ABSTRACT
Successful information integration and sharing data across disparate systems and designs are required for fast access to and interpretation of many types of geospatial information. Spatial data are highly heterogeneous - not only they differ from data representation and storage methods, but they also differ in the way of querying the data. Finding and accessing spatial data in an environment like this is a crucial task. Enterprise geographic information system (E-GIS) is an organization-wide approach to GIS implementation, operation, and management. The main focus of the paper is to integrate diverse spatial data repositories for geographic applications using service-based methodology. We have adopted service-oriented architecture (SOA) for the discovery and retrieval of geospatial data. The architecture uses a central ontology as metadata information, which acts as service broker. Ontology-based discovery and retrieval of geographic data solves the problem of semantic heterogeneity, the major bottleneck for spatial interoperability. The implementation is in compliant with the Web Map Service (WMS) and Web Feature Service (WFS), the web service standards proposed by OGC. The need for loosely coupled service-based access of data in the spatial domain has been exploited. A query processing mechanism in distributed environment of spatial data sources has been discussed at the end. The proposed system has been implemented and fully tested.

Categories and Subject Descriptors
H.2.5 [Heterogeneous Databases]: Data translation; H.3.5 [Online Information Services]: Data sharing, Web-based services.

General Terms
Management, Design, Languages

Keywords
Geospatial web service, Ontology, Service Oriented Architecture, Spatial Data Interoperability

1. INTRODUCTION
Spatial data is an important part for any Decision Support System (DSS), not just in Earth Sciences. Spatial information is the key to effective planning and decision-making in a variety of application domains. Today, there is a huge amount of data gathered about the Earth, computers throughout the world are connected, and the use of GIS has become widespread. Although spatial information systems have been characterized as an integration tool, GIS interoperability is far from being fully operational.

Since most of the GISs are not originally designed to work in cooperation, several interoperability problems arise while integrating these diverse spatial data sources. Each GIS provides its own proprietary data format as well as its specific query language [2]. In addition, geographic resources are designed for a variety of different purposes. Orthogonal directions in the design of geographic resources may affect the semantics of the data they contain and impair their integration [3, 4, 5]. These discrepancies make the integration of different geographic resources significantly complex.

Enterprise geographic information system (E-GIS) is an organization-wide approach to GIS implementation, operation, and management. An E-GIS provides access to the shared geospatial information and analysis resources for a large number of concurrent users from different locations [22]. E-GIS can also be defined as an effort to design integrated geo-spatial management techniques to serve a complex institution. The problem of bringing together heterogeneous and distributed computer systems is known as interoperability. Interoperability has to be provided on a technical and informational level. In short, information sharing not only needs to provide full accessibility to the data, it also requires that the accessed data can be processed and interpreted by the remote system.

One important initiative to achieve GIS interoperability has been taken by Open Geospatial Consortium (OGC) [6]. This is an association of government agencies, research organizations, software developers, and systems integrators, looking to define a set of requirements, standards, and specifications that will support GIS interoperability. The specifications provided by OGC enable syntactic interoperability and cataloguing of geographic information. However, although the OGC defined catalog support discovery, organization and access of geographic information, they do not define any method for overcoming semantic heterogeneity problem. The semantic heterogeneity problem presents a major challenge in spatial data discovery and retrieval in an open and distributed environment of geospatial data domain.
In this paper we propose an approach that focuses on the spatial data integration based on Service Oriented Architecture (SOA) using web services for integrating diverse repositories of spatial data. Providing data repositories as services over the Internet makes the integration more flexible, in the sense that a data repository can be configured to the Service Brokering agency with minimum possible bindings making it less coupled to the integrated system. Two standard Web Service techniques proposed by OGC - Web Feature Service (WFS) [8] and Web Map Service (WMS) [10], have been incorporated effectively in developing a service-oriented integrated system for spatial data repositories. Also with SOA technique the query capability of the integrated system gets enhanced, and the underlying inherent complexity for query processing can be abstracted from the user.

The paper focuses on the task of service discovery and retrieval, which is a crucial task in the open and distributed environment of spatial web services. Often, effective service discovery requires an extensive search for appropriate services across multiple application domains. Catalogs support discovery, organization, and access of geographic information and thus help the user to find information that exists.

The query capability of the system has also been extended to that offered by OGC-compliant catalogs. The query mechanism uses XML&KVP (Keyword Value Pair), a SQL like query language derived from the OGC CQL (Catalog Query Language).

The rest of the paper is organized as follows: Section 2 gives some motivating examples for spatial data integration. Section 3 gives a brief description of general service oriented architecture and its use in spatial domain. Ontology-based discovery and retrieval mechanism has been discussed in section 4. Section 5 gives a brief description of the SOA nature of the integrated system. Finally, a conclusion is drawn in section 6.

2. RELATED WORK

Current efforts to integrate geographic information embrace the idea of metadata standards as the key to information sharing and analysis. These include the Federal Geographic Data Committee (FGDC) and the National Spatial Data Infrastructure (NSDI), Geospatial One-Stop, and the U.S. Geological Survey’s The National Map as well as standards from the International Standards Organization (ISO) for geospatial metadata [11]. The NSDI attempts to bring together geographical information sources from all levels of government and other organizations into a single point of entry for easier access to data.

Several ways of integrating the spatial data repositories have been proposed in the literature. Most current approach for an integrated co-operative interoperable Spatial Information System or E-GIS is warehousing the data from heterogeneous repositories in a clean and consistent form, which takes care of both the semantic and syntactic heterogeneities among the data sources. [12] proposes a spatial data warehouse based technique and employs middleware technology for data exchange from the spatial data warehouse.

There are some works [13, 14] in the geo-spatial domain using Open standards proposed by OGC. An approach is proposed in [13], which uses a WFS [8] based mediation approach with the help of derived wrappers. It provides an approach of a multi-tier client-server architecture and uses standard WFS wrappers to access data. The wrappers are further extended by derived wrappers that capture additional query capabilities. Although they offer an efficient integration and querying mechanism among heterogeneous data repositories, they don’t offer the effectiveness of service-based technology in heterogeneous system. In this architecture the data sources are tightly coupled with the mediator.

In this paper an SOA architecture using Web Service has been proposed. An initiative has been taken towards enterprise-wide integration and sharing of spatial data. The advantages and benefits of a service method have been fully exploited for designing an E-GIS architecture. The discovery and retrieval methodology of SOA makes the data repositories less coupled to each other making the system maintenance more cost effective.

3. SOA AND SPATIAL DATA

3.1 General SOA methodology

The emerging Service Oriented Architecture (SOA) based method using Web Services is gaining lots of interest in the way of seamless integration of information systems spreading across several organizations. As shown in Fig. 1, taken from [15], service oriented architectures involve three different kinds of actors: service providers, service requesters and discovery agencies. The service provider exposes some software functionality as a service to its clients. In order to allow clients to access the service, the provider also has to publish a description of the service. Since service provider and service requester usually do not know each other in advance, the service descriptions are published via specialized discovery agencies. They can categorize the service descriptions and provide them in response to a query issued by one of the service requesters. As soon as the service requester finds a suitable service description for its requirements at the agency, it can start interacting with the provider and using the service.

![Figure 1. Basic Methodology of Service Oriented Architecture (SOA)](image)

3.2 Geospatial domain

The open and distributed GI domain opens a wide range of possibilities for acquiring, processing and analyzing geographic information without the need of GIS expert knowledge. In an environment where services are previously unknown, a service that is appropriate for answering a given question from among a large number of available services has to be discovered first. Service discovery, thus, is a crucial task that will become even more important with the emerging Semantic Geospatial Web.

Although the SOA technology using Web Services is evolving in the information system domain, where legacy or already existing systems are integrated to form an organization-wide information system, the application of SOA in comparatively highly...
heterogeneous spatial data domain has not been tried much. Only [2, 16] have provided a service-based methodology for Geographic Information (GI) integration. Keeping in mind the highly heterogeneous spatial data, the complex query mechanism required for processing them, a service-based technique could possibly be one of the best solutions for co-operative integration of spatial data.

Two major problems that exists in highly heterogeneous geospatial domain are as follows:

- Improper or insufficient documentation makes it difficult for outside users to discover the data sets those are useful for their task.
- Datasets in different format often requires to be converted in order to be used in other system. This problem is taken care of (as suggested by OGC) by providing the data in a vendor-neutral formats like GML [9].

OGC WFS provides a set of protocols to provide standardized service interfaces for the geospatial data sources. Through this services distributed geospatial data can be accessed and processed across administrative and organizational boundaries. As the data sources are less coupled to the integrated system, they can be created and managed locally, which leads to increasing quality and efficiency. The integrated system can easily be extended to include new services and/or data sets.

4. SEMANTIC HETEROGENEITY AND ONTOLOGY

Although standards forming bodies like the OGC provide the basis for syntactic interoperability the usability of information that is created in one context is often of limited use in another context because of insufficient means for meaningful interpretation. This problem is referred to as the need for “semantic interoperability among autonomous and heterogeneous systems”. Problems caused by semantic heterogeneous descriptions play a crucial role during the task of finding relevant information within a GI web service environment [17]. In [2], semantic heterogeneity is defined as the consequence of different conceptualizations and database representations of a real world fact.

4.1 Central Catalog

A central catalog is absolutely necessary in an integrated system where services and service consumers are arbitrarily distributed and unknown of each other. A catalog’s task is to allow a service consumer to find and access resources at the service provider’s end. This information can be described as metadata information for the system. In general each service provider has to register or publish its offerings by means of metadata to the catalog to enable it being discovered by a service consumer.

The metadata-based approach for catalog definition may lead to problems when querying and interpreting search results of different distributed catalogs. This may lead to semantic heterogeneity problem existing among diverse data services. Current free-text search in catalogs is not sufficient to overcome the semantic heterogeneity.

4.1.1 Ontology-based Catalog

Accepting the diversity of geographic data domain, explicit context models can be used to reinterpret information. Ontologies have become popular in conceptual modeling of a system as they can be used to explicate contextual information. The term “ontology” has been used in information sciences with several meanings. [18] introduced the term “ontology” to mean an “explicit specification of a conceptualization”. In order to make the ontology machine-readable, it has to be formalized in some representation language.

Explication of knowledge by means of ontologies is a possible approach to overcome the problem of semantic heterogeneity. The Bremen University Semantic Translator for Enhanced Retrieval (BUSTER) [1] has addressed the task of ontology-based information brokering. Among other things, this system provides a hybrid ontology-based approach [7] with logical reasoning on metadata for retrieving information sources.

4.1.2 Structure of Ontology

Hybrid ontology description [7] approaches have been used to construct ontology of the system (figure 2). The semantics of each source is described by its own ontology (application ontology). But in order to make the local ontologies comparable to each other they are built from a global shared vocabulary (domain ontology). Shared Vocabulary is based on the assumption that members of the domain share a common understanding of certain concepts. Once a shared vocabulary exists, those terms can be used to make the application ontology for a source. The global ontology contains basic terms of a domain, which are combined in the local ontologies in order to describe more complex semantics. At the central part of our SOA architecture is the ontology-based Service Broker.

In order to register an information source, a metadata description in form of a Comprehensive Source Description (CSD) is needed. Each CSD consists of metadata that describe technical and administrative details of the data source as well as its structural and syntactic schema and annotations [19]. We have adopted XML-Based Language for the CSD description.

The XML-based language description is very similar to that of an XML Schema description. It is relatively compatible with existing Web standards since many of them are designed to facilitate machine understandable web representation. We need to be able to write down ontologies in a way that enables systems to process them. This is essential, as we want to automate parts of the information sharing process. The hierarchical structure of the ontology is shown in figure 3.
Communication Link
Rocks
Metalled Roads
Unmetalled Roads

(\texttt{define-concept roads\_class}
(\texttt{Roads\_Metalled\_Roads\_General\_Roads})
(\texttt{Roads\_Metalled\_Roads\_Traffic\_Roads})
(\texttt{Roads\_Metalled\_Roads\_National\_Highways})
(\texttt{Roads\_Unmetalled\_Roads\_Village\_Roads})
)

\begin{figure}
\begin{center}
\begin{verbatim}
<featureType datastore = "roads" >
  <name>road</name>
  <SRS>4326</SRS>
  <title>Road Data</title>
  <abstract>
  This is some road data states
  </abstract>
  <numDecimals value = "0" />
  <keywords>road, metalled, unmetalled</keywords>
  <latLonBoundingBox dynamic="false"
    miny="24.955967" maxy="49.371735"
    minx="-124.731" maxx="-66.969849" />
  <styles default="simple-roadtype"/>
</featureType>
\end{verbatim}
\end{center}
\end{figure}

Figure 3. Hierarchical description of ontology

4.2 GI Services
In order to provide the geospatial data to the user communities who have little or know knowledge of the data sets, they must be provided as service units. Services can communicate in an interoperable way with each other if they have well-defined standard interface. This means that they must have agreed upon some well-defined service standards and interface requirements. WFS have been used as wrappers at the data source end to exactly serve this purpose.

The use of standardized service interface is an important requirement for integrating the heterogeneous data repositories. It allows the services to be classified as well known service types (features) that provide the services through specified interface. Several testbed experiments has been carried out by OGC for several service types:

- Web Map Service (WMS) [10] for producing digital maps.

A service consumer having knowledge of these interface specifications can access the services over the Web after discovering services. The ontology-based annotation for the information sources describes them as features. Thus the data sources can provide feature service, which is defined through its application schema. The shared vocabulary of domain ontology is used to describe the detail of the properties included in the schema. The application schema can be requested through WFS DescribeFeatureType operation.

4.3 Publishing the Services
Publication of a service actually refers to registering a service to the ontology and providing necessary information for it being discovered. Generally a service is characterized through several properties, e.g. input, output, functionality etc. While publishing the web services in the ontology, this features of services needs to be considered. Registration of service is a three step procedure [17]

- Specifying some shared vocabulary relevant to the service. These shared vocabularies is a set of agreed upon vocabularies among the data providers’ community
- The application ontology for the service is defined in such a way that a requester can understand the general concept of the service. A reference of the application ontology has to be provided in the sources description (CSD) of the service.
- The source description for the service contains a reference to the corresponding ontology. In addition to that several other detail of the service like administrative and technical details.

Once a service is published it can be discovered by a requester using an ontology-based reasoner.

4.4 Service Discovery and Retrieval

4.4.1 Service Discovery
If an information request is submitted to an information broker (ontology), the broker has to decide which of the registered sources it should use to answer the request. The actual search for a service is performed by mapping the query concept and concepts in different application ontologies. This is achieved through the application of a terminological reasoner, e.g. RACER (Reasoner for A-Boxes and Concept Expressions Renamed) [20].

For each application schema there is one application ontology that is described with the shared vocabulary of the corresponding domain. These ontologies provide the formal description of the application schema of a data source. Once the ontologies of the system have been provided, the services can be discovered through GetCapabilities request. A sample output for the service discovery request is shown in figure 4.

4.4.2 Service Retrieval
Users pose queries for retrieving data from services in terms of features of the discovered services. As a consequence, the system must contain a module that uses the resource descriptions in the ontology in order to translate a user query into sub queries that refers directly to the schemas of the data resources. For this purpose, we establish a correspondence between each resource and the central ontology. The generic format for users’ query for the system is based on CQL, an XQuery like query language. A user generally request one or more collections of features of data and may specify some filtering operation on the data. As an example, a user might request for road networks of some place ‘kgp’. The road could be of various types – depending on the shared vocabulary in the ontology, the road types are identified. Each of these road types is included in the query for accessing road network data. The equivalent query in CQL is as shown in
During query processing, the CSD descriptions for the related feature service are parsed.

```xml
<FeatureType>
  <Name>kpg:road</Name>
  <Title>road_Type</Title>
  <Abstract>Generated from shape file</Abstract>
  <Keywords>road shape file</Keywords>
  <SRS>EPSG:4326</SRS>
  <LatLongBoundingBox minx="-0.0014" miny="-0.0024" maxx="0.0042" maxy="0.0018"/>
</FeatureType>

<FeatureType>
  <Name>kpg:cultivation</Name>
  <Title>cultivation_Type</Title>
  <Abstract>Generated from GML file</Abstract>
  <Keywords>Cultivation GML file</Keywords>
  <SRS>EPSG:4326</SRS>
  <LatLongBoundingBox minx="0.0014" miny="-0.0011" maxx="0.0042" maxy="0.0024"/>
</FeatureType>
```

The **Query Processor** (figure 6) translates the query in sub-queries and dispatches them to the appropriate WFS module at the data providers end. A WFS request consists of a description of a query or data transformation operations that are to be applied to one or more features. The WFS is invoked to service the request. The WFS layers at each data source end are in charge of querying the data sources: it receives requests from the **Query Processor**, executes them, converts the data in standard GML [9] form before transferring them to the **Query Processor**. The feature data, that is sent to the **Query Processor** is processed on the basis of users specified filter operation.

```xml
<wfs:Query typeName="kgp road_network">
  <wfs:PropertyName>ROAD_NAME</wfs:PropertyName>
  <wfs:PropertyName>ROAD_TYPE</wfs:PropertyName>
  <wfs:PropertyName>the_geom</wfs:PropertyName>
  <Filter>
    <PropertyTypeLike>
      Metalled Road
    </PropertyTypeLike>
    <PropertyTypeLike>
      Unmetalled Road
    </PropertyTypeLike>
  </Filter>
</wfs:Query>
```

**Figure 4. Service discovery form the ontology-based system**

On the syntactic level, wrappers are used to establish a communication channel to the data source(s), which have been found, that is independent of specific file formats and system implementations. Each generic wrapper covers a specific file- or data format. For example, generic wrappers may exist for ODBC data sources, XML data files, or specific GIS format. Still, these generic wrappers have to be configured for the specific requirements of a data source. We have used WFS as wrappers at the information source. These wrappers maintained at each data provider’s end accepts query in the form of WFS GetFeature request, i.e., a XML&KVP (Keyword Value Pair) or as GQuery expression, a SQL like query language derived from the OGC CQL (Catalog Query Language). [21] gives a detailed description of the GQuery Language.

```xml
<wfs:Query typeName="states">
  <wfs:PropertyName>STATE_NAME</wfs:PropertyName>
  <wfs:PropertyName>LAND_KM</wfs:PropertyName>
  <wfs:PropertyName>the_geom</wfs:PropertyName>
  <Filter>
    <PropertyIsBetween>
      <PropertyName>LAND_KM</PropertyName>
      <LowerBoundary>
        <Literal>100000</Literal>
      </LowerBoundary>
      <UpperBoundary>
        <Literal>150000</Literal>
      </UpperBoundary>
    </PropertyIsBetween>
  </Filter>
</wfs:Query>
```

**Figure 5. A sample query for data retrieval**

**Figure 6. The Query Processor Module**

**Figure 7. A query with more enhanced data manipulation option**

A sample query from the service consumer, which performs a selection in the feature data to find the states with an area between 100,000 and 150,000 along with the attributes STATE_NAME, LAND_KM, and geometry, is shown in figure 7.

**5. REALIZATION OF SOA**

A Service broker (or catalog service) is an intermediary service whose responsibility is to bring a service requestor and a service provider together, and thus may be considered the core of any GI web service environment. After a service offered by a provider is identified to match the service requestor’s requirements it is
bound to the service requestor, the service then is executed, passing data and instructions across common interfaces. In the previous sections we have provided a detailed discussion of the implementation of the system and its different components. As described previously, SOA involves three different kinds of actors: service providers, service requesters and discovery agencies. The role of these actors in the developed system has been performed by Data sources, Data users and the Ontology respectively.

In the domain of spatial data integration using SOA we have viewed these three components (figure 8) as follows:

- **Service Providers** – They are the actual data providers wrapped up in a WFS/WMS interface. These interfaces accept request form the Query Processor and perform the operation on the proprietary repositories. The retrieved data in sent back in GML format.

- **Service Requesters** – General users who can discover a service by requesting the central ontology through WFS GetCapabilities request. This operation provides a description of the service, its operations, parameters and data types. It is used for the clients to identify if a service provides the needed functionality and how to access it. The different features that can be queried along with the operations that can be performed on them are listed in a metadata format.

- **Discovery Agency** – The Ontology plays the role of Discovery Agency/Service Broker. It holds the global schema for all the repositories and provides a standard interface for the Service Requesters. Service available in the integrated system can be discovered by querying the Ontology.

![Figure 8. The Actors in a SOA application of spatial data](image)

### 6. CONCLUSION

Geographic data is increasingly becoming available on the Internet, allowing a large number of users to share and access the rich databases that are currently being maintained in several organizations. However, GIS data is immensely heterogeneous, being available in various formats and stored in diverse media (flat files, relational database). In this paper we have discussed SOA architecture for the Enterprise-wide GIS (E-GIS) system. Current GI standards developed by OGC have been use as the basic building blocks of the service-based system. This paper describes the central role of a geographic ontology in the development of an integrated information system. An initiative has been taken towards enterprise-wide integration and sharing of spatial data. The advantages and benefits of a service method have been fully exploited for designing E-GIS architecture. The added flexibility of adding and removing data sources to the integrated system makes it appropriate solution for the users who needs to access data from repositories, which may not be defined priory. The discovery and retrieval methodology of SOA makes the data repositories less coupled to each other making the system maintenance more cost effective.

The querying capability to the data has also been achieved through XML based query language XQuery. Data stored as flat files or in relational DBMS has been integrated for interoperable access. Designing a suitable client interface would bring ease for users in querying/accessing data. This issue is being taken up as future work of the project.

### 7. REFERENCES


