictions on the area and technology choices appropriate for the interested investors and social services.
- The elected bodies (municipal, district councils etc.) and authorities' decisions for investments implementations.
- More efficient processes for the relevant authorities' audits and management.
- Planning policies for environmental protection and enhancement.
- Management studies and research preparation for environmental protection.

The system will provide information about:

- The environmental status in each location with respect to water resources, soil and atmosphere.
- The existence of significant point (e.g. heavy industries, aquaculture) or diffused (e.g. informal industrial concentration, mining, fish farms) pollution sources.
- The existing surveys, studies and measurements for the high risk areas under examination.
- The land use and legal status to any location.
- The infrastructure networks.

In this paper, we present a Database Design that supports several of the above mentioned objectives and information provision. In Section 2, we present examples of user queries on such data that the system should be able to answer, for extraction of useful information. In Section 3, we present the basic categorization of data that will be maintained by the system. In Section 4, we present the data model of a Database that is able to support such data maintenance. In Section 5, we examine how existing indexing structures can be utilized for efficient processing of such queries. In Section 6, we summarize the contribution of this work and discuss future research directions.

2 USER QUERIES

After discussing with environmental-management and spatial-planning specialists, a number of example user queries that the EGIS should be able to answer were gathered. These queries, in general, consist of a series, or combination of conventional (including conditions on 1-d data) and spatial (including conditions on 2-d / 3-d data) database queries. Processing of spatial queries is more demanding than that of conventional ones (Corral and Vassilakopoulos, 2009). A representative list of such queries is presented in the following.

- Identify areas where logging activity can take place, in combination with the restrictions of reforestable geographical units.
- Identify coastal sites where aquaculture activities could seamlessly develop, under possible restrictions, due to various pollution burdens of the marine environment that evolved during the last 10 years .
- Which are the statutory quarrying areas, in the Region of Central Greece .
- Identify sites where there is possibility of operating oil refineries, given the current restrictions on air pollutants or disposal of generated waste.
- Identify sites where marinas of tourist vessels, or/and identify hinterland sites where heliports or airports of small private aircrafts can be operated, according to the National and Regional Spatial Framework.
- Identify sites where the installation and operation of antennas for mobile telephony is allowed based on the related regulatory framework.
- Identify sites suitable for installation and operation of a wind farm, according to the existing spatial planning.
- Show the production units in the Region of Central Greece that have annual assets exceeding X€ and operate outside of organized industrial areas and the conditions of environmental protection that were set for their operation.

3 DATA CATEGORIZATION

Based on the EGIS requirements and objectives presented in the Introduction and by analysing the detailed list of user queries and related regulatory frameworks and by repetitively interviewing environmental-management and spatial-planning specialists, a categorization of the data collection to be maintained was reached. In summary, the EGIS should maintain data for the:

Natural Environment,
- protected areas (Natura/Ramsar conventions, national parks, areas of outstanding natural beauty, game and wildlife reserves), natural monuments and archaeological/ historical sites, natural processes data (rainfall, earthquakes, disasters and landslides), special cases of negative impacts (pollution, permanent environmental nuisances), land use (Corine 2000 programme, wind-measurement/ insolation data, hydrographic network), and the

Anthropogenic environment,
- industry/artisanship facilities, interventions in coastal areas, quantitative characteristics of point/linear pollution sources, quantitative characteristics and performance calibration of biological treatment plants, areas and installations of livestock facilities, air pollution, tourism facilities, residential centres, spatial plans of residential organization and General Urban Plans, urban perspectives and street plans, town planning and perspectives, population and qualitative characteristics, major construction projects, transport facilities, ports.

The basic elements of the data categorization in more structured and detailed manner is presented in the following.

A. Primary Sector

1. Agricultural crops
- Classification - characterization
- Infrastructure: networks mapping
- Adequacy / quality of water resources
- Possible impacts

2. Livestock / Poultry units

A. Specific information
- Location, area – property, starting year, type
- Method
- Licenses
- Environmental terms and possible impacts
- Impacts ascertained

B. General information
- Incompatibilities with other uses
- Reports and studies on the impacts

3. Aquaculture

A. Specific information
- Location
- Area, property
- Starting year
- Type, capacity
- Licenses
- Environmental terms and possible impacts
- Impacts ascertained

B. General information
- Incompatibilities with other uses
- Directions of spatial planning for Aquaculture
- Reports and studies on the impacts

B. Secondary Sector

1. Production Units

A. Specific information
- Location, area – property, starting year, type of activity
- Licenses
- Environmental terms and possible impacts
- Impacts ascertained
- Other complaints and results

B. General information
- Incompatibilities with other uses
- Reports and studies on the impacts

2. Energy units and Renewable Energy Sources (RES)
- Units in operation
- Units licensed
- Units applied for license

3a. Mining activities: mining, mineral researches, mineral rights

3b. Quarrying activities: quarrying areas, quarries

A. Specific information
- Location, area of activity, ownership, starting year
- Stage:
  a) active
  b) inactive
  c) with definite concession
  d) under licensed mineral exploration
- Mining type: Surface / underground
- Restorations made
- Restorations provisioned
- Installation location photos
- Licensing (start/end, renewals)
- Environmental terms and possible impacts
- Impacts ascertained

B. General information
- Incompatibilities with other uses
- Guidelines by national and regional spatial plans
- Reports and studies on the impacts

C. Tertiary Sector and Infrastructure

1. Water management

A. Specific information
- Location
- Licenses
- Environmental terms and possible impacts

B. General information
- Incompatibilities with other uses
- reports and studies on the impacts

2. Waste management

A. Specific information
- Location, starting year , provisioned duration, settlements/population serviced, area, description, annual capacity, competent authority
- Licenses
- Environmental terms and possible impacts
- Impacts ascertained

B. General information
- Regional planning for waste management
- Incompatibilities with other uses
- Possible impacts

3. Liquid waste management

A. Specific information
- location, starting year, area, equivalent population, description, management body
- Licenses
- Impacts ascertained

B. General information
- Incompatibilities with other uses
- Possible impacts

4. Energy networks
- Mapping, schematizing / dimensioning of natural gas pipelines, power transmission lines characterization, existing substations in outer-urban space

5. Road networks
- Road category

6. Port facilities

A. Specific information
- location , starting year, area description, management body
- Licenses
- Impacts ascertained

B. General information
- Incompatibilities with other uses
- Possible impacts

7. Decentralized Administration Services: Regions, Municipalities, Chambers, other services
- Location , name, services provided, telephones, e-mails

8. Markets and Auction Houses
- Location, name, services provided, telephones

9. Aggregates of small touristic units
- Location, description, type, capacity

10. Major touristic units
- Location, name, type, capacity

11. Permanent exhibitions - Exhibition halls
- Location, title, services provided, telephones , e-mails

12. Centers for education, research, innovation and technology
- Location, title , services provided, telephones, e-mails

Natural Environment and Problems

1. Soil / subsoil
- Description
- Measurements, researches
- Problems (erosion, pollution, desertification)

2. Watersheds / Water Resources Status (adequacy, quality)
- Description of basin
- Location, rivers, lakes, artificial water systems
- Supplies of main water systems
- Problems

3. Forests, Protected Areas Status, Species Biodiversity
- Description, measurements, surveys

4. Marine Environment - Fish Stocks
- Description, measurements, surveys

5. Atmosphere, Noise pollution
- Description, measurements, surveys

Urban Environment

1. Settlements Limits - Urban Data- Traditional / Touristic Settlements - Archaeological Sites - Historical Sites
- General Urban Plan, decisions of settlements delimitation

2. Settlements outside the General Urban Plan

Institutional Environment

1. Statutory Areas

2. Industrial Area / Industrial Park
- Ownership / Management Body
- Foundation year / changes / history
- Area
- Established businesses, operational infrastructure
- Provisioned infrastructure
- Possible impacts and environmental terms in accordance with the Decisions of Environmental Permits

3a. Nominated areas of absolute protection and protection of nature

3b. Listed Monuments of Nature

3c. National Parks

3d. Landscapes of outstanding beauty

3e. Notable wetlands
- Management Bodies

4. Natura Areas

5. Spatial Plans – regulations – provisions of national (special) Spatial Plans
- Listing of regulations regarding:
- National Spatial Plan
- Specific RES Spatial Plan
- Tourism Plan
- Industry Plan
- Aquaculture Plan
- Regional Spatial Plan
- Open-City Spatial Plan / General Spatial Plan

7. Approved Master Plans, General Urban Plans , Open-City Spatial Plans, Urban Control Zones

Year of preparation, area of reference, competent service of implementation

8. Legal information
- Legal Environmental Guide, European Directives
- Competent bodies / institutions
- Appeals - major convictions

4 DATABASE MODEL

The database design follows the Entity-Relationship model (Thalheim, 2000), with two entity sets being at the core. The first one consists of all the *basic objects* (*B*), representing economic activity, environment etc., while the second one concerns *legislation* (*L*). In the sequel we present both and discuss the overall schema as depicted by the Entity-Relationship (ER) diagram of Figure 1. For presentational reasons the ER diagram only shows the core dependencies, not including attributes, full entity categorization etc.

4.1 Basic Objects

Entity set B essentially implements the categorization presented in Section 3. Therefore, a basic object in the database belongs to one of the following categories: primary, secondary and tertiary sectors, natural and urban environment and infrastructure. These categories are further split into subcategories depending on object specifics. For instance, the primary sector is split into: agriculture, livestock/poultry and aquacultures. The refinement of entity-sets is presented in Subsection 4.4

A basic object might “expire” for various reasons. For instance, an industry might seize to operate. Therefore, aside from the domain specific attributes a basic object has (outlined in Section 3), it also has a lifetime captured as start and end dates. We expect that data about expired objects will be kept for historical purposes.

Furthermore, a basic object might belong to a greater group, e.g., a specific industry might belong to the industrial zone of a city. We model the above through the self relation “Group”.

Last, certain objects might have time dependent attributes. For example, in an agricultural unit, the owner as well as the type and volume of production, can change from time to time. We model this through the entity set “Time Dependent Attributes” and the relationship “BOTDL”. It is worth noting, that the entity set “Time Dependent Attributes”, in fact consists of multiple tables (at least one for each subcategory of “Basic Object”). It is also hard to determine in advance for each case all the time dependent attributes that will be required. Therefore we will examine the use of Entity-Attribute-Value model (Dinu and Nadkarni, 2007) that permits such variability of attributes, but might affect query performance.

4.2 Legislation

In the entity set “Legislation” we model the related laws and regulations. Fully capturing the semantics of legislation on environmental and development issues is a formidable task on its own, and cannot be accomplished within the EGIS budget limitations. However, we plan to include key aspects of the legislation as sets of rules that either force or restrict concerning: (i) specific values on basic object attributes and (ii) the coexistence of basic objects. Thus, we will be able to model to a large extend, laws regarding the permissible activities in areas of environmental interest, allowable pollution levels etc. Here too, we plan to examine the use of Entity-Attribute-Value model in implementing the entity set “Rules”

4.3 Other Design Issues

In the ER diagram, the relationship “Concerns” is tertiary between “Legislation”, “Rules” and “Reference Category”. This is done in order to tackle situations where regulations are too vague to be associated directly with specific basic objects (“BOLeg” relationship) or with specific

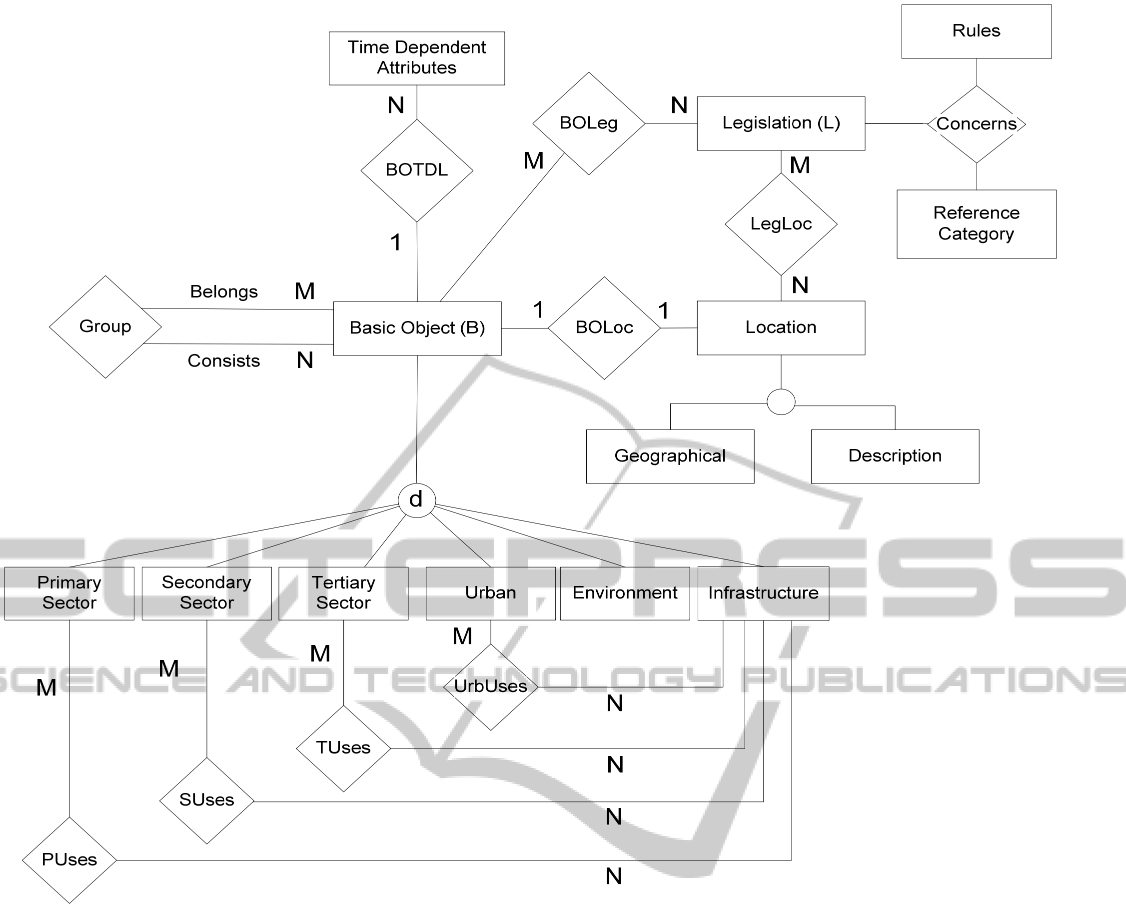

Figure 1: Basic ER diagram.

geographical areas (“LegLoc” relationship).

Another design decision we took was to enable both text descriptions and geographical data to be part of a “Location”. This was deemed necessary since much of the data provided by Central Greece prefectures have no clear geographical reference. Concerning the storage of geographical infor-mation per se, depending on the case, it can be a single point, a line segment, a polygon, or a set of the previous elements.

Last but not least, we model the fact that human activities often require some sort of infrastructure, through the relationships “XUses”. Since different kind of infrastructure is used in different ways, volumes, metrics etc., it is expected that the relevant to “XUses” tables will be sparse making again an Entity-Attribute-Value table implementation worth investigating.

4.4 ER Diagram Refinement

To have a more complete depiction of the ER model, in the following, we present specializations of the basic objects entity-sets. For the sake of figure clarity and space, attributes are not presented. In Figure 2/3/4, the analysis of the Primary / Secondary / Tertiary Sector entity set is presented, while in

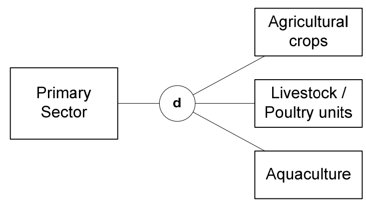

Figure 2: Analysis of the Primary Sector entity-set.

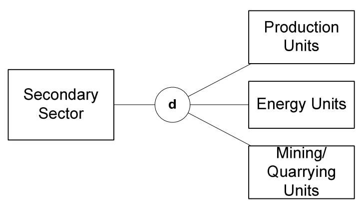

Figure 3: Analysis of the Secondary Sector entity-set.

Figure 5/6, the analysis of the Environment/Infrastructure entity set is presented. The Ubran entity-set is not presented, since we consider that there is no need to further refine it.

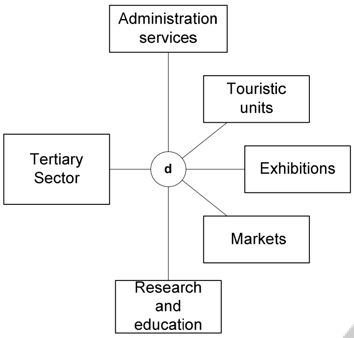

Figure 4: Analysis of the Tertiary Sector entity-set.

5 INDEXING TECHNIQUES

Out of the two basic entity sets, objects B, and objects L, the representation of the first one of them requires a Geographical Property, since such an object possesses a certain geographical position. The second object set refers to the first one, through relationship sets of the ER diagram of Figure 1. To efficiently process queries, techniques for indexing these object sets and implementation of the relations between them are needed. We choose to store all objects in a spatially enabled Relational Database Management System, e.g. PostgreSQL/PostGIS (Regina and Hsu, 2011). More specifically, the key attributes of the objects (object IDs) will be indexed by an 1-d access method, like the B-tree, or the B+-tree (Comer, 1979), supported by all relational DBMSs.

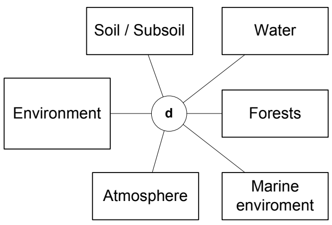

Figure 5: Analysis of the Environment entity-set.

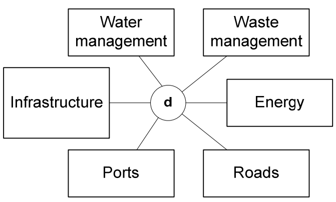

Figure 6: Analysis of the Infrastructure entity-set.

PostgreSQL/PostGIS supports spatial indexes, like R-trees (Manolopoulos et al, 2006), for indexing the spatial attributes of the objects. The R*-tree (Beckmann et al, 1990) is a more efficient member of the R-tree family for query processing than the original R-tree and if the RDBMS used in implementation does not support it, external spatial indexing of the object IDs could be a choice. This can be accomplished either through RDBMS embedded access methods, or through external indexing, the spatial attributes of the objects will be indexed by R-trees, or R*-trees.

When the spatial attribute of an object is a point, this attribute per se will be inserted in the spatial index. However, when the spatial attribute is a line segment (e.g. a river, or road), a polygon (e.g. a lake, or field), or a set of points (e.g. a group of weather monitoring sensors), since R-trees are based on Minimum Bounding Rectangles (MBR) of objects and internal nodes, the MBRs of objects will have a non-zero area (an MBR might even cover the reference area, the Region of Central Greece) and overlap between MBRs within the tree structure will be high, reducing performance of query processing. In this case, a new technique of indexing is required that would partition the spatial characteristic of objects in a controlled way, to keep query processing performance high. Possibly, a member of the Quadtree family that partitions space in a hierarchical regular fashion would be appropriate. Quadtrees have been recently shown to be competitive to R-trees (Kim and Patel, 2010, Roumelis et al, 2011).

When the user query (or one of the individual queries to which a complex user query is analyzed to) is based on a condition for the object IDs, it will be submitted to the RDBMS through SQL, taking advantage of the index on objects IDs. When this query is based on a condition for the spatial property of objects, it will be submitted to the spatial index and the objects IDs satisfying this query will be returned. Subsequently, the objects IDs will be used for retrieving further properties of the objects from the DBMS, through SQL, by taking advantage of the index on objects IDs. This two step process could also be reversed (accessing the non-spatial index, retrieve spatial object characteristics and subsequently the spatial index).

6 CONCLUSIONS

In this paper, based on existing modelling and analysis techniques and existing techniques of database indexing and query processing, we designed the data model and outlined the required indexing that can lead to efficient query processing of a Geo-Environmental Information System for the Region of Central Greece. This systems aims to be utilized for environmentally aware political decision making, development, investments and spatial planning and economic activities monitoring / auditing, or, in other words, as a tool for combining the protection of the environment and economic growth, in a region with rich natural resources, existing environmental problems and high potential of development.

In the future, we plan to implement this design in a pilot EGIS that will be assessed by final users and domain experts (environmental-management and spatial/development-planning specialists). Based on this feedback, the INSPIRE directive (http:// inspire.jrc.ec.europa.eu/) and related literature (e.g. Paolino et al., 2010), the design will be updated and enhanced. Additionally, the use of Entity-Attribute-Value model (Dinu and Nadkarni, 2007) and further incorporation of the time dimension in the EGIS data model will be examined. Moreover, we will evaluate the indexing and query processing techniques embedded in the pilot system and, based on spatial indexing and query processing literature (related reviews appear in Corral and Vassilakopoulos, 2009, and in Vassilakopoulos and Corral, 2009) we will develop new techniques (like access methods that partition objects with non-zero area), aiming at increased efficiency of demanding query processing.

ACKNOWLEDGEMENTS

Work funded by the "Development of a Geo-ENvironmental information system for the region of CENtral Greece" (GENCENG) project (SYNERGASIA 2011 action, supported by the European Regional Development Fund and Greek National Funds); project number 11SYN 8 1213.

REFERENCES

Beckmann, N., Kriegel, H.P., Schneider, R., Seeger, B., 1990. The R*-tree: an Efficient and Robust Access Method for Points and Rectangles. In *SIGMOD Conference,* 322-331.

Comer, D., 1979. The Ubiquitous B-tree. *ACM Computing Surveys* 11(2), 121–137.

Corral, A., Vassilakopoulos, M., 2009. Query Processing in Spatial Databases. In *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends,* Vol II, 269-278, Information Science Reference.

Dinu, V., Nadkarni, P., 2007. Guidelines for the effective use of entity-attribute-value modeling for biomedical databases, *International journal of medical informatics*, 76 (11–12), 769–779.

Gandhi, V., Kang, J. M., Shekhar, S., 2009. Spatial Databases. In *Encyclopedia of Computer Science and Engineering,* Cassie Craig (Eds.), Wiley.

Kim, Y.J., Patel, J., 2010. Performance Comparison of the R*-tree and the Quadtree for kNN and Distance Join Queries. *IEEE Transactions on Knowledge and Data Engineering* 22(7), 1014–1027.

Manolopoulos, Y., Nanopoulos, A., Papadopoulos, A.N., Theodoridis, Y., 2006. *Rtrees: Theory and Applications,* Springer, London.

Regina, O., Hsu, L., 2011. *PostGIS in action,* Manning Publications Co., Stamford.

Roumelis, G., Vassilakopoulos, M., Corral, A., 2011. Performance Comparison of xBR-trees and R*-trees for Single Dataset Spatial Queries. In Proc. of ADBIS 2011, 228-242.

Thalheim, B., 2000. *Entity-Relationship Modeling: Foundations of Database Technology,* Springer-Verlag New York, Inc. Secaucus.

Vassilakopoulos, M., Corral, A., 2009. Spatio-Temporal Indexing Techniques. In *Handbook of Research on Innovations in Data base Technologies and Applications: Current and Future Trends*, Vol II, 260-268 Information Science Reference.

Wainwright, J., Mulligan, M. (Eds), 2013. *Environmental Modelling: Finding Simplicity in Complexity,* John Wiley & Sons, Chichester, 2nd ed.

Paolino, L., Sebillo, M., Tortora, G., Vitiello, G., 2010. Integrating Discrete and Continuous Data in an OpenGeospatial-Compliant Specification. *Transactions in GIS*, 14 (6), 731-753.