FEDERATING HETEROGENEOUS DATABASES IN A MULTI-LEVEL DECISION SUPPORT SYSTEM FOR WATERSHED MANAGEMENT - A CLIENT/SERVER APPROACH

Mustafa Radwan., Yaser Bishr, Edison Espinoza., and Tankiso Mabote.
Department of Geoinformatics
International Institute for Aerospace Survey And Earth Sciences, ITC
P.O.Box 6, 7500 AA Enschede, The Netherlands

Commission IV, Working Group 6

KEY WORDS: Decision Support System, Information sharing, Watershed management, Multi-level Decision making

ABSTRACT

Environmental decision making in general and watershed management in particular involve three main components: ecological, economic and social. These components are highly correlated and interdependent. Ideally we should consider the impact of changing one component on the other in the decision making process. The assessment of the impact of a proposed program should consider two main aspects: the interrelationship between the components of the underlying watershed (ecological, economic and social) and the hierarchical level of decision making (local, regional and global). The latter strictly affects the relation between the environmental variables.

In this paper a proposal for developing a multi-level decision support system for watershed management is presented. The system provides a link between the three levels of decision making. The emphasis is on the aspects related to linking the component databases which support the three levels. These databases play an important role in the multi-level analysis of watersheds. The link must allow information and decision transfer and resolve databases heterogeneity. Problems as well as their solutions, which are pertaining to linking these spatial databases, are explained.

1. MUTLI-LEVEL WATERSHED MANAGEMENT

In order to eventually propose the proper components and architecture, the first phase in developing multi-level decision support system for watershed management, MLDSS, is to analyze the functional relationship between the three levels of interest, i.e., national, regional, and local. The MLDSS should be able to support three main activities in watershed management: 1- **monitoring** of watersheds and basin status in order to keep records of the type and rate of degradation; 2- **analysis** of watersheds in order to investigate the causes of their degradation and propose plans for sustainable management; 3- **management** of watersheds, where the actual execution of the plans is executed.

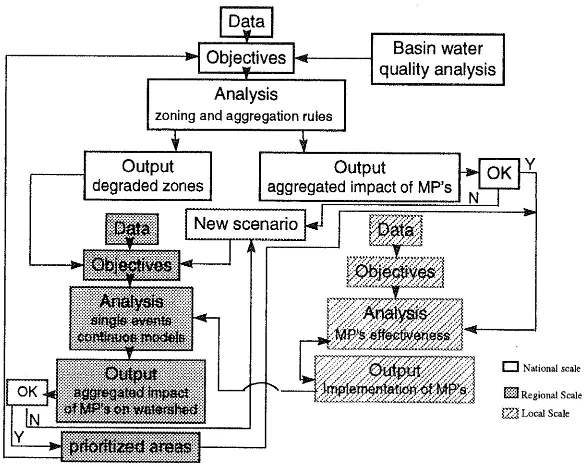

Figure 1 Interrelations between the three scales

Figure 1 shows the interaction and relationship between the three levels. They are characterized by an extensive flow of information and decision between the three levels. This can be summarized as follows:

- ♦ *At the national level*, decisions concerning the guidelines and constraints for initiating an environmental sustaining project are defined for the whole country..Decisions are taken based on the information provided by the regional level. The information is used to analyze the underlying watershed in order to locate and identify degraded areas. The areas are then ranked according to the rate and degree of degradation, and their social and economic impact. At this level, political factors are likely to be considered and even sometimes can overrule other considerations. Information on the impact of the new management practices on each watershed, at the regional level, is used for further analysis at the national level in order to improve decision making.
- ♦ *At the regional level*, decisions are mostly dealing with identifying the proper combination of management scenarios. Usually the objective at this level is to have maximum positive environmental impact on the whole watershed. This combination and their corresponding impacts are aggregated and quantified at the national level for approval as mentioned above. Subcatchments are also analyzed and ranked at the regional level. The ranking is used to prioritize sites for further analysis and protection [Bishr et al., 1995]. Scenarios with highest ranks are implemented at the local scale.
- ♦ *At the local level*, scenarios with the highest priority are implemented. Their results and impacts are used to provide a feedback to the regional scale for modifying and improving the scenarios.

Despite all developments in computer based tools, many mangers are not benefiting from these competent tools. The main reason for this is the difficulty of developing an integrated data model which supports DSS components. Management support systems and decision support systems in particular are technologies designed to overcome this dilemma. A decision support system (DSS) refers to a collection of computerized technologies whose objective is to support managerial work and especially decision making [Turban E., 1993]. The architecture proposed in this paper is designed such that it allows transfer and exchange of information across levels. In the next, a generic architecture for spatial decision support system for watershed management, is shown.

2. MULTI-LEVEL DECISION SUPPORT SYSTEM

According to the functionality between the three levels of decision making mentioned section 1, spatial decision support system, SDSS, should maintain the following capabilities:

a) Ability to retrieve, process, format, display, and store data sets and information using current technology and appropriate models that help managers at each level in their processes of monitoring, analysis and management of watersheds.

b) Ability to provide an integrated data model which supports all DSS components.

c) Ability to communicate with the different hierarchical management levels by exchanging data, knowledge, and decisions. This particular possibility is an extended capability of SDSS which gave rise to the concept of Multi-Level Spatial Decision Support System, MLSDSS as proposed by [Radwan M., et al., 1995].

2.1 Components Of The Architecture

Figure 2, shows an architecture which provides a link between several decision support systems. The architecture is based on the client-server architecture [Boar, H.B., 1993]. It is also built on the assumption that each decision level has its own SDSS. The detailed components of such SDSS is explained in section 3. In this architecture we distinguish between two types of databases. One is called the basic database which is required by all levels of decision making. The basic data are forming the lowest level of the aggregation hierarchies of each decision level. For this reason we provided a dedicated server for transferring these basic data to each SDSS. The second type of database, called the spatial decision support system database, is the one which is supporting its underlying SDSS. As mentioned in section 1, each level has its own objectives and consequently has its own view of geographic objects, i.e., of the basic databases, and intuitively different aggregation hierarchies. Having different aggregation hierarchies gives rise to a problem known as databases heterogeneity. A multi-level global server is

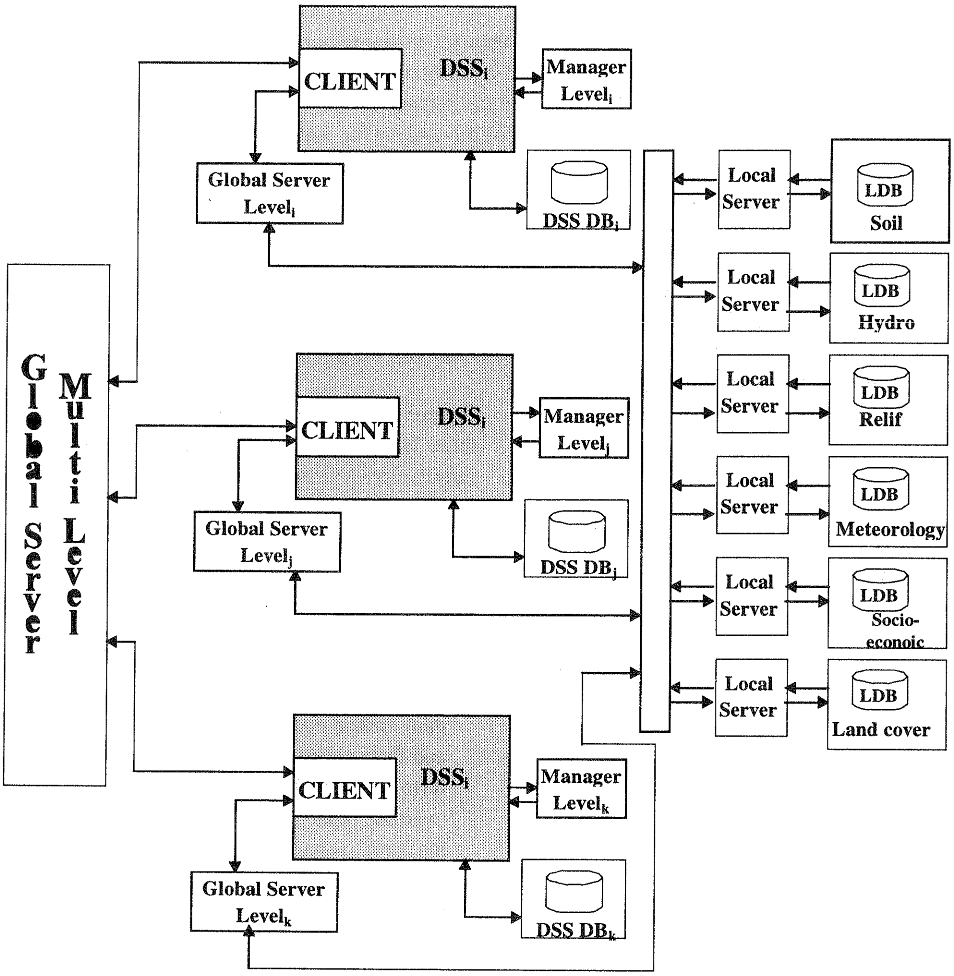

Figure 2 Proposed Architecture Of MLSDSS

provided in order to resolve the heterogeneity between the different SDSS databases. In this section we only emphasis on the system architecture for resolving the heterogeneity. In section 4, aspects of heterogeneity presented.

LOCAL SERVER (LS):

The local server represents the gate for communication of the local database with the global server. The local server contains data that each participating basic database is willing to share with the federation. The sharable data is supported by a metadata which contains information about the data stored in the database [Bishr Y., 1996].

The metadata specifies the characteristics of the data items: format, quality, collection and processing procedure, etc. The metadata should also contain schema and mapping definitions to enable the abstraction of the sharable data to a global export schema.

The role of the LS is to put at the disposal of clients a common object-oriented interface on top of the local DBMS. The view is then object oriented and is implemented through methods. In a sense, the LSs are abstractions of the local DB. It is responsible for accessing and retrieving data as requested by the Client through the Global Server. For the present work, the server hosts elementary databases of single data type, e.g., soil, hydrology, land cover/use, relief, etc., at a detailed map scale.

GLOBAL SERVER (GS)

The global server has a global external schema which provides a mechanism for resolving the different aspects of heterogeneity, section 4. In the proposed architecture, each level of decision making: i, j, or k, from now on called client, is provided with a GS, i.e., GS_i or GS_j or GS_k, which support all users at their underlying level.

To support each level, the corresponding GS links its client with the LSs. Upon the receipt and acceptance of the client's request, the following operations are performed:

- Analyze the request to identify and locate the required data items and send the corresponding messages to the appropriate LSs.
- Receive the data sets from the LSs and process them to provide the adequate data items.
- Reply to the client by sending the requested data.

Additionally, the global server has the following tasks:

- Control the transactions with the clients and data sources.
- Maintain the global directory, i.e., information about the data available within the federation: location, information on specific data sets, ownership, format, cost, etc. This is achieved by storing a comprehensive metadata in the global server.
- Execute data conversion: units, formats, etc.

MULTILEVEL SERVER

The multi-level server is responsible for linking the different decision making hierarchies to establish the corresponding feedback between the three decision levels in terms of data, knowledge, and decisions necessary for their activities. Its tasks are:

- Control the communication between clients (management levels),
- Access and retrieval from the corresponding DSS database in a specific level.

In this section, the system architecture and the functionality of each component are explained. In the next section the components of the DSS are outlined. In section 4, more emphasis is given on the different aspects of heterogeneity of the databases which support such an architecture. section 0 shows methodology followed for implementing such an architecture. Then the paper is concluded in section 6.

3. ARCHITECTURE OF THE SDSS

The proposed system, Figure 3, is a hybrid system which incorporates an expert system, a GIS software, a remote sensing software, and application Models. The selection of components of the proposed system is the result of the classification and definition of the problem types to be solved in WSM. The system integrates the supporting databases of each of those components into a single data model. The system has the following Components:

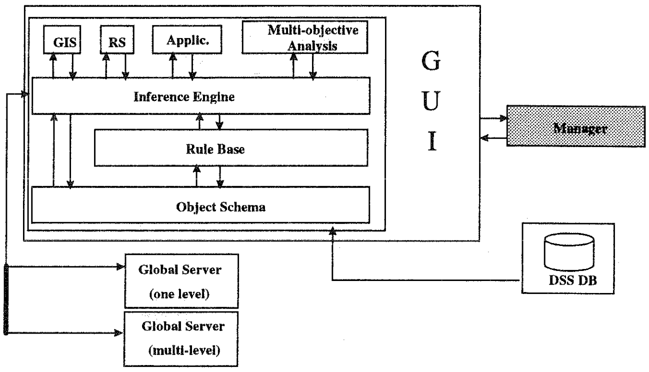

Figure 3 Components of SDSS

GIS Component

GIS functionality is accessed through its interface with the DSS. The possibility to use geographic information systems enables DSS users to handle spatial and non-spatial data from different disciplines for the purpose of spatial data management, analysis and presentation.

RS Component

Remote sensing capabilities are accessed through its interface with DSS. This capability provides the benefit of capturing and analyzing remote sensing data and the opportunity to perform monitoring operations of particular aspects of the environment relevant to watershed management.

Application software

In order to provide specialized analysis in watersheds, there are several simulation models which are designed to analyze different aspect of watersheds, e.g., soil erosion and water quality.

Multi-objective/Multi-criteria Analysis Module
For the task of analyzing the conflicting goals/objectives of the different decision levels and disciplines involved in watershed management.

Inference Engine
Manipulates the knowledge represented in the knowledge base to infer a solution to the problem(s) described by the information in the databases.

Rule-based system
This is the knowledge representation scheme. It implements the concept of data-driven and goal-directed reasoning. The rule-based system contains the knowledge about a particular domain that has been acquired from domain experts.

DSS Database
This database contains specific data (initial and derived data when solving the problem) which describe all facts that are known about the problem. Data may be the result of: some GIS operations; data automatically collected by a data acquisition system; sets of answers provided by the end user in response to the queries posed by the inference engine; or data exported from remote databases.

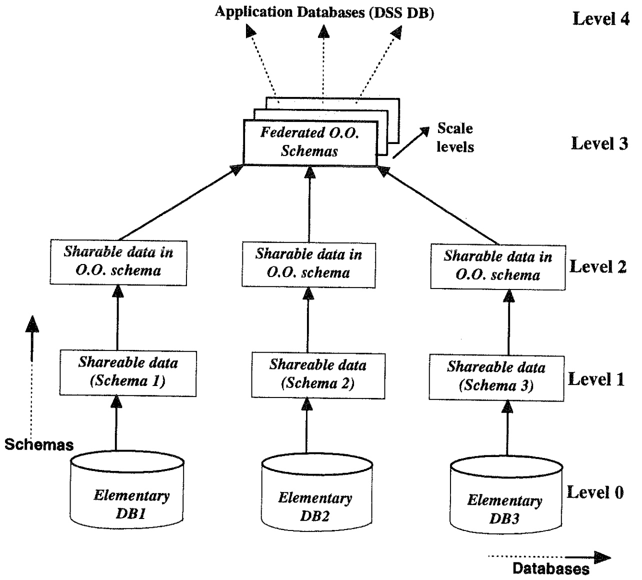

Figure 4 Multi-level federate database

Object schema
The object schema describes the entities in the domain (in terms of classes, objects, properties, methods, inheritance, etc.). It also enables the retrieval of the object properties from the SDSS database using the Object-Oriented data model. The object schema provides an integration layer that supports the requirements of the different SDSS components mentioned above.

GUI
It provides users with an intuitive interface for interacting with the SDSS and its various components. The basic objective of the GUI is to allow users to access GIS, RS, application programs, rule base and multi-objective analysis components from a single interface.

In this section, the components of the MLDSS were shown. First the overall architecture of the MSLDSS were explained, then the components of the SDSS were presented. The supporting databases of such an architecture are heterogeneous in many respects. The next section shows how the data models are organized and structured in the MLDSS.

4. HETEROGENEITY IN DISTRIBUTED GISs

[Worboys, M., et al., 1991] classified semantic heterogeneity as generic and contextual. The earlier occurs when different GIS applications use different data models of representing their spatial information. For example one may use a layer-based approach while a another may use an object-based approach. The contextual heterogeneity occurs when the semantics of schemes depend upon the local conditions at particular GIS. For example two spatial databases contain two different objects which have two different meanings, though they refer to the same real world entity, e.g., agricultural fields in environmental database are different from those in a cadastral database. [Spaccapietra, S., et al., 1991] listed 4 classes of heterogeneity or conflicts: semantic conflict, descriptive conflict, data model conflict, and structural conflict. The semantic conflict occurs in the situation where two sets of objects from two schemes are representing sets of real world entities which are related by a set comparison operators other than equality. Descriptive conflict occurs when two database objects, representing the same real world entity, are described with different sets of properties. Data model conflict is the situation where two schemes are defined with two different data models, e.g., relational and object oriented models. The situation where two related objects are represented using different data structures is called structural conflict. For example a designer represents a component X of an object O either by creating a new object type X or add it as a property of O.

A relatively similar classification of types of database heterogeneity is presented by [Saltor et al., 1993]. They provided more comprehensive classification of heterogeneity to which we are more inclined. Their classification has three aspects: syntactic, schematic, and semantic. Descriptive and structural conflicts are equivalent to schematic heterogeneity, while the data model conflict is equivalent to syntactic conflict.

1. Syntactic: each database may be implemented in a different DBMS with a different data model, e.g., relational model Vs object oriented model. Syntactic heterogeneity is also related to the geometric representation of geographic objects, e.g., raster and vector representations.
2. Schematic: where objects in one database are considered as properties or metadata in the other, or object classes of the same real world entity have different hierarchies and descriptors in different databases.
3. Semantic: a real world entity may have two different meanings in their underlying databases in order to serve various applications, giving as a consequence semantic conflicts. For example a road network in a GIS for transportation has different semantics from that in a GIS for

topographic mapping.

An architecture for sharing data stored in disparate geographic information systems is shown in Figure 4. It is a four levels architecture:

- At the first level, level 1, schemes are representing the part of the database that each member in the federation is willing to share. These schemes are represented with the native language of the host.
- The second higher level of schemes is to model the same data posted at the first level but with a uniform data model (e.g. Object Oriented).
- The third level is a global schema which provides a uniform perspective of the information posted at the second level.
- The federated schemes are supporting different scale levels, i.e., national, regional, and local.
- Another fourth level of schemes is representing the application specific views which can be partly retrieved form the federation (e.g. application database).

The object oriented sharable schemes in level 2, Figure 4, are represented in the local servers, Figure 2. The federated object oriented schemes, Figure 4, are represented in their corresponding global servers GS_i, Figure 2. The multi-level global server in Figure 4, is the one which provides the link between the applications, i.e., level 4 in Figure 2.

As shown in Figure 5, there could be several watershed management projects running in parallel. Each individual project can have positive impact when viewed locally. However, they might have negative impact on the basin, when their overall impact is screened. For this reason management projects proposed at watershed level should be assessed at the basin level. The implementation of the system architecture proposed in Section 5 aimed at showing this case. Management practices where introduced into watersheds. The management practices proved to minimize the soil erosion. This impact was quantified using AGNPS model. The impact of these management practices was then analyzed at the basin level. This impact was quantified using DUFLOW. In order to achieve this objective, the multi-level decision support system should allow data and decision transfer.

In the previous sections the system architecture of the multi-level decision support system for watershed management is shown. Moreover, an architecture for resolving the aspects of heterogeneity in the databases supporting such system was proposed in this section. A prototype which implements the system is presented in the next section. Only the system components and functionality are shown. The supporting data models are outside the scope of this paper.

5. SYSTEM IMPLEMENTATION

Four main software packages were used to implement MLSDSS the client-server operations:

- Nexpert Object as the object-oriented shell and rule-base system.
- Arc/Info as the GIS platform.
- AGNPS as the erosion and water quality simulation model at the regional level, i.e., catchments and subcatchments.
- Duflow as the erosion and water quality simulation model at the national level, i.e., basin. It is used to analyse the overall impact of management practices on the basin.

Nexpert Object supports a rich range of representation features. In Nexpert, the domain is modelled in terms of objects, classes, and properties. Rules are used to manipulate the objects and class structures. The specific properties of objects and classes are called slots. Meta-slot attributes are used to describe certain characteristics of the slot. Nexpert Object shell is used to build an object oriented shell around the sharable data and represent the object network for the client, local servers, and global server. The rule-base is used to design the global server and the database bridge will be used by the local servers to transfer data between dBase and Nexpert representation. Nexpert object will have several roles in our system. First it is used for integration and implement the components of the SDSS, i.e., ARC/Info, AGNPS, and Duflow, [Espinoza E., 1995 and; Mabote T., 1995] gave a detailed description of this system. Second Nexpert is used for developing the schemes of the MLDSS and as a mediator for resolving the heterogeneity between the supporting databases. In this paper we only emphasis the technique followed for implementing the servers.

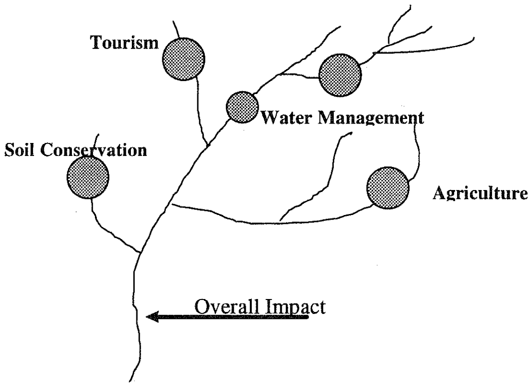

Figure 5 Several management projects in one basin

Servers Implementation

To simulate the servers in the proposed architecture, corresponding Nexpert's Knowledge Bases are developed, one for each server.

LOCAL SERVER

Using NEXPERT, an object oriented shell is put on top of the local database. This shell is basically the object oriented view of the sharable data schema using the canonical or common data model. To provide the server functionality it is necessary to develop methods that enable the communication between the GS and the Local Database. Using these methods, messages can be passed from the GS to the LS in order to retrieve data, as requested by the client, from the underlying database. The retrieval asked by the GS is originally an Object Oriented query that the LS has to translate into an equivalent query for the local DBMS.

GLOBAL SERVER

The sharable data schemes, posted by the local databases in the federation, are integrated to form a global conceptual schema (GCS) to serve watershed management. The GCS is held by the GS and is intended to support, among others, DUFLOW .

The operations indicated for the GS in section 2.1 require the implementation of methods and rules. The rules are needed to select, upon the receipt of the Client request, the servers where the required data are located and the procedures to get the data using the information contained in the metadata. The methods, and the corresponding messages, are needed for: requesting information from the LS, aggregating data, and sending data to the client.

MULTILEVEL SERVER

The conceptual model developed for the integration of the management levels is implemented as a schema in the Multilevel Server. It is composed of methods, and the corresponding messages, to enable the communication between the decision making levels and the transfer of data, knowledge and decisions amongst them. There are methods for communication, retrieval of data, sending decisions.

6. CONCLUSIONS

The main objectives of this research were to develop a federated database system able to support watershed management at different levels of decision making and to develop a model that enables the interaction between these levels for environmental management.

Object-Oriented modelling was selected for the development of the servers of the federation. Its powerful constructs: classification, generalisation, association, and aggregation, together with the other features provided by Object-Oriented concepts, like inheritance, encapsulation, operations, and message passing, renders a powerful tool with great expressiveness and semantic contents. In the present research, it was realised how an object oriented shell on top of the local DBMS facilitates the federation of heterogeneous databases. Peculiarities of the data structures provided by the local DBMS are encapsulated, and in this way are hidden from the user who would be provided, additionally, with the more powerful constructs of Object-Oriented model for his/her developments.

The implemented prototype gives an indication of the advantages that the client-server computing can offer in the GIS environment. Furthermore the prototype showed that multi-disciplinary environmental applications can be supported through the FDBS architecture.

The use of the information provided by the Federated Database would prove to be more useful if, as it has been shown, all management levels can profit from it by means of easy access and use, and facilities for interchange of information, data, knowledge, and decisions. In this context we are currently developing a mechanism for semantic data sharing across heterogeneous databases.

RFERENCES

1. Bishr Y., 1996 : "*A Hierarchical Spatial Canonical Data Model – Towards Federating Heterogeneous GISs*". Proceedings of the ISPRS Conference IWG III/IV-Conceptual Aspects of GIS, Vienna.
2. Bishr Y., and Radwan M., 1995: "*Preliminary Design Of A Decision Support System For Watershed Management*". ITC Journal, 1995-1.
3. Bishr Y., Molenaar M., and Radwan M. 1996 "*Spatial Heterogeneity Of Federated GIS In A Client/Server Architecture*". In Proceedings of the GIS'96 Vancouver Canada. To Be Published.
4. Boar, H.B., Implementing Client/Server Computing, A strategic perspective. McGraw-Hill, Inc. 1993. ISBN 0070062153.
5. Radwan M., and Bishr, Y., 1995: "*Towards A Multi-Scale Decision Support System For Watershed Management*". *Paper Presented in Africa GIS'95, Ivory Coast.*
6. Saltor F., Castellanos M.G., and Gracia-solaco M., 1993 "*Overcoming Schematic Discrepancies in Interoperable Databases*". In Proceedings of the IFIP WG2.6 Database Semantics Conference on Interoperable Database Systems (DS-5), Lorne, Victoria, Australia, 16-20 November, David K. Hsiao, Erich J. Neuhold, and Ron Sacks-Davis (eds), pp. 191-206.
7. Spaccapietra S. and Parent C., 1991 "*Conflicts And Correspondence Assertions In Interoperable Databases*". SIGMOD Record, Vol.20, No. 4, pp49-54.
8. Turban E., 1993, "*Decision Support And Expert Systems, Management Support Systems*", 833 pages. ISBN 0-02-422691-7.
9. Worboys, M.F., and Deen, S.M., December 1991: "*Semantic Heterogeneity in Distributed Geographic Databases*". SIGMOD Record, Vol. 20, No 4.
10. Espinoza E. J., 1995: "*Development Of A Federated Database System In A Distributed Environment For Environmental Decision Making At Different Scales*". Unpub. M.Sc. thesis, ITC, Enschede, The Netherlands.
11. Mabote T., 1995 : "*The Design Of A Prototype To Support Watershed Management Using The Client-Server Concept In A Distributed Environment*". Unpub. M.Sc. thesis, ITC, Enschede, The Netherlands.