Geodata integration in distributed environment using OLE DB data provider

Aleksandar Stanimirović[1], Leonid Stoimenov[2] and Slobodanka Đorđević-Kajan[3]

Abstract - **This paper describes implementation of OpenGIS standards for uniform access to heterogeneous and distributed data sources. These standards are based on Microsoft Universal Data Access specification and OLE DB technology. The basic architecture of Ginis OLE DB Data Provider and schema definition for GIS data is also shown in this paper.**

Keywords – GIS, OLE DB, OpenGIS, Ginis, provider

I. Introduction

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. Interoperability means openness in the software industry, because open publication of internal data structures allows GIS users to build applications that integrate software components from different developers.

Integrating geodata from various sources increasingly becomes important because of growing environmental concerns, pressures on governments and businesses to perform more efficiently, and simply because of the existence of a rapidly growing body of useful geodata and geoprocessing tools [1] Increasing number of geodata producers and users have expressed the need for the integration of geodata from distributed information sources and for interoperable GISs [2]. The systems that own this data must be capable of interoperation with systems around them, in order to make access to this data to become feasible.

Process of designing that kind of systems can become very complicated. It is very useful, in order to avoid this complication, to realize data access in separate component. In this way, application becomes completely independent from underlying data access technique. That kind of components will provide means for development of new applications, capable to interchange data with systems from distributed environment. At the same time this component must provide standard interfaces for data access. In that way existing applications can also use this component without any change.

[1]Aleksandar Stanimirović is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Serbian and Montenegro, E-mail: alex@elfak.ni.ac.yu

[2]Leonid Stoimenov is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Serbian and Montenegro, E-mail: leni@elfak.ni.ac.yu

[3]Slobodanka Đorđević-Kajan is with the Faculty of Electronic Engineering, Beogradska 14, 18000 Serbian and Montenegro, E-mail: sdjordjevic@elfak.ni.ac.yu

The paper is structured as follows. In the second section, we shortly describe related work. The goals of our research activities and `GinisFrame` component framework along with are described in the third section. Fourth section of this paper describes our efforts in defining schema definition for GIS metadata information.

II. Related work

In the past few years the OpenGIS Consortium (OGC), as a consortium of GIS vendors, agencies, and academic institutions, has emerged as a major force in the trend to openness. According to OpenGIS Abstract Specification for Features and OpenGIS Simple Feature Specification access to spatial data should be seen in the context of database problems [3]. Spatial geometry should be seen as one of the object attributes and feature collections could be easily represented as tables in relational database. Several feature collections (tables) also form one relational database.

This point of view is supported by current trends in database systems. Modern relational database management systems introduce new technologies for spatial data storage. OpenGIS Consortium has proposed OpenGIS Simple Features Specification For SQL [4] to standardize storing geo-spatial data in relational databases. Unfortunately there exists large base of spatial data stored in proprietary formats (mostly binary data in various GIS formats). Chances that this data will be converted to some standard database format are very small. For this reason OpenGIS Simple Feature Specification For OLE/COM provides unified method for accessing data spatial data stored in both databases and proprietary formats. This technology is based on Microsoft OLE DB standards. This standard provides data in unified table-oriented manner to any system or application.

Universal Data Access (UDA) is the Microsoft strategy for providing access to all types of data across information system [5]. Microsoft UDA provides access to a variety of information sources including relational and non-relational data (e-mail and file system stores, text, graphical, geographical data and more).

The Universal Data Access provides component architecture. This kind of architecture allows components to implement only required set of functionalities over data sources. UDA also assumes existence of service components that implements additional functionalities on the top of less capable components.

The Microsoft UDA strategy is based on OLE DB. Microsoft defines OLE DB as a specification for a set of data access interfaces [6]. This set of interfaces expose data from a variety of sources using OLE Component Object Model (COM). These interfaces provide applications with uniform access to data stored in diverse information sources.

OLE DB architecture distinguish three kinds of database components:

- *Data providers* – components that represent various data sources. Providers expose information uniformly using a common abstraction called the rowset,
- *Data consumer* – applications or tools that consumes data exposed by data providers and
- *Data services* – components that consume and produce data at the same time. Data services retrieve data from existing providers, transform data and expose it, using the same set of interfaces, to other components.

The Data provider is basic component that must be implemented in order to allow data to be exposed to and shared among different applications. Data providers are only required to implement minimal set of OLE DB interfaces. Data providers wishing to allow simple creation and modification of data must implement additional OLE DB interfaces. Sophisticated data providers may also expose custom COM interfaces in order to implement functionalities specific for their data sources.

III. GINISFRAME – COMPONENT BASED FRAMEWORK

Component-oriented programming is the predominant software development methodology today [7][8]. Components can be considered as stand-alone service providers. When a system needs some service, it calls on a component to provide that service. In that case system doesn't take care about where component is executing or about programming language that was used for component development. Component could be considered as a completely independent executable entity. Components publish their interfaces and all interactions are through that interfaces. Components may exist at different level of abstraction, from a simple library subroutine to an entire application.

Basic architecture of **GinisFrame** component framework [9] is shown in Fig. 1. This framework is part of **GeoNis** framework for interoperability of GIS applications. **GinisFrame** component framework was used for development of **GeoJP** application, which is part of **GeoNis** framework.

GinisFrame is component-based framework developed in CG&GIS Laboratory at Faculty of Electronic Engineering in Nis. This framework provides greater flexibility and efficiency in development of desktop and Internet GIS applications. GIS applications can utilize services provided by components from framework using standard interfaces. In this way GIS application is completely independent from implementation of a component.

Ginis OLE DB Provider is part of **GinisFrame** component framework, and is responsible to provide access to diverse data sources. In order to achieve this functionality this component use existing commercial OLE DB providers. This component is also responsible for providing data consumers with GIS specific functionality.

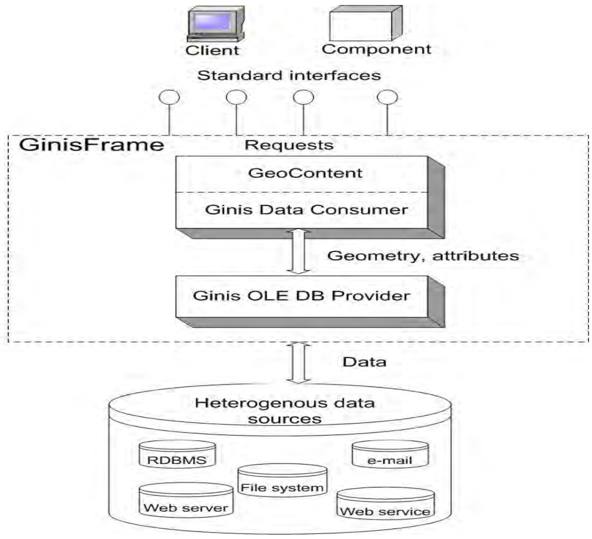

Fig. 1 Architecture of GinisFrame component framework

The basic architecture of **Ginis OLE DB Provider** component [9][10] is shown in Fig. 2.

GIS OLE DB provider consists of three components:

1. **Data Manager Component** – read and writes data from various data sources,
2. **OLE DB Component** - implements Data Source, Session, Command and Rowset objects that expose standard interfaces to data consumers according to Microsoft specification and
3. **GIS DB Component** – provides GIS specific functionality extending or implementing new interfaces besides interfaces defined by Microsoft specification.

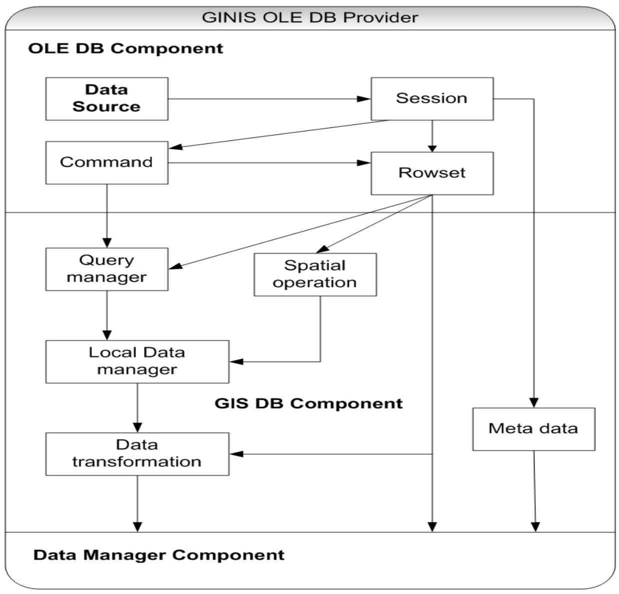

Fig. 2 Basic architecture of Ginis OLE DB Provider component

OLE DB standard defines facilities by which consumers and providers communicate. In order to achieve interoperability OGC has defined standard that allows GIS consumers and GIS providers to exchange GIS information (OpenGIS Consortium Inc., 1999). This standard makes a GIS data consumer capable to determine the GIS capabilities of a GIS data provider. Using this standard GIS data consumer can also retrieve GIS data in predictable manner.

These standards involve:

- *Open GIS data provider registry entries* – data provider must register his support for GIS data so that consumer can distinguish him from common data providers,
- *GIS metadata* – required and optional GIS metadata that describes to clients that GIS data exists and how to get it,
- *IColumnsRowset* – additional GIS columns are defined for the Rowset returned by calling GetColumnsRowset method,
- *Geometry* – methods are described for acquiring geometry from Column or Field as WKBGeometry,
- *Spatial Reference Information* – methods are described for acquiring Spatial Reference Information from Session and Rowset objects and
- *Spatial Filter* – standard spatial filter parameters are defined for use with Command object. Parameters are spatial filter, spatial operator and Geometry Field/Column name.

Although this standard is needed for communication between GIS consumer and GIS provider there is no strict level of compliance for Open GIS OLE DB providers. The minimum level of support is that a provider must provide registry entries for Open GIS OLE DB provider. Clients are supposed to use standard techniques to determine which functionalities provider supports and to take actions according to that.

V GIS Metadata

`Ginis OLE DB Provider` component must provide additional GIS Metadata information in order to tell consumers that GIS data exist and to explain how to get this data and how to interpret it. This additional metadata contains following information:

- which tables are considered GIS *feature tables*,
- which columns contain geometry, type and spatial reference of the geometry,
- spatial references of the Data Source and
- spatial operators supported by the provider.

According to OpenGIS Simple Feature Specification For SQL Simple geospatial feature collections conceptually can be stored as tables with geometry valued columns in a Relational DBMS (RDBMS). Each feature will be stored as a row in a table. A table whose rows represent Open GIS features is referred to as a *feature table*. In order to support this specification OpenGIS Simple Feature Specification For OLE/COM has extended schema information provided by standard OLE DB provider.

`Ginis OLE DB Provider component` contains three additional *SchemaRowsets* and one additional *PropertySet*:

- *DBSCHEMA_OGIS_FEATURE_TABLES Rowset* - this rowset indicates those tables that the consumer can query as features,
- *DBSCHEMA_OGIS_GEOMETRY_COLUMNS Rowset* - this rowset identifies the feature columns in the *feature table* that contains geometry information,
- DBSCHEMA_OGIS_SPATIAL_REF_SYSTEMS Rowset – contains information of *Spatial Referent Systems* supported by data provider and
- *OGIS Property Set* – contains information of spatial operator supported by data provider.

In order to provide this kind of information to consumers **`Ginis OLE DB Provider`** must obtain this data in the first place. If GIS application use **`Ginis OLE DB Provider`** on the top of relational database, designed according to OpenGIS Simple Feature Specification For SQL, this data can be obtained from additional tables that extends standard database Schema Information. But in **`Ginis OLE DB Provider`** cannot rely on this.

Modern GIS applications have to deal with distributed data sources. *Feature table* is not necessary located in one data source. Columns that forms *feature table* can be distributed in great number of data sources. Also *feature* data is not necessary in relational databases. Spatial data can be stored in great number of heterogeneous data sources. Ginis OLE DB provider must access all this data sources in order to build up *feature table* and to provide data from it to his consumers. In this situation obtaining GIS metadata information is very difficult and sometimes impossible.

In order to solve this problem we extended **`Ginis OLE DB Provider`** in the way that he can receive GIS metadata information from consumer during his initialization. **`Ginis OLE DB Provider`** still fully supports OpenGIS Simple Feature Specification For SQL. When `Ginis OLE DB Provider` is used for accessing data in relational database, designed according to this specification, additional schema information from database can be used if not specified different. In all other cases **`Ginis OLE DB Provider`** relies on data provided by consumer during initialization.

`Ginis OLE DB Provider` expects this GIS metadata information in the form of XML document. This XML document is built according to schema in Fig. 3.

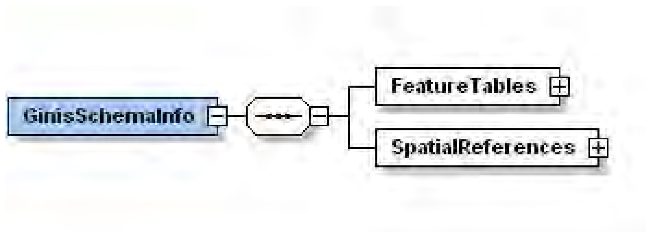

Fig. 3 Schema for document with GIS Metadata information

Every schema element contains information specified in OpenGIS specification [3] plus some additional information that are specific for **`Ginis OLE DB Provider`** component.

Metadata XML document contains collection of *feature table* and *spatial reference system* definitions. *Spatial reference system* is defined according to schema in Fig. 4.

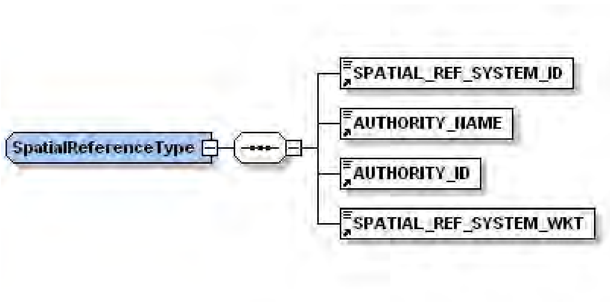

Fig. 4 Definition of Spatial Reference System

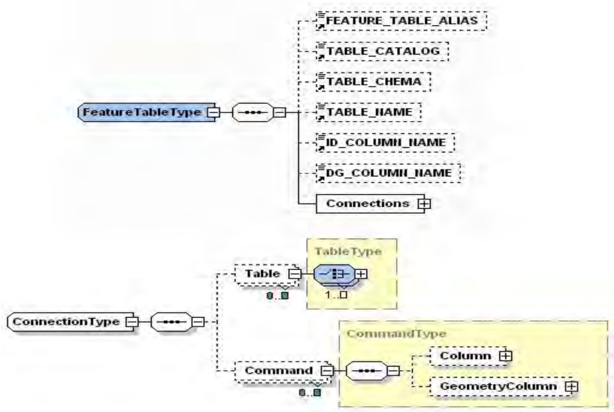

Fig. 5 Definitions of Feature table and Connections

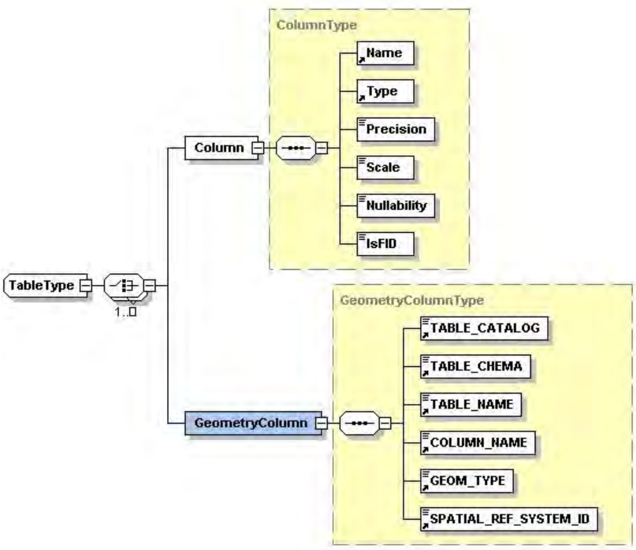

Fig. 6 Definition of columns and geometry columns

Every *feature table* is defined as a collection of connections. Each connection defines one data source and contains tables from that data source or commands that can be executed against data source. Each table or command contains collection of columns and geometry columns.

Schema for this definition is shown in Fig. 5 and Fig. 6. Feature table is treated as distributed table partitioned vertical. Every distributed source only has certain columns of *feature table*. Every vertical fragment includes primary key so that full feature table can be reconstructed. For this reason some columns can be defined to participate as a part of feature identifier (FID).

VI CONCLUSION

To provide integrated access to various distributed geo-information sources, we have developed Ginis OLE DB Provider component. This component is a part of GinisFrame component framework. Component is based on OLE DB technology and OpenGIS specification. OLE DB technology provides system architecture model related to interoperability. This model specifies uniform paradigm for accessing spatial and non-spatial data in distributed data sources. We also developed schema definition for specifying data sources and metadata information in distributed environment.

ACKNOWLEDGEMENT

This research was partially supported by the project “Geographic Information System for Local Authorities based on Internet/WWW Technologies”, funded by Ministry of Science, Technology and Development, Republic of Serbia, Contract No. IT.1.23.0249A.

REFERENCES

[1] Buehler R., and McKee L., 1998, The Open GIS Guide, Third Edition, OpenGIS Consortium, Inc, http://www.OpenGIS.org/techno/specs.htm

[2] Stoimenov L., and Đorđević-Kajan S., 2002. Framework for semantic GIS interoperability, FACTA Universitatis (Niš), Series: Mathematic and Informatics, 17(2002), pp.107-125.

[3] OpenGIS Consortium Inc, OpenGIS Simple Features Specification For OLE/COM, 1999, http://www.opengis.org

[4] OpenGIS Consortium Inc, OpenGIS Simple Features Specification For SQL, 1999, http://www.opengis.org

[5] Rauch S., Manage Data from Myriad Sources with the Universal Data Access Interfaces, Microsoft System Journal, September 1997

[6] Microsoft Press, Microsoft OLE DB 2.0 Programmer’s Reference and Data Access SDK, 1998

[7] Stojanovic Z., and Dahanayake A.N.W., 2002, "A New Approach to Components", accepted for the 2002 Information Resources Management Association (IRMA) International Conference, to be held in Seattle Washington, USA, May 19-22, 2002.

[8] Szyperski, C., 1998, "Component Software: Beyond Object Oriented Programming", ACM Press, Addison-Wesley.

[9] Stoimenov L., Stanimirović A., Đorđević-Kajan S., Realization of Component-Based GIS Application Framework, Accepted for 7th Agile Conference on Geographic Information Science, Heraklion, Crete, 29 April-1 May, 2004

[10] Stanimirović A., Đorđević-Kajan S., Stoimenov L., Ginis OLE DB data provider, accepted for Etran conference, Čačak, May 2004