FEDERATED MULTI-DATABASE INFRASTRUCTURE FOR GIS INTEROPERABILITY - THE DELTA-X PROJECT

Mosaad Allam
Cherian Chaly

Geomatics Canada
615 Booth Street, Ottawa, Ontario, Canada K1A 0E9

Ekow Otoo

Interobject Spatial Research Inc.
Ottawa, Ontario, Canada

ISPRS Commission II, Working Group 3

KEY WORDS: Interoperability, Data-Federation, Metadata, Spatial-Browser, Client-Server, Infrastructure.

ABSTRACT

Access to a federated set of dispersed spatial databases over a wide area network is a major issue. Considering that there can be a variety of GISs and DBMSs residing on a heterogeneous array of platforms, this precipitates problems of data transfer from system to system. From the users point of view, the data should be current, available in readily usable format, easily accessible, supported with database management functionalities and allow selective feature retrievals. The user should have the option to use whatever database or GIS which is appropriate for the application.

Delta-X, a code name for a federated spatial information management system, is an approach to providing interoperability in a network of heterogeneous databases. It establishes interoperability between different relational DBMS's, simple files and object repositories. Data is made accessible to users through a distributed virtual global schema definition. By this means, the independence and autonomy of control, of the individual datasets registered in the federation, are still respected, while still maintaining some integrated uniform access with respect to a particular GIS environment.

To assist the users in identifying the source of data required for their applications, the MetaView/GIS spatial browser was developed. MetaView/GIS provides access to metadata of various databases. The Delta-X and MetaView provide a unique technology for the development of a Canadian spatial datainfrastructure.

1. INTRODUCTION

Geomatics Canada Mandate is to provide accurate and current spatially georeferenced information about Canadian landmass. The Geographic Information Systems and Services, a Division of Geomatics Canada, was set-up in 1987 to develop applications to promote the use of the Department data, and to explore research that will aid in the growth of GIS technology. Initially, our main concern was to improve access to GIS data through improved communications facilities, and developing the technology required for building a spatial information infrastructure in the Sector to facilitate the ordering and distribution of GIS data.

In 1988, the Division established the National GIS Technology Centre and acquired several GISs from several vendors. The GIS software was installed on personal computers (PCs), and UNIX workstations. The various systems were interconnected into an ethernet local area network (LAN). The heterogeneous GIS environment was selected due to the wide variety of GIS software and hardware platforms available on the market, and to simulate conditions in large organizations, where different GISs are used to store their spatial data in different databases.

The operation of the GISs in this simple LAN, in which a number of workstations (nodes), each runs a different vendor GIS software based system, and our efforts to retrieve data from various databases residing in other LANs in the Sector quickly revealed a number of operational problems, such as:

- the transfer of data between a number "n" of GISs requires "n^2" conversion packages,
- the different GIS have underlying database management systems (DBMS) that are based on different data models: relational, network, and object-oriented,
- the data required by GIS applications resides in several database in different data types: vector, raster, structured text, free text, and knowledge base,
- the available data is generally not in the format required by the GIS, and some processing and restructuring are required before it can be incorporated into the GIS application environment,
- checking data out of the server and restructuring it into its local GIS data storage is cumbersome,
- checking the data back onto the server, either as a new version or as an updated and revised form of an existing one posed more problems for database administrator,
- GIS databases available in agencies and institutions that are geographically dispersed

- in procuring data from other agencies, user must purchase all the information compiled for that map sheet and extract the feature relevant to the applications, and
- large volumes of datasets are currently not maintained with any DBMS. They are maintained simply as large sequential files on reels of tapes.

These and other related problems are what the federated multi-database spatial information management system (code named the Delta-X project) is concerned with. Such a model of operation is not exclusive to GIS. Similar problems have been addressed in purely corporate database environments. The Delta-X is a multi-database system with a common integrated global conceptual schema definition. The global database is intended to achieve interoperability between DBMS that have been specialized for three distinct data types: spatial data (vector and raster), structured text and free-text. Our design provides an integrated access to data stored in relational databases, object oriented databases, simple file systems and information retrieval and document management systems. The current implementation does this through *Remote Procedure Calls* by formulating queries in the format of the target systems. Where the target systems are relational DBMS, SQL is used as the language for remote data access. Simple file systems are mapped into relational tables. Thus, Delta-X establishes interoperability between relational DBMSs and GISs.

To provide a user accessible data warehousing user assistance for inventory and metadata review, a complementary system: MetaView/GIS Spatial Browser (MV/GIS) was developed as a front end to the Delta-X. The MV/GIS and Delta-X systems are the basic component of our Division's spatial data infrastrucure and provide a solution to Geomatics Canada data warehousing.

2.. THE DELTA-X SYSTEM

2.1 The Delta-X Common Spatial Data Model

This is the conceptual data model into which other GIS internal representations can be mapped. For vector data, the Delta-X spatial data model maintains up to degree four topology. This global conceptual scheme forms the intermediate transition schemes for data exchange between different GISs. The global conceptual data model has a mapping onto either a relational database, as a collection of relational tables, or into an object-base as a colony of categories. A Delta-X server maintains the vector data in a relational database if the underlying database management system is relational, e.g., Ingres, Oracle, Informix, etc. Similarly, the Delta-X server retains the data in a set of equivalent category classes in an object oriented database, if the underlying DBMS is object-oriented, e.g., ODE, ObjectStore, Objectivity. The significant idea is that Delta-X server transparently delivers data to and receives data from clients' GIS environment.

The illustration with the vector data shows the general approach for handling data in Delta-X. This approach, where by a common defined global data model is materialized for actual representation in local databases management system, is extended to handle the various classes of data-type required in GIS. For example, raster data such as remote sensed images, structured text that are related to spatial objects, and free text are all represented in the common global schema definition. Since some of these data types have internationally defined standards of representation, e.g., GIF, JPEG, MPEG, SGML, HTML, etc., these are maintained as files in their respective standard formats, and related to the defined spatial features through spatial indexes. A number of such non-vector data types are related to the spatial features using index techniques.

2.2 Spatial Indexing

Even over a small area of coverage, maintaining all the relevant information for all data types, quickly grows into a significant large database. Delta-X organizes space into hierarchical tessellated regions and maintains two distinct levels of index schemes for identifying a feature in defined space. The first level of index identifies a regular polygonal cell of the region of coverage that has been hierarchical tessellated into a near unicorn grid. The second level of index associates features within each cell with other information types. The first level uses a quadtree-like method of spatial index. The second level draws from a number of one-dimensional and multi-dimenesional index techniques depending on the data-type to be indexed.

The partitioning of the space into cells (tiles), by the first level of index enables the large volume of data to be accessed and controlled in manageable units. By this means, all data of a specific region or cell, and controlled by a particular organization or agency, can be grouped together. Geographic coordinate system (longitude and latitude) forms the common and basic coordinate reference system by which regions and features are related across multiple databases. The local database may derive the actual projection and coordinate reference system used in referencing spatial objects within that particular environment.

2.3 The Delta-X Features and Services

Delta-X performs multiple client-server roles. First, Delta-X servers control the data storage in commercial DBMS systems. Second, it is a client-server transaction processing system, and handles message and data exchange between Delta-X database and/or other specialized servers and Delta-X clients. Third, Delta-X performs specialized server functions, such as data conversions to and from the internal data interchange format. Finally, it acts as a proxy client on behalf of GIS systems that produce or consume data - it is the source or destination of data in Delta-X transactions

A single Delta-X server can be accessed by many Delta-X clients at the same time. Even a single Delta-X client can start many transaction on the same Delta-X server. In our current implementation, all conversion processes share the same CPU. The Delta-X server architecture, however, can also be implemented on a parallel machine or on a cluster of workstations on a dedicated LAN, where each transaction process runs on a different machine. All transaction processes share the same database.

Delta-X has a client component and a server component. On a client, Delta-X provides a GUI that enables GIS users to start and control transfer/conversion transactions and a software which communicates with the server. On a server, Delta-X provides the client-server and server-server communication software, a module that transfers data, a network management agent, an interface to a database, and a set of data conversion routines. The server's network management module facilitates setting of server's operational parameters, authentication keys, the entering and leaving of the Delta-X federation, and it also raises alarms to network management when the server malfunctions. Data on the server can be stored in any commercial database model - relational or object-oriented.

The exact internal format of data stored in Delta-X servers is database-specific, but all database servers share the same data model and logical schema, and they can store the same types of data. The internal database schema used at a particular server is transparent to all other servers in the federation. A separate database interface module adapts each internal database schema to a uniform data interface used within Delta-X. Conversion routines are also database-specific and GIS-specific. Note that the conversion routines and the database schema are the only entities in Delta-X that are specific to standard data formats or GIS systems.

The clear separation of data storage functions (the Delta-X server) from data processing functions (GIS) in a network not only facilitates data sharing and creation of logically organized application-specific data repositories, but it also enables to take advantage of the latest development in both of the DBMS and GIS worlds.

The Delta-X server is a node in a network of Delta-X servers which has a data repository containing a part of shared data stored in the Delta-X federation. As shown in Figure 1, the Delta-X comprises modules to perform the following functions:

- Transaction management and surveillance
- Data storage, access, and management
- Data conversion
- Client-server and server-server communication over a LAN or WAN
- Naming and authentication for Delta-X clients

The Transaction Monitor (TM) module performs transaction management tasks, such as transaction scheduling, transaction monitoring, authentication of transaction requests, locking of data, resource administration, and transaction commits, aborts and recoveries. TM receives requests for new transactions from Delta-X clients or requests to process already running transactions from fellow servers in the Delta-X federation. The Delta-X server validates the requestor's permissions for accessing the requested data and notifies the clients about the status of transactions in progress. TM also logs accounting data in order to support billing of users who access data stored in the federation.

If required, TM will relocate the transaction to another server to, for example, perform a conversion or finish a data transfer. After the data conversion has been finished, TM will initiate the transfer of converted data, temporarily stored on the server, to its final destination. There is one Convertor module for the conversion of each GIS data format to and from the Delta-X internal data format. All Convertors are database and GIS dependent.

The Data Mover module is responsible for sending or receiving data between the clients and servers and/or between the servers themselves. It is implemented on top of the TCP/IP protocol stack.

2.4 Network Configuration

As shown in Figure 2, Delta-X is a loosely coupled network of servers and clients. Clients and servers attached to the same LAN form a cluster. Clusters are connected to each other via a WAN, which forms the backbone of the Delta-X system. Servers and clients can also be connected directly to the WAN or, via a dial-up line, to one of the servers. The set of all clusters, clients and servers connected directly to the WAN, and clients remotely connected to cluster servers forms the Delta-X federation. Communication between clients and servers in the federation is TCP/IP-based (both over the LAN and the backbone WAN). Any client can request a transaction from any server in the federation. A dedicated server in a cluster performs name and authentication services for all clients in the cluster. A server can also perform name and authentication services for clients remotely attached to the server. Although having a dedicated server for certain functions, a client can connect to and request data from any server in the federation.

One dedicated server in the federation "the Delta-X Master ", monitors and coordinates the activities of the other servers, and holds the configuration of the Delta-X federation. When a server wants to join or leave the Delta-X federation, it must contact the Delta-X Master Server, which updates its configuration tables, and distributes them to all other in the federation. The Delta-X Master Server is duplicated for uninterrupted availability.

A server is essentially a database server which is also capable of converting data from client into an internal Delta-X format and vice versa, and storing the converted data in its database. When another client wishes to import this data, the Delta-X server translates the data into the target client's format and transfers the data to the target client. The target

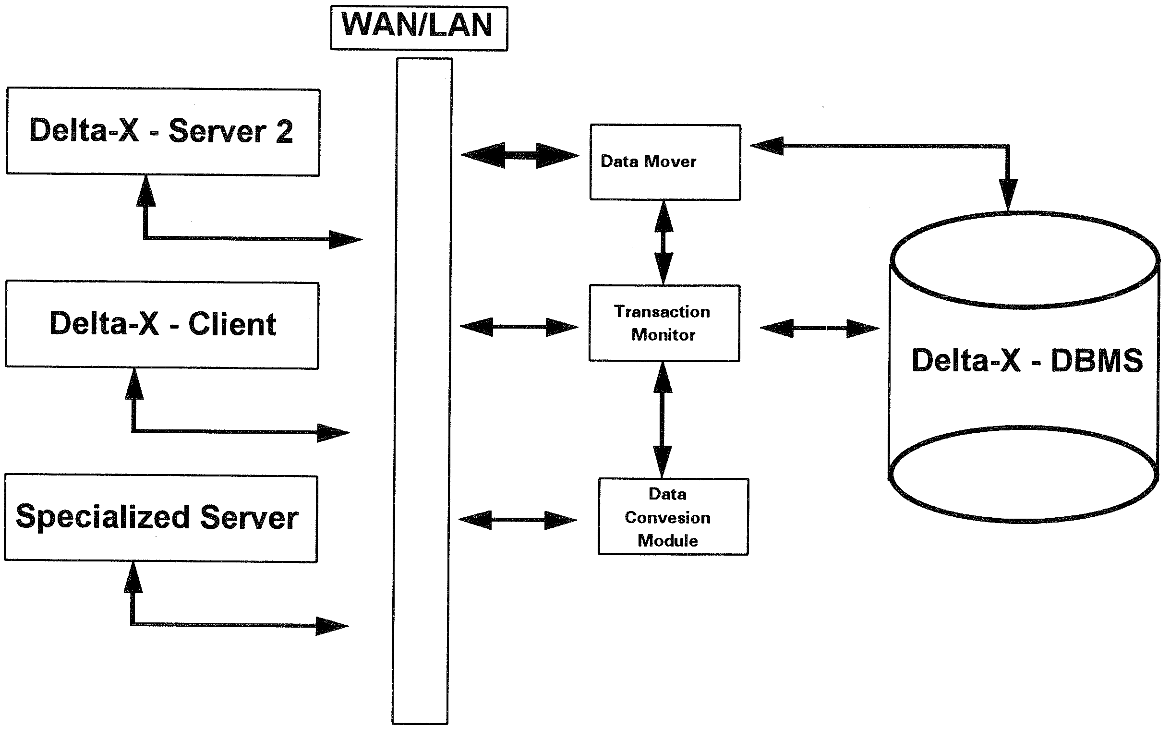

Figure 1: The Delta-X Server

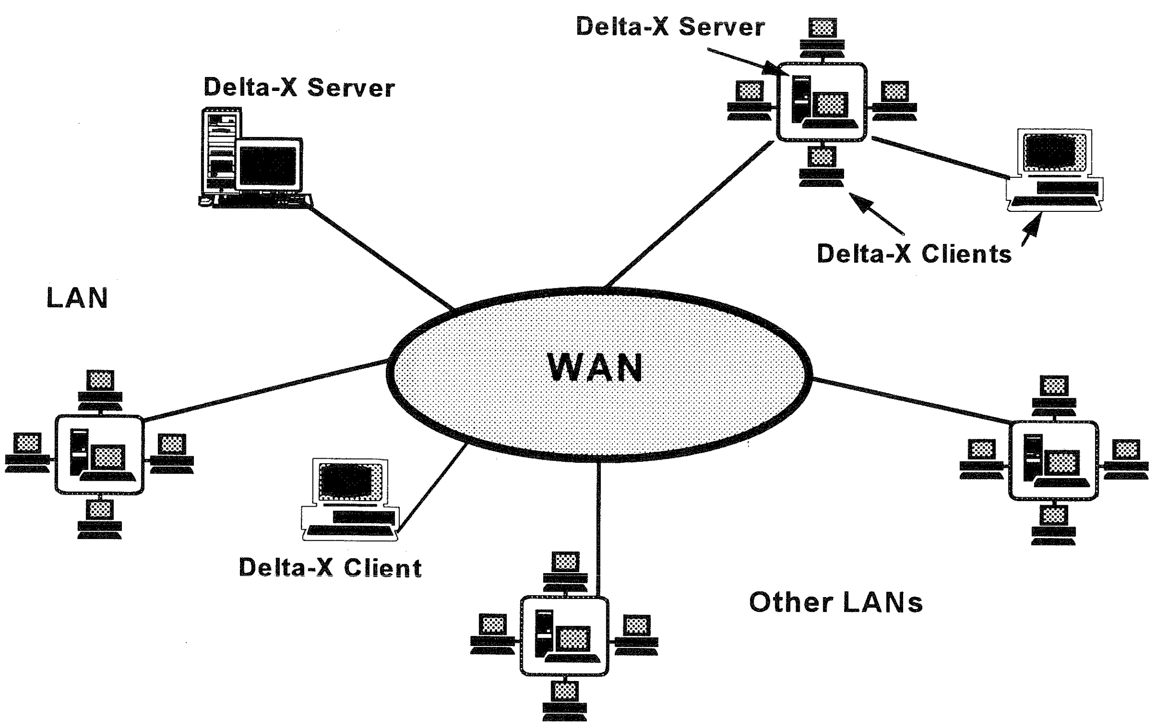

Figure 2: The Architecture of Delta-X Sytem

format can be any common GIS format and the target machine can be either a UNIX workstation or a PC.

The Delta-X server handles requests both from clients and from other servers in the federation. A transaction may start on one server, but it may migrate to another server, for example, to carry out a conversion from one data format into another, or to finish a data transfer. When a GIS wants to import or export data to Delta-X, it relinquishes control to the Delta-X client. The Delta-X client initiates the data transfer/conversion transaction. In data import operations, after a transaction has been completed, the GIS will use the data imported in the transaction.

The mode of communication between a Delta-X client and a Delta-X server or between Delta-X servers is asynchronous. Delta-X supports conversational transactions, which proceed as follows. A client connects to a server, requests a transaction, and then disconnects from the server. When the server finishes the requested transaction, or when it has to report the transaction's status, the server connects to the client and delivers requested data or reports the status.

The main reason for introducing the asynchronous communication paradigm into Delta-X is the long duration of the conversion/transfer transactions. A typical transaction can take some time to complete and it would not be reasonable to maintain a connection between a client and a server and thus to tie up network resources for the whole duration of the transaction.

The asynchronous communication between the server and the client has several other advantages. First, the server can simultaneously process multiple transactions from multiple clients without running out of communication channels. Second, asynchronous communication between Delta-X entities facilitates nesting and chaining of Delta-X transactions. The implementation of client-server and server-server asynchronous communication is based on the SUN remote procedure calls over TCP/IP.

2.5 Client Services

Delta-X client is a software that runs on the user's machine and enables Delta-X users to start and control Delta-X data conversion and movement transactions. Users can either export their data to Delta-X or import data from Delta-X. The Delta-X client also supports administrative functions, such as joining or leaving the Delta-X federation or access rights control, and viewing of data stored in Delta-X databases. Raster and vector data can be viewed. Display of data by a user prior to importing the data is important since a data conversion transaction can take a long time to complete.

As shown in Figure 3, the Delta-X client comprises the following modules: the Administration and Data Transfer and Conversion graphical user interfaces (GUIs), the Transaction Management module, and the Network Communication module.

The Administration GUI enables the user to perform administrative tasks. The administration GUI is only packaged on Delta-X clients used by Delta-X site administrators. A Delta-X client can simultaneously control multiple transactions on multiple servers. The Data Transfer and Conversion GUI also supports viewing of textual, raster and vector data to which the user has access rights and which is stored on any server in the Delta-X federation.

The network communication module performs all communication tasks with Delta-X servers. The communication functions can be either requests to view spatial data stored in one of Delta-X data repositories, or requests to perform data transfer/conversion transactions.

3. METAVIEW/GIS SPATIAL BROWSER

MetaView/GIS Spatial Browser was developed as a front end to the Delta-X to assist the GIS system users in identifying the source of data required for their applications. MetaView/GIS facilitates access to metadata of various database, e.g.: information on specific datasets, ownership, geographic coverage, format, availability, access mode, cost, etc.

MetaView/GIS is configured as a client-server model to run over Internet. The client is a user UNIX workstation, e.g. a GIS, and the server component runs on MetaView/GIS site. A number of other client-server directory services exist for locating and retrieving information across the Internet. Such directory services include Wide Area Information Server (WAIS), World-Wide Web (WWW), Archie and Gopher. WAIS, Archie and Gopher provide the user with an overview of likely places to find the desired information, and then help the user locate the specified information items. In WWW, a GUI software known as mosaic allows a user to navigate through databases with a mouse click in a hypertext mode.

MetaView/GIS differs from these systems in a number of ways. First, MetaView/GIS, operationally, serves as a partial front end of a commercial service Delta-X. Second, it provides an X11 graphical interface at the user's client which relies on commercial databases management system and not a hypertext system. Third, users can specify spatial queries. Fourth, MetaView/GIS provides more information related to a data set than those existing in these directory services. However in designing MetaView/GIS, the services provided by these systems were taken into consideration and in fact are utilized by the MetaView/GIS system. For example, if a data set is available free at a site, one may choose to retrieve it by employing the services of the other directory services.

3.1 MetaView/ GIS Operational Concepts

One of the main difference between MetaView/ GIS and the other directory services is the ability for a user to specify queries using region selection. MetaView/GIS maintains two types of databases: one contains map data, the other contains the metadata. The map database contains data for displaying maps of the world. The other database maintains the metadata, directory, catalogue, etc., of data sets in other databases. The first database is internal to the MetaView/GIS system, and users cannot therefore influence its organization and/or modification. The metadata database is however available for updates by certain category of users. There are two main users of the MetaView/GIS system; those who will maintain the data in the database and those who will interact with the database to access information about data sets and databases.

The MetaView/GIS software is organized on a client/server model. The server has the MetaView/ GIS databases. A client runs at the user's site. Queries by the users are bundled and transferred to the server, which retrieve the necessary data from the database and return it to the client. Similar to the Delta-X, the communication between the client and the server is done in an asynchronous mode over a communication network. As shown in Figure 2, the Delta-X clients and servers, are basically, the MetaView/GIS servers and clients.

3.2 MetaView/GIS Server and Clients

The MetaView/GIS server interfaces between the MetaView/GIS clients and the databases of the system. Similar to the Delta-X, the MetaView server is responsible for transaction management; this includes the monitoring and scheduling of transactions, the authentication of client requests, the administration of resources and the commit/abort of transactions and recovery. Further, other MetaView Servers has the capacity to store, execute and manage data requests, connect to a client to transmit data, communicate with other MetaView/GIS servers for assistance when necessary and authenticate MetaView/GIS privileged clients.

The MetaView client includes the GUI portion of the MV/GIS software that runs at the user's site. When MV/GIS is started it first displays its top level window as shown in Figure 4. At the top level a user has the option of connecting to Delta-X to convert data stored in one format and to transport it to users.

3.3 Queries

The Spatial Browser is used to search the database by any of the following options: keywords, data set name, category code, category name, agency/owner name, contact person name, database name, application name or project name. The Summary button gives a summary of the data sets in the database.

Visual queries involve first invoking a display of a globe, and then determining a region or an area of a globe where information is desired. A user can invoke this facility of MetaView/GIS by opening the map database. A user may zoom in on an area to facilitate more precise region selection. The selected area is interpreted as the bounding coordinates and is transmitted to the "Search region" in the Database Information Window. A user can also specify the search region manually for searches to be constraint within the region. The provision of visual spatial querying capability is one of the main distinguishing features of MetaView/GIS.

3.4 Metadata Update Functions

There are essentially, two classes of users of MetaView/GIS. A general public user and a privileged user. The privileged clients are the database owners that provide the metadata of their databases/data sets to the MetaView/GIS database. The owners of the databases or the data sets need to insert data in MetaView/GIS database, and to update existing information. MetaView/GIS provides a more flexible approach and friendly entry forms for these purposes. Updates made by authorized user are logged in temporary tables in the database. These changes are actually effected by the database administrator (DBA).

3.5 Other MetaView/GIS facilities

The main window of the browser, allows users to make request by using either and on-line or off-line tool. A user can ftp or telnet to another machine to transfer data. These systems use third party networking software for their functions. An off-line request can also be made by a FAX. The system prepares a fax report for the user based on the data set that is presently being accessed. A handy mail tool is also available for users. These functions can be accessed with the menu Request/Off-line/Fax and Request/Off-line/Mail.

Also, other MetaView spatial browser services include SQL, for advanced users who may wish to interact with the MetaView/GIS databases directly, and Delta-X for the access, retrieval of data sets from databases in the federation. MetaView provides access to other directory services.

4. CONCLUSIONS

Although the Delta-X and MetaView/GIS spatial browser work with geospatial data, its architecture, design and implementation can be applied to any type of data. The successful implementation of both systems in our Division, provided unprecedented capabilities to all the GIS workstations on our network. Our users can browse through the directory and metadata database of the federal government data holdings developed by the Inter Agency Committee on Geomatics. Data transfer between the various GIS databases and standard file formats are further proof of

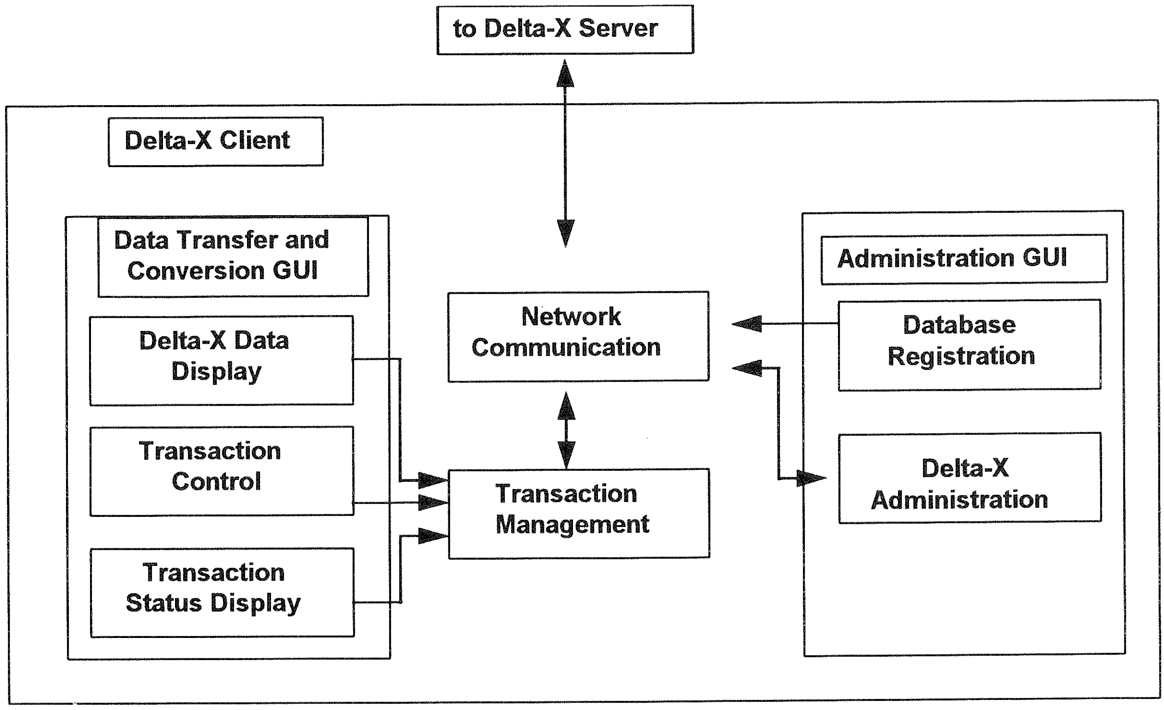

Figure 3: The Delta-X Client

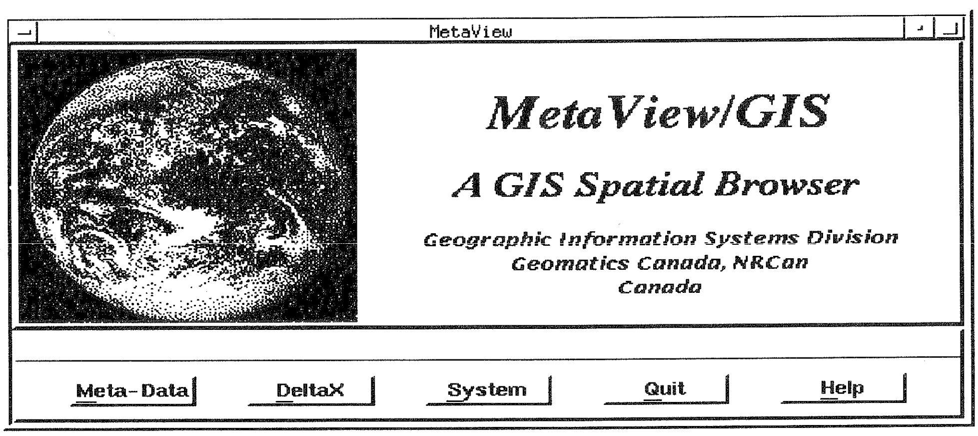

Figure 4: Top Level Window of MetaView/GIS

the validity of our concept and the success of our implementation, and the use of the Delta-X and MV/GIS in the Canadian geospatial data warehouse.

The Delta-X operational paradigms describe a methodology for the integration of multi-database management concepts and data interchange in a network of heterogeneous databases. The system as designed is a sufficiently open system that freely admits new participants, either as new databases management systems or new GIS technology, with minimum overhead. The only essential requirement is that new database or GIS vendors must provide the mapping functions that translate information between their proprietary data format and our global conceptual data model.

We have described the essential functionalities in Delta-X and we have shown the software modules required to be integrated to achieve a successful implementation. GIS database development is an exercise being carried out as national projects in several countries. We believe the approach taken by the Delta-X design will form the reference model by which similar projects will emerge to establish eventually, a global network of GIS databases.

Although our efforts are directed toward the implementation of the system in the Department, and the commercialization of the system, further developments and enhancements are in progress. The prototype implementation will be gradually extended to enhance the system security, to make Delta-X fault-tolerant and highly available by adding host stand-by servers and/or introducing data replication across the servers in the federation. Further, support for parallelism and load-balancing within the cluster and across the whole Delta-X federation is necessary.

5. REFERENCES

Allam, M., and Otoo, E., 1991, A Spatial Topographical Data Model: A First Step Towards Standardization of GIS Data Exchange. International Archives of Photogrammetry and Remote Sensing, Vol. 30, Part 2, Ottawa, June 1994.

Anklesaria, F., McCahill, M., Lindner, P., Johnson, D., John, D., Torrey, D., and Alberti, B., "The internet gopher protocol (a distributed document search and retrieval protocol)", Request for Comments (Information) RFC 1436, Internet Engineering Task Force, March 1993.

Berners-Lee, T., Cailliau, R., Groff, J.-F., and Pollermann, B., "World-wide web: the information universe", Electronic Networking: Research, Applications and Policy, vol. 2, pp. 52--58, Apr. 1992.

Boar, Bernard H.: Implementing Client/Server Computing.A Strategic Perspective. McGraw Hill, New York, 1993.

Chaly, C.K., Zhu, W., and Effah, S., 1994, MetaView: A GIS Spatial Browser - Functions and Services. International Archives of Photogrammetry and Remote Senisng, Vol. 30, Part 2, Ottawa, June 1994.

Croswell, P.L., 1991, Obstacles to GIS implementation and guidelines to increase the opportunities for success. URISA Journal, 3, pp.43-57.

Fisher, Sharon: 'Riding the Internet Highway: A complete guide to 21st Century Communication', New Riders Publishing, 1993.

Gray Jim, Reuter, Andreas: Transaction Processing: Concepts and Techniques, Morgan Kaufmann Publishers, San Mateo, California, 1993.

Masser, I., and Campbell, H., 1991, Conditions for the effective utilization of computers in urban planning in developing countries. Computers Environment and Urban Systems, 15, pp. 55-67.

Masser, I., and Campbell, H., 1992, Geographic information systems in organizations: some conditions for their effective utilization, in Lepper M. de, Scholten, H.J. and Stern, R.M. (Eds). The Added Value of Geographic Information Systems in Public and Environmental Health, Dordrecht, Kluwer.

Medved, J., Petras, J., 1994, The Client-Server Architecture of Delta-X. International Archives of Photogrammetry and Remote Senisng, Vol. 30, Part 2, Ottawa, June 1994.

Onsrud, H. and Rushton, G. (Eds), 1992. Institutions sharing geographic information Tech Report 92-5, National Centre for Geographic Information and Analysis, Santa Barbara University of California.

Otoo, E., 1994: A Federated Spatial Information System: Delta-X. International Archives of Photogrammetry and Remote Senisng, Vol. 30, Part 2, Ottawa, June 1994.

Rhind, D., 1992, Data access, charging and copyright and their implications for GIS. Int. Journ. GIS, 6, pp.13-30.

Schwartz, M.F., Emtage, A., Kahle, B., and Neuman, C.B., "A Comparison of Internet Resource Discovery Approaches", Computing Systems (The Journal of the USENIX Association), vol.5, pp. 461 - 493, Fall 1992.

Stanfill, C., "Massively parallel information retrieval for wide area information servers", in lottesvProceedings of the IEEE International Conference on Systems, Man, and Cybernetics, (Charille, Virginia), IEEE, Oct. 1991.

Tveitdal, S. and Hesjedal, O., 1989, GIS in the Nordic countries - market and technology, strategy for implementation - a Nordic approach. proc. GIS 89, Vancouver, Canada.