GEONIS - FRAMEWORK FOR GIS INTEROPERABILITY
Integration of GIS data sources in Telecom Serbia

Leonid Stoimenov, Slobodanka Djordjevic-Kajan
Faculty of electronic Engineering, University of Nis, Beogradska 14, Nis, Yugoslavia
Email: leni@elfak.ni.ac.yu, sdjordjevic@elfak.ni.ac.yu

Keywords: Interoperability, GIS, mediation, middleware integration

Abstract: This paper presents research in Geographic Information Systems interoperability. Also, paper describes our work in development, introduces interoperability framework called GeoNis, which uses proposed technologies to perform integration task between GIS applications and legacy data sources over the Internet. Our approach provides integration of distributed GIS data sources and legacy information systems in local community environment.

1 INTRODUCTION

In recent years, a large number of diverse, distributed and heterogeneous information sources (databases, knowledge bases, collections of documents, etc), are available over the Internet. The exchange of information has become a crucial factor in today's economy. Many activities in business world involve different organizations that have to work together, and use the existing information whenever is possible, in order to reach a common goal. Similar situation is also in the Geographic Information Systems and their applications.

Popularity of GIS in government and municipality institutions induce increasing amount of available information (Stoimenov et al., 2000). Information that exists in different spatial databases may be useful for many other GIS applications. But, information communities find it difficult to locate and retrieve data from other sources, in reliable and acceptable form. In such systems reuse for geodata is very often difficult process, because of poor documentation, obscure semantics, diversity of data sets, and the heterogeneity of existing systems in terms of data modelling concepts, data encoding techniques and storage structures (Devogele et al., 1998). Also, available information is always distributed and no one wants to share his own information in public without commitment. In that case, centralized control is not applicable and not practical, since the ownership of data is in domain of organizations whom they belong, and no one wants to share his own information with public (Akker and Siebes, 1996). The problem of bringing together heterogeneous and distributed information systems is known as interoperability problem.

The paper is structured as follows. In the first part, we describe related work in interoperability and mediation. The goals of our research activities, described in the third part of this paper, are defining GeoNis architecture for integration of distributed and heterogeneous GIS data sources and adding the integration technology to the existing framework. Also, this paper presents the GeoTT – a Telecom GIS application, with the special accent on its use in the telecommunication network maintenance. GeoTT uses ORHIDEA extensions of GeoNis to perform integration of different data sources in local Telecom services.

2 RELATED WORK

The research in information systems interoperability is motivated by the ever-increasing heterogeneity of the computer world. Heterogeneity in GIS is not an exception, but the complexity and richness of geographic data and the difficulty of their representation raise specific issues for GIS interoperability.

The need to share geographic information is well documented (McKee and Buehler 1996; Vckovsky, 1998; Stoimenov et al, 2000). The first attempts to obtain GIS interoperability involve the direct translation of geographic data from one vendor or standard file format into another. However, these

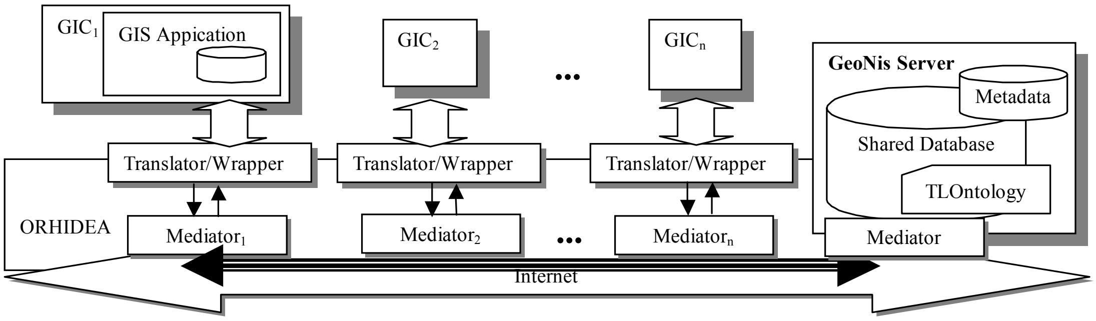

Figure 1: GeoNis framework

formats can lead to information loss. The alternatives that avoid this problem are usually more complex, like standards Spatial Data Transfer Standard (SDTS) and the Spatial Archive and Interchange Format (SAIF). A broader discussion of geographic information exchange formats can be found in (GDEStandards, 2001). One of important strategies for interoperability is conversion of different data formats in common data structure. Such data structure should be forth defined, and usually is based on one of existing GIS standards (McKee and Buehler 1996; GDEStandards, 2001).

The one important initiative to achieve GIS interoperability is the OpenGIS Consortium. This is an association looking to define a set of requirements, standards, and specifications that will support GIS interoperability. The objective is technology that will enable an application developer to use any geodata and any geoprocessing function or process available on 'the net' within a single environment and a single workflow (McKee and Buehler, 1996).

A number of proven and well-established methods exist to allow heterogeneous data sources to communicate, including federated databases and schema integration (Larson, 1998), object-oriented approaches (Chawathe et al., 1994), data warehousing (Voisard and Juergens, 1999) and mediators and ontologies (Wiederhold, 1994; Guarino, 1998; Fonseca and Egenhofer 1999; Stoimenov et al., 2000). Recent reviews of GIS interoperability and integration efforts can be found in (Abel et al., 1998; Laurini et al., 1998). Making the local geographic datasets available publicly, and establishing a common interoperability framework over shared data interchange protocols are important parts of this research. However, there are institutional and technical problems of geodata sharing and interoperability. These problems have become the focus of international research and infrastructure efforts (Onsrud and Rushton, 1995; Masser, 1998).

The mediator-based system is important for spatial data interoperability architecture (Stoimenov et al., 2002). The 3-level architecture of mediator-based systems is constructed from an application layer, and large number of information sources (heterogeneous data sources with wrappers), communicating with each other over a standard protocol (Wiederhold, 1994). A wrapper is a program that is specific to every data source (ARPA I3, 1995; Roth and Schwarz, 1997; Stoimenov et al., 1999). The wrapper extracts a set of tuples from source file and performs translation in the data source format. The most important fact is that data integration system lets users focus on specifying what they want rather than thinking about how to obtain the answers. As a result, it frees them of combining data from multiple sources, interacting with each source and finding the relevant sources.

Applying the mediator framework to the Intranet/Internet environment solves the difficult problem of gaining access to real world data sources. Internet provides the underlying communication layer and protocols for mediation of distributed systems. Mediator-based systems are constructed from a large number of relatively autonomous sources of data and services, communicating with each other over a standard protocol and enabling "on-demand" information integration. Nowadays, the mediation concept is a part of the ARPA I3 (Intelligent Information Integration) reference architecture (ARPA I3, 1995). The I3 reference architecture should be seen as a vision of how vast amount of heterogeneous information can be incrementally pulled into a gigantic, reusable library of information resources.

3 GEONIS INTEROPERABILITY FRAMEWORK

The GeoNis is a framework for interoperability of GIS applications that have to provide infrastructure for data interchange in the local community environment. Data sources are local services and offices that own geodata in some format. Specified communities have own GIS application, often created with different GIS tools and with different underlying database management systems.

To achieve interoperability, the first prerequisite is that individuals and organizations (i.e. GIC - Geographic Information Community) know each other and the data they possess. Second, there must be a willingness to make data available to users outside the source organization. Given that an organization is open to interoperability, it must announce its existence and willingness to exchange information. Then other individuals can discover the organization and assess whether there may be interest in accessing information. The goal of the GeoNis framework is to make the use of different data sources in their GIS applications simple for users.

For data exchange between GIS applications and underlying databases we define the following presumption (modified from (Levinsohn, 2000)):

- Simple - users should not have to understand all details about the data or their source system to import and use them.
- Transparent - complexities associated with data transfer should be hidden for users.
- Open - interoperability should apply to all systems, and data exchange should be independent of the used technology.
- Equal and independence - systems are equal and autonomous, and they have exclusive right to control its information and information processing without centralized control.
- Effective - data transfer should be reliable, and the resultant data should be useful for the intended purposes.
- Universal - all geospatial databases should be accessible.

Our solution is based on single, mediator-based architecture for interoperability in local community environment, OpenGIS Simple Feature as common model, and local ontologies for resolving semantic heterogeneity of data sources.

GeoNis is based on ORHIDEA (Stoimenov et al., 2000) data integration platform. The ORHIDEA platform has been developed in order to perform intelligent integration of information from multiple heterogeneous GIS (spatial and geographic), and non-spatial (thematic) data sources. The ORHIDEA is a middleware, mediator system that provides data interchange and access to data sources, distributed over the Internet, without changing how or where data is stored. The ORHIDEA platform uses agent-wrapper translators and mediator technology to allow communications between GIS applications over the Internet/Intranet.

The Mediators in ORHIDEA are “complex”, higher level component, that combine data from different data sources. The mediators should hide existence of various (heterogeneous) data sources. The mediator and agent-wrapper approach also allows maintaining these applications (data sources) and incorporate a new sources, as they became available (Stoimenov et al. 1998).

The basic architecture of GeoNis framework is shown in Figure 1. Each of GIC (i.e. local service or office) contains GIS application and corresponding (spatial) database. In addition to domain oriented GIS applications, there is one common GIS server that maintains all shared/common geographic data. Those data are public available and could be used by GIS clients in every GIC or citizens through the available public services on user demand. Data in local spatial databases are accessible in dependency of user privileges. The requests for specific data set are forward through local mediators or GeoNis server.

The GeoNis server also contains information about registered GIC and their access rights. Every new GIC who wants to participate in exchanging data must register on GeoNis server in order to allow access to his public available data and local ontology. A local ontology is referring only to public available data. After that, registered GIC have access to all available data from the other public GIC databases (with possible given rights for access), and access to shared data on GeoNis server. GeoNis uses shared ontology, one for each source of data i.e. for each GIC, and single top-level ontology located on GeoNis server.

For each data source there is a translator (or wrapper), which logically converts the basic data objects to common information model. The wrappers provide access to the data in the data sources using a common data model, in our case that is OpenGIS Simple Feature. The mediators provide coherent views of the data in the repositories by performing semantic reconciliation of the data representations provided by the wrappers. The set of information wrappers implements the middle-layer linking services. This layer performs mediator functions, which include transformation of data and mapping between data models. In order to make this logical translation, translator converts queries to requests (in XML format) through information from common model and top-level ontology. If it is necessary, query request is forwarded to GeoNis server, which resolves problem of semantic heterogeneity in used notation by ontology. The GeoNis server forwards that demands for data in XML format to appropriate data source(s) (local GIS applications). Translator on destination GIC application converts XML request to local application data model (SQL or API). The data source application may execute these requests. Also,

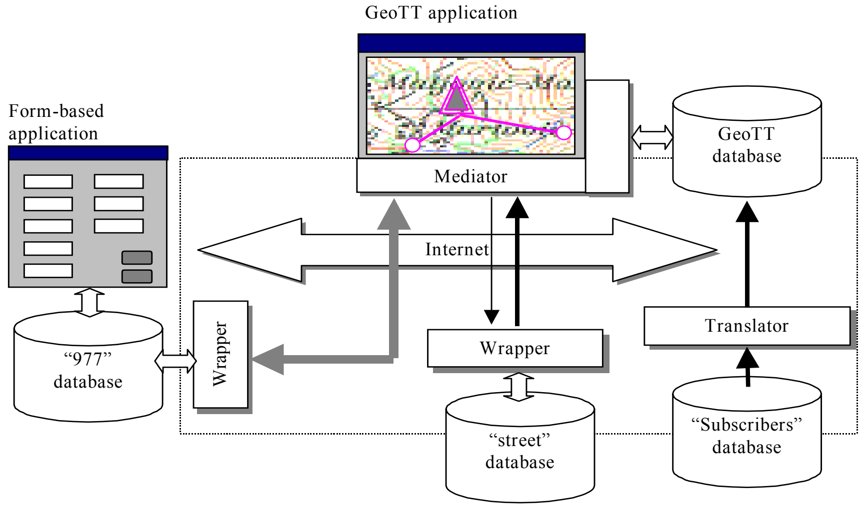

Figure 2: GeoTT integration environment

the translator converts query results from local application data model to a common model. As a result, translator/wrapper generates a GML (GML, 2001) document and then forwards data to application mediator who asked for them. The GIS application, which received data in GML format, could display them or convert them to local GIS format by translator (in keeping with defined styles).

4 GEOTT PROJECT

The National Telecommunications Company of Serbia "Telecom Srbija", department of Nis, is responsible for providing local, long distance and international telephone services. The operational area of the Department of Nis covers 4600 km2 and includes about 60 larger inhabited places. There are 65 telephone exchanges, which provide local, long distance and international telephone services for about 100000 subscribers in this area. The total TT (telephone telegraph) cable length is 995.88 km, with 960 terminal blocks and 5495 manholes, and 266356.3 km of cable pipes. About 160 people work in the Department for local network exploitation and maintenance and 10 people work in the section for technical documentation maintenance. This section maintains documentation for 172 cable routes.

The GeoTT is a specialized system for graphical inventory of the local TT networks. This system provides the following activities: (1) geographic

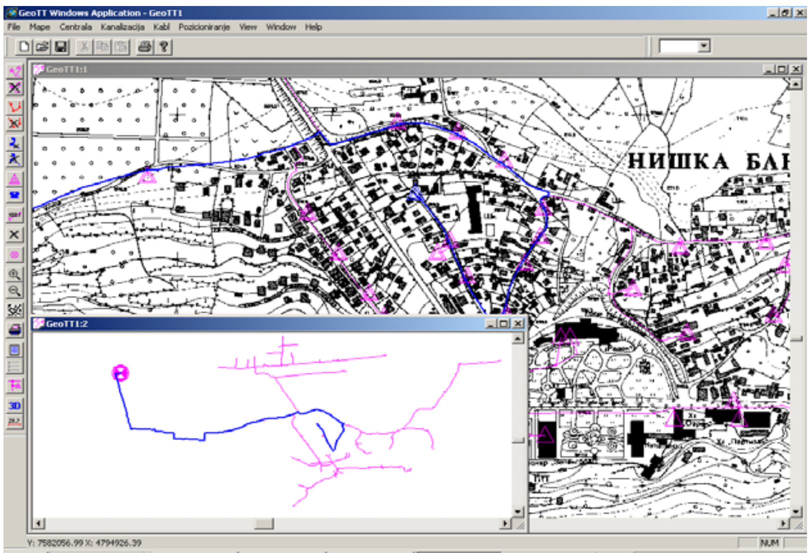

Figure 3: Route of TT cable from position of the fault to the telephone exchange

background preparing; (2) input, maintenance, retrieving, and displaying local TT network data; (3) document generation; (4) statistical report production; (5) user's requirement processing; (6) coupling GeoTT system with other information systems from the environment.

The GeoTT uses data from the other information sources in Telecom. The data sources, which should be recognized by our system, are GeoTT spatial database, relational databases and other legacy information resources in local Telecom. The data sources are distributed over a communication network such as the Internet and Intranet. The GeoTT spatial database and ontology server is based on GinisNT (Mitrovic et al., 1996), an object-relational GIS environment for the development of end-user applications.

The data stored in databases of GeoTT system involve (Figure 2):

- Scanned and georeferenced maps of the 018-network group, in 1:300000, 1:25000, 1:5000, and 1:1000 scales, located on GeoTT data server (raster maps, located in GeoTT spatial database).
- Cable network with the appropriate cable equipment, located in GeoTT (GinisNT-based) spatial database.
- Subscribers' names and addresses, so-called "988" database (988 is a telephone number of appropriate service/department in local telecom), located in relational database in "988" (Relations with Subscribers) department. Subscriber's data from this database are available through the translator.
- Photographs, video and audio records of the TT network, manual for network maintenance, and various warnings about TT network condition (local multimedia database in GeoTT spatial database).
- Data on faults and corrections on TT network, so-called "977" database, located in relational database in "977" department (Department of Fault Denunciation) of local telecom.

The TT structure is permanently evolving. The new subscribers' telephone numbers are added, new cables are laid, new exchanges are put into operation, new external terminal blocks, and connection and junction boxes are also added. All these changes are directly entered into the appropriate databases and reflect on GeoTT system behaviour.

In the sections for TT network maintenance, the GeoTT system helps with fault locating, as well as with a statistical management. The GeoTT enables operators to navigate trough entire TT network structure to pinpoint the TT network part on local maps. The operators could choose this part on the basis of any data on TT network or subscribers (for example, telephone number, external terminal block address, subscriber's name or address), or on the basis of any geographical data (coordinates or view list). Map-based screens allow the network to be viewed 'in place' on the relevant urban maps (Figure 3).

On the basis of the information from the database and the data from the measuring devices, the GeoTT system could determine particular telephone cable pair and the exact location of faults. The GeoTT system registers faults (telephone number, registration time and alphabetic description of faults). Based on the data obtained by means of the measuring device, GeoTT finds the accurate location of faults blinks this position and highlights cable route from the telephone exchange to the position of the faults (Figure 3). GeoTT system provides paper map of the interested area marking the position of the faults.

All the changes in the technical documentation made in the process of removing faults are evidenced and stored into the database. The work order that is not closed indicates that the fault is being removed or that the changes made are not evidenced.

The GeoTT system can respond quickly to user's queries and subscriber's complaints as well as help the technician with making decision. GeoTT helps in evidencing and collecting changes, which are made in the process of TT network reconstruction, or in removing faults.

5 CONCLUSION

We present here an ongoing case study and development of the framework for interoperability of GIS applications. This framework is aimed to resolve interoperability problem in the local, municipality environment. The project introduced a mediation-based system designed to allow data source reusability, system scalability and extendibility. The principles behind the ontology/mediation framework described in this paper are extensibility, relative autonomy of infrastructure nodes, and universal access to heterogeneous data sources from a variety of portals. The system should provide actualisation of client/server applications, using the Internet and Web technologies as under-layer for the network service and the integration of distributed data sources. Mediators are efficient and cheap way of integrating data from heterogeneous information systems. Logical consequence is that mediators provide evolving system development and use the existing software investments in software and databases.

The GeoTT system presented in this paper is intended to input, store, retrieve and show data on the TT network. The GeoTT is open, object-oriented, specialized information system based on a new accomplishment in the GIS technology and mediator and ontology-based interoperability. The tool, which could be effectively used in the TT network evidencing, management and maintenance is obtained thanks to these technologies.

In the future, we will focus on research and technical problems of spatial mediation on the global scale, describe the main components of the interoperability architecture, and make decision about ontology internal representation (ontology representation language) and implementation.

REFERENCES

Abel D.J., et al., 1998, Towards Integrated Geographical Information Processing, *International Journal of Geographic Information Science*, Vol. 12, No. 4, 1998, pp.353-371.

van den Akker J., A. Siebes, 1996, DEGAS: A temporal active data model based on object autonomy, *Computer Science/Department of Algorithmic and Architecture, CWI Technical report, CS-R9608*, Amsterdam, Netherlands, 1996.

ARPA I3, 1995, *Reference Architecture for the Intelligent Integration of Information*, ARPA, Prepared by the Program on Intelligence Integration of Information (I3), version 1.0.1, May 19, 1995.

Chawathe S., Garcia-Molina H., Hammer J., Ireland K., Papakonstantinou Y., Ulman J., Widom J., 1994, The TSIMMIS Project: Integration of Heterogeneous Information Systems, *In the Proceedings of ISPJ Conference*, Tokyo, Japan, pp.7-18, October 1994.

Devogele T., Parent C., Spaccapietra S., 1998, On Spatial Database Integration, *International Journal of Geographic Information Science*, Vol. 12, No. 4, 1998, pp.335-352.

Fonseca F., Egenhofer M., 1999, Ontology-Driven Information Systems, *7th ACM Symposium on Advances in GIS, Medeiros C.B. (ed.)*, Kansas City, MO, 1999, pp.14-19.

Geographic Data Exchange Standards, 2001, http://www.diffuse.org/gis.html.

GML, 2001, *OGC GML 2.0 Recommendation paper*. Open GIS Consortium, 2001. http://www.opengis.net/gml/01-029/GML2.html

Guarino, N., 1998, Formal Ontology in Information Systems, *In N. Guarino (ed.) Formal Ontology in Information Systems. Proceedings of FOIS'98*, Trento, Italy, 6-8 June 1998. IOS Press, Amsterdam, pp.3-15.

Larson A.P., 1998, Federated Databases Systems for Managing Distributed, Heterogenous, and Antonomous Databases, *ACM Computing Surveys,* Vol 22, 1998. pp.183-236.

Laurini R., et al., 1998, Spatial multi-database topological continuity and indexing: a step towards seamless GIS data interoperability, *International Journal of Geographic Information Science*, Vol. 12, No. 4, 1998, pp.373-402.

Levinsohn A., 2000, Geospatial Interoperability: The Holy Grail of GIS, *GeoEurope*, http://www.geoplace.com/gw/2000/1000/1000data.asp

Masser, I., 1998, *Governments and Geographic Information*, Taylor&Francis: London, 1998.

McKee L., and Buehler R., 1996, The *Open GIS Guide*, OpenGIS Consortium, Inc, 1996, http://www.opengis.org/

Mitrovic D., Mitrovic A., Djordjevic-Kajan S., A., 1996, scalable, object-oriented GIS Framework, *In Proc. of ISPRS Int. Workshop "Hardware and Software Aspects of GIS"*, Milan Italy, pp. 20-21, 1996.

Onsrud, H., Rushton G., 1995, Eds. *Sharing Geographic Information*, Rutgers University Press: New Brunswick, NJ, 1995.

Roth M.T., Schwarz P., 1997, Don't Scrap It, Wrap It! A Wrapper Architecture for Legacy Data Sources, *In: Proceedings of the 23rd VLDB Conference*, Athens, Greece, 1997, pp.226-275.

Stoimenov L., Mitrovic A., Djordjevic-Kajan S., 1998, An Object-Relational Environment For Developing Object-Oriented GIS Applications Based On Mediator, *16th IASTED International Conference on Applied Informatics,* Garmisch- Partenkirchen, Germany, pp.168-171.

Stoimenov L., Mitrovic A., Djordjevic-Kajan S., Mitrovic D., 1999, Bridging objects and relations: a mediator for an OO front-end to RDBMSs, *Information and Software Technology*, Elsevier, 1999, Vol 41, No. 2, pp. 59-68.

Stoimenov L., Djordjević-Kajan S., Stojanović D., 2000, Integration of GIS Data Sources over the Internet Using Mediator and Wrapper Technology, *MELECON 2000*, Cyprys, 2000.

Stoimenov L., Djordjevic-Kajan S., Milosavljevic A., 2002, Realization of Infrastructure for GIS Interoperability, *YUINFO '02*, Kopaonik, 2002, (in serbian).

Vckovsky A., 1998, *International Journal of Geographic Information Science - Special Issue: Interoperability in GIS*, Vol 12, No 4, 1998.

Voisard, A., Juergens, M., 1999, Geographic Information Extraction: Querying or Quarrying?, *In Interoperating Geographic Information Systems, M. Goodchild, M. Egenhofer, R. Fegeas and C. Kottman (Eds.)*, Kluwer Academic Publishers, New York, 1998.

Wiederhold, G., 1994, Interoperation, Mediation and Ontologies, *Int. Symp., on Fifth Generation Computer Systems*, Tokyo, Japan, 1994.