IW-SOSE 2006

ICSE 2006, Shanghai

An Approach for Service Oriented Discovery and Retrieval of Spatial Data

Authors: Manoj Paul, S.K. Ghosh

Dr. S. K. Ghosh

School of Information Technology

Indian Institute of Technology, Kharagpur, India

skg@sit.iitkgp.ernet.in

Outline

- Spatial Data – Importance
- Spatial Data Infrastructure
- General WS vs. SWS
  - SOA
- UDDI and Ontology
- System Architecture: Service Discovery and Retrieval

Spatial Data

- Refers to information that identifies the geographic location of natural or constructed features and boundaries on the earth
- Also known as geospatial data or geographic information
- Key information for effective planning and decision-making in a variety of Decision Support Systems (DSS)

Spatial Data

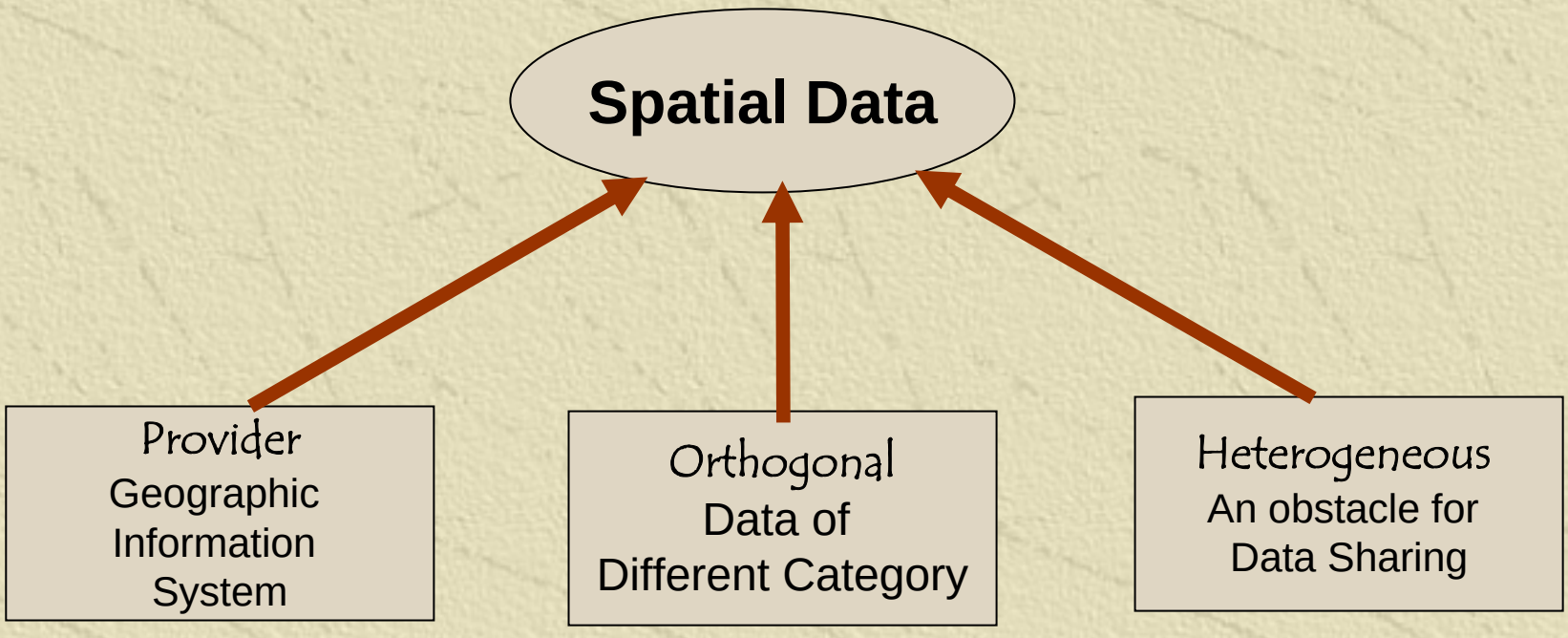
Types of Spatial Data:

- Geospatial data
  - Polygons
  - Points
  - Lines
  - Images/Grid
- Socioeconomic data
  - County/Province statistics
  - Census data
  - Social surveys

Spatial Data Sources:

- Geographic data (polygons, points and lines)
- Arc/Info data
- Shape files (*.shp, *.shx, and *.dbf)
- Grid
- Image data (ERDAS Image, JPEG, TIFF, BMP and Arc/Info Image)
- Tabular data (dBASE, INFO and TEXT)
- SQL
- SDE (Spatial Data Engine)

Spatial Data: Characteristics



Spatial Data: Integration

- There is demand for GIS in large computational environments, with a large number of users, spread over numerous locations
- Sharing large volumes of spatial data is an important task
- GIS are no longer simple departmental tools
  - Corporate tools
  - Governmental tools
  - Society-oriented applications

Spatial Data: Integration

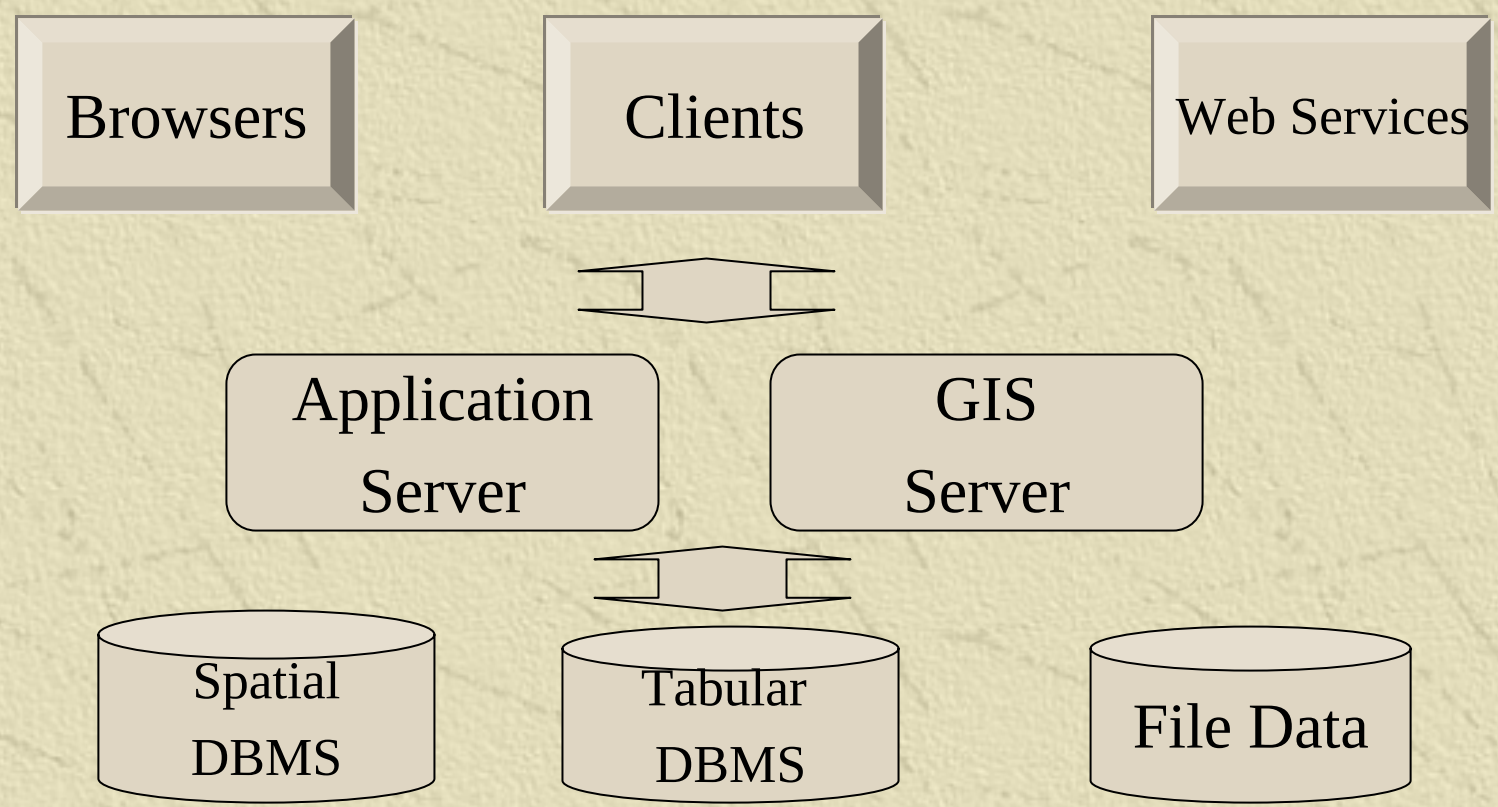
Factors influencing sharing:

- The fast development of networking technology and the Internet
- The success of Web-based GIS
- The wide applicability of mobile and ubiquitous computing
- Challenge
  - Heterogeneity (system, modeling and schematic) as well as need to support autonomy posed main challenges; major issues were data access and connectivity

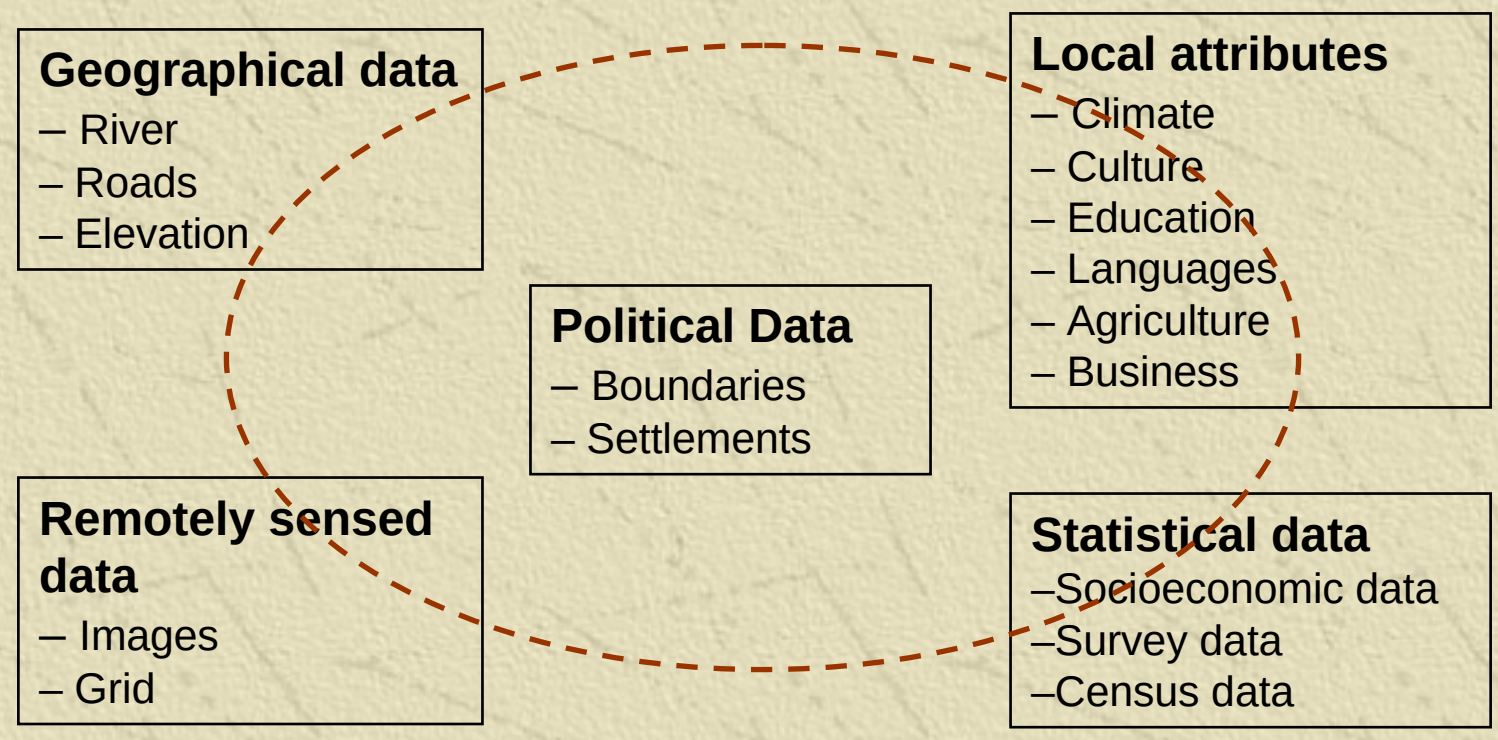
Enterprise-GIS (E-GIS)

- Growing demands for sharing of spatial data at the enterprise level
  - Enterprise GIS (E-GIS)
- An enterprise GIS is an integrated, multi-departmental system composed of interoperable components
- It provides broad access to geospatial data, a common infrastructure upon which to build and deploy GIS applications, and significant economies of scale

E-GIS


Browsers
Clients
Web Services
Application Server
GIS Server
Spatial DBMS
Tabular DBMS
File Data

E-GIS


Geographical data
– River
– Roads
– Elevation
Local attributes
– Climate
– Culture
– Education
– Languages
– Agriculture
– Business
Political Data
– Boundaries
– Settlements
Remotely sensed data
– Images
– Grid
Statistical data
–Socioeconomic data
–Survey data
–Census data

Integration: Issues

- GISs uses proprietary mechanisms for data storage and access
- Interoperability problems arise while integrating these diverse spatial data sources
- 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

Integration: Issues

- Geographic information domain is currently characterized by a paradigm shift
  - from monolithic systems to open and distributed GIS and their use processes.
- A move from standardized data formats to specifications of geographic information (GI) service interfaces
- The number of GI services available on the web is rapidly and continually increasing

Integration: Issues

- Till date geospatial data sharing and exchange among several organizations is hard to achieve

- It involves *atleast*
  - A *lot* of political negotiation
  - Agreements on standards
  - Agreements on costs and cost sharing
  - Agreements on maintenance
  - ….

Spatial Data Infrastructures

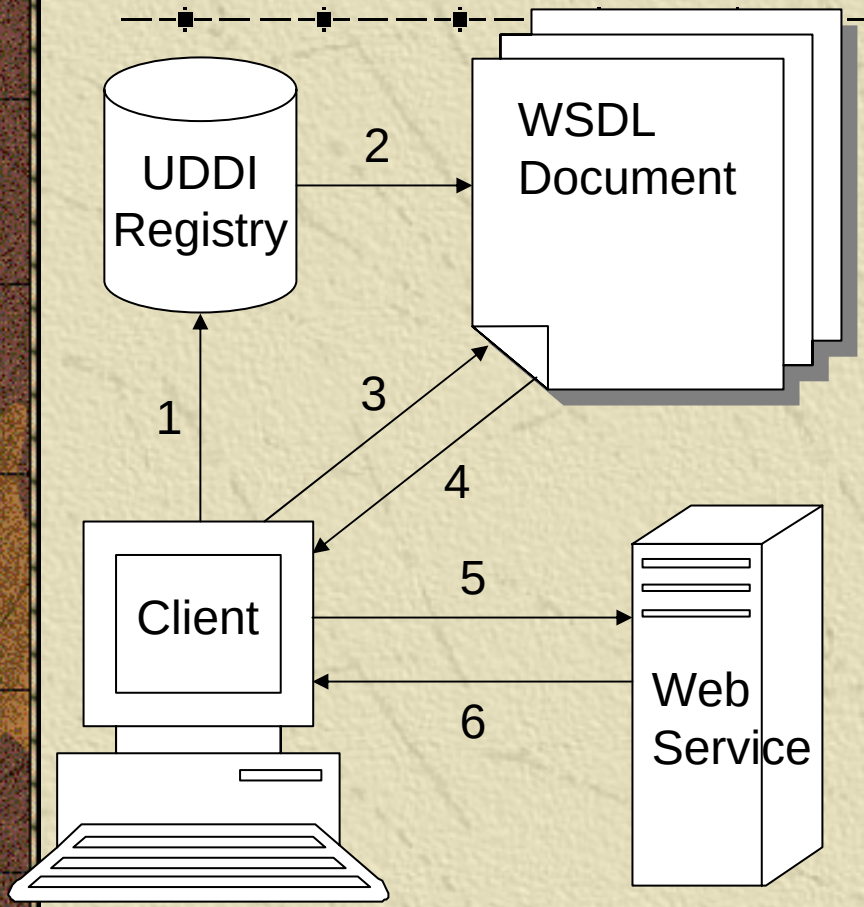
- An approach for integration of Spatial data sources at the national level
- *Infrastructure* implies that there should be some sort of coordination for policy formulation and implementation
- Definition (NSDI):
  - "Technologies, policies, and people necessary to promote sharing of geospatial data through all levels of government, the private and non-profit sectors, and the academic community"

Spatial Data Infrastructures

SDIs should

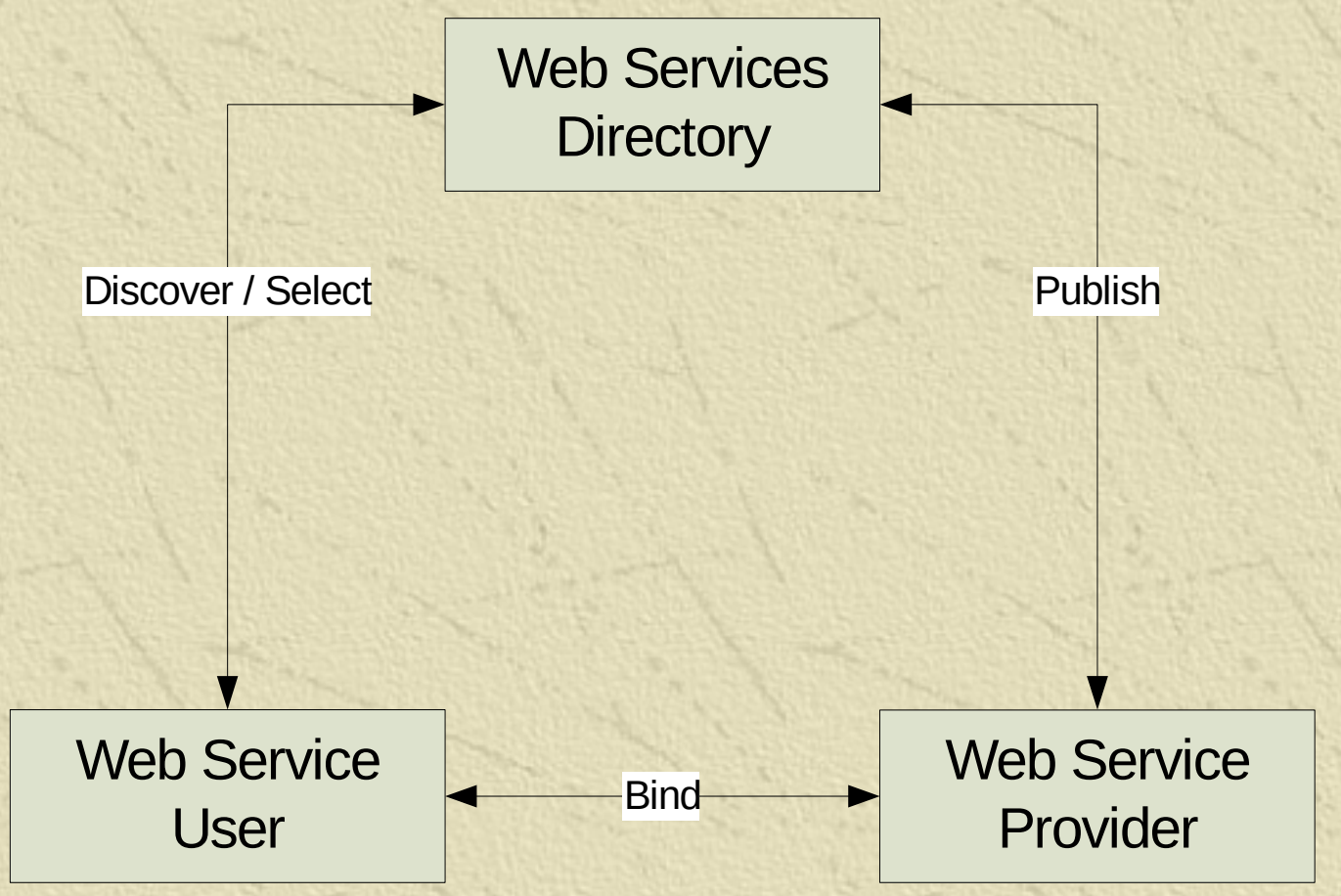
- Provide standardized access to data
- Have multiple participants, in the role of information services providers and/or users
- Have a broad thematic scope
- Facilitate data sharing

General Web Services



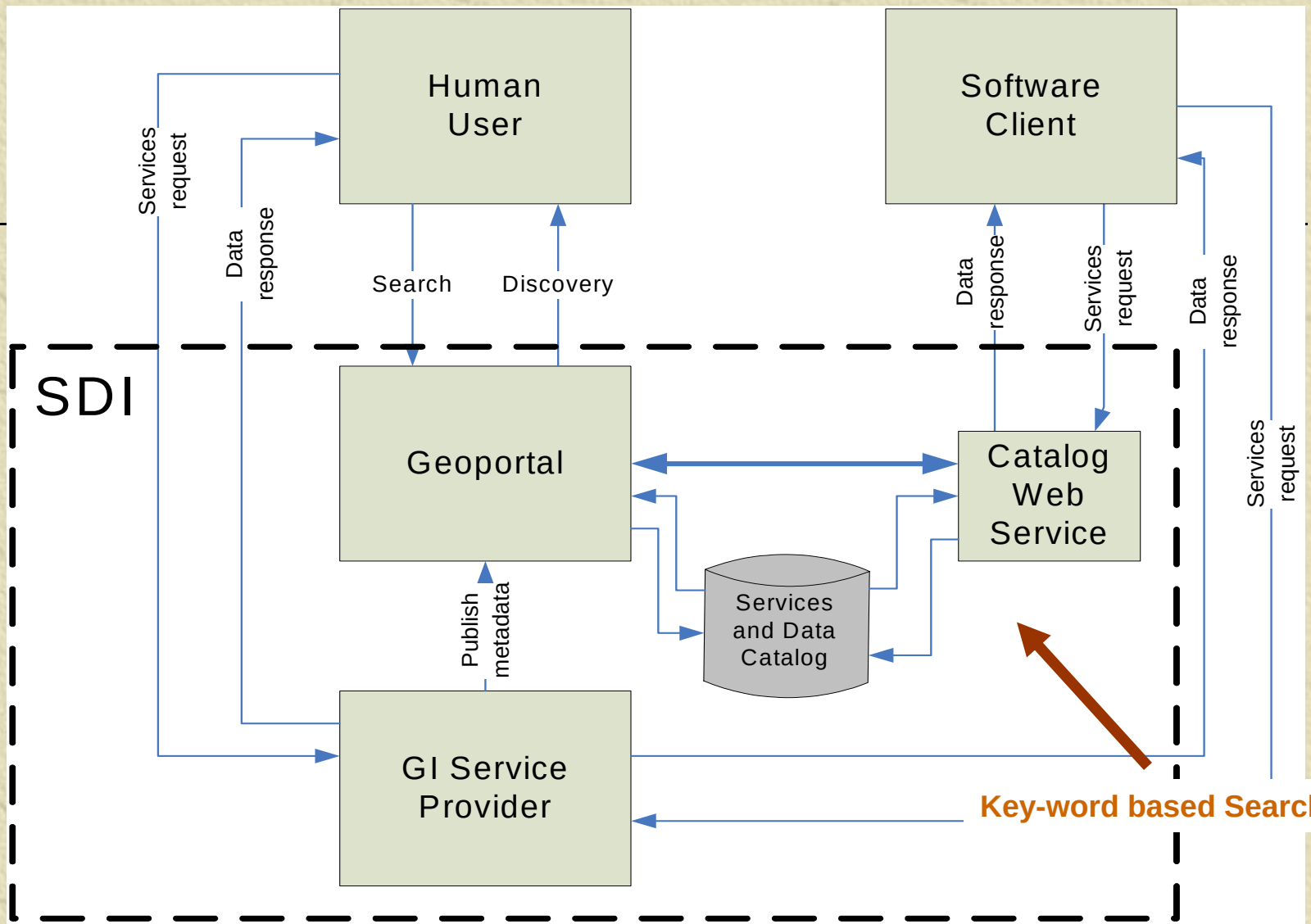
1. Client queries registry to locate service.
2. Registry refers client to WSDL document.
3. Client accesses WSDL document.
4. WSDL provides data to interact with Web service.
5. Client sends SOAP-message request.
6. Web service returns SOAP-message response.

SOA


Web Services Directory
Discover / Select
Publish
Web Service User
Bind
Web Service Provider

Geospatial Services

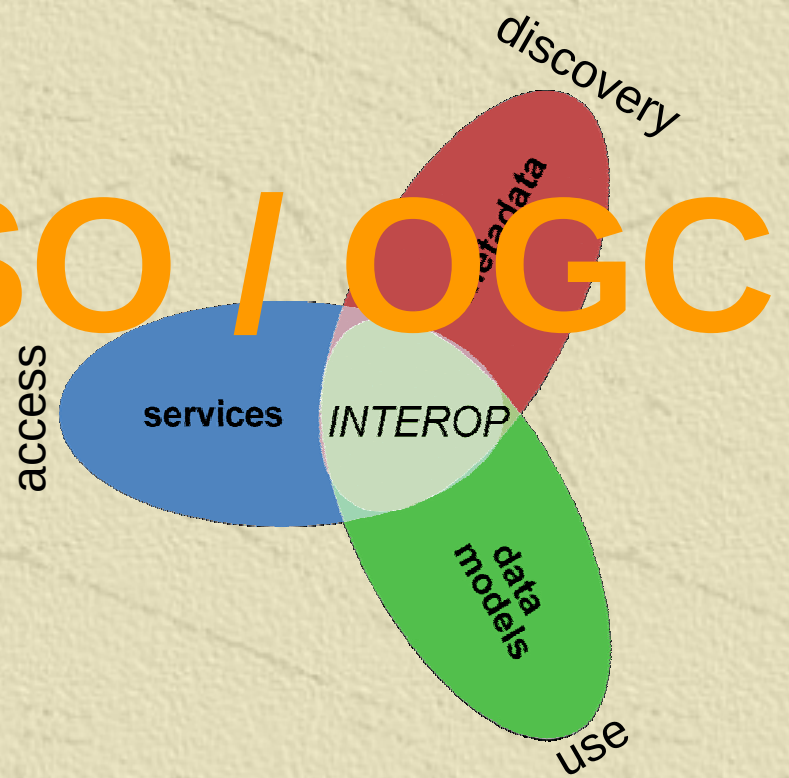
- Geospatial data can be interfaced in the web as Services for better accessibility among users
  - Geospatial data can be accessed in real-time over Internet as if it were local data
  - Timeliness of the data is preserved
- Multiple organizations can benefit from the data being staged and maintained once and used many times
- SDI can be built over the these Spatial Web Services (SWS)



Courtesy: PUC Minas

Standards

- **Dimensions of interoperability for earth science data**



Spatial Web Services - OGC

- Open Geospatial Consortium (OGC):
  - International consortium of nearly 300 companies, government agencies and universities participating in a consensus process to develop publicly available geo-processing specifications
- OGC web services
  - Web Map Service
  - Web Feature Service
  - Web Coverage Service

OGC: WMS

- Web Map Service (WMS)
- Three operations:
  - GetCapabilities
  - GetMap
  - GetFeatureInfo
- GetCapabilities request
  - <Service> metadata for service as a whole (ISO 19115 compliant)
  - <Capability> metadata describes request bindings, exceptions, and:
  - <Layer>s: title, name, SRS, bounding box, keywords, style; nested layers inherit from parent

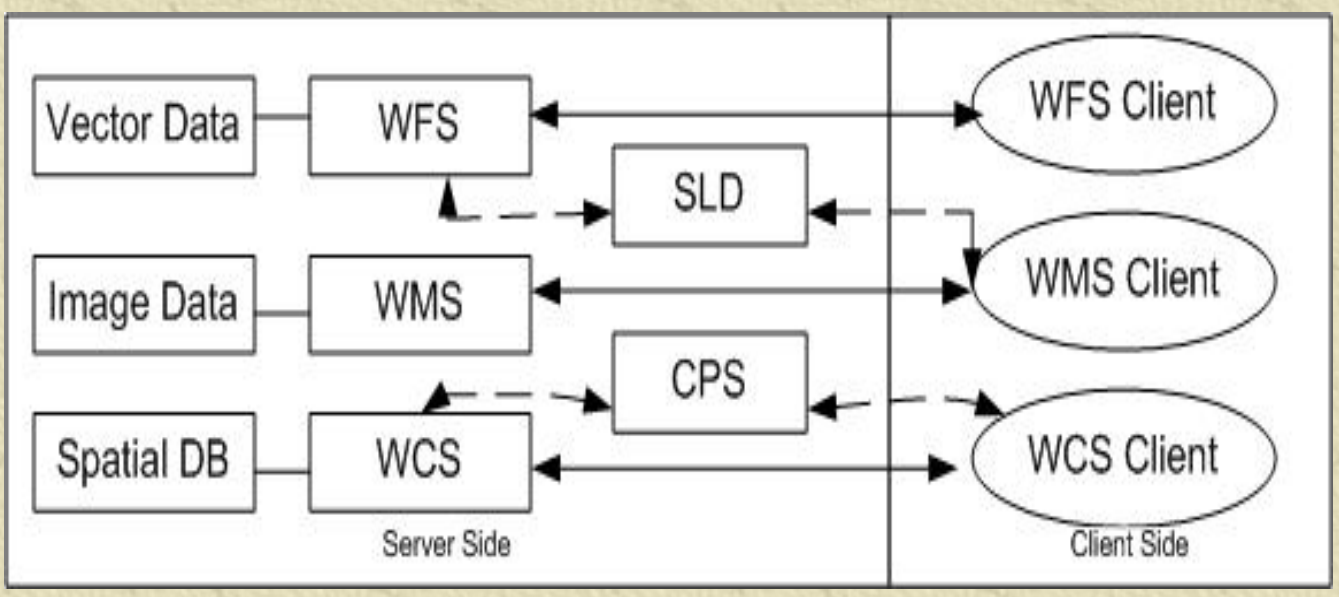
OGC: WFS

- *Web Feature Service (WFS)*
  - **Provides access to geographic *feature instances***
- **Features are defined by *application schema* compliant with the Geography Markup Language (GML)**
- **Operations:**
  - DescribeFeatureType: **Returns XML schema for the feature**
  - GetFeature: **Allows retrieval of features in XML compliant to the feature schema. Selected features may be constrained through Xpath expressions**
  - GetCapabilities: **describes capabilities of WFS (as for WMS)**

OGC Web Services

- **Web Map Service (WMS, ISO 19128)**
  - GetCapabilities
  - GetMap
- **Web Feature Service (WFS)**
  - GetCapabilities
  - DescribeFeatureType
  - GetFeature
  - Transaction, LockFeature
- **Web Coverage Service (WCS)**
  - GetCapabilities
  - DescribeCoverage
  - GetCoverage
- **Catalog Service for the Web (CSW) – new!**

OGC Web Services



SWS *vs* WS

- SWS do not necessarily use the usual W3C standards, including SOAP and WSDL
- Instead of UDDI, OWS propose the use of catalog services
- SWS have a particular interface for binding
- SWS use GML, and not plain XML

⟹ Focus on Service Discovery

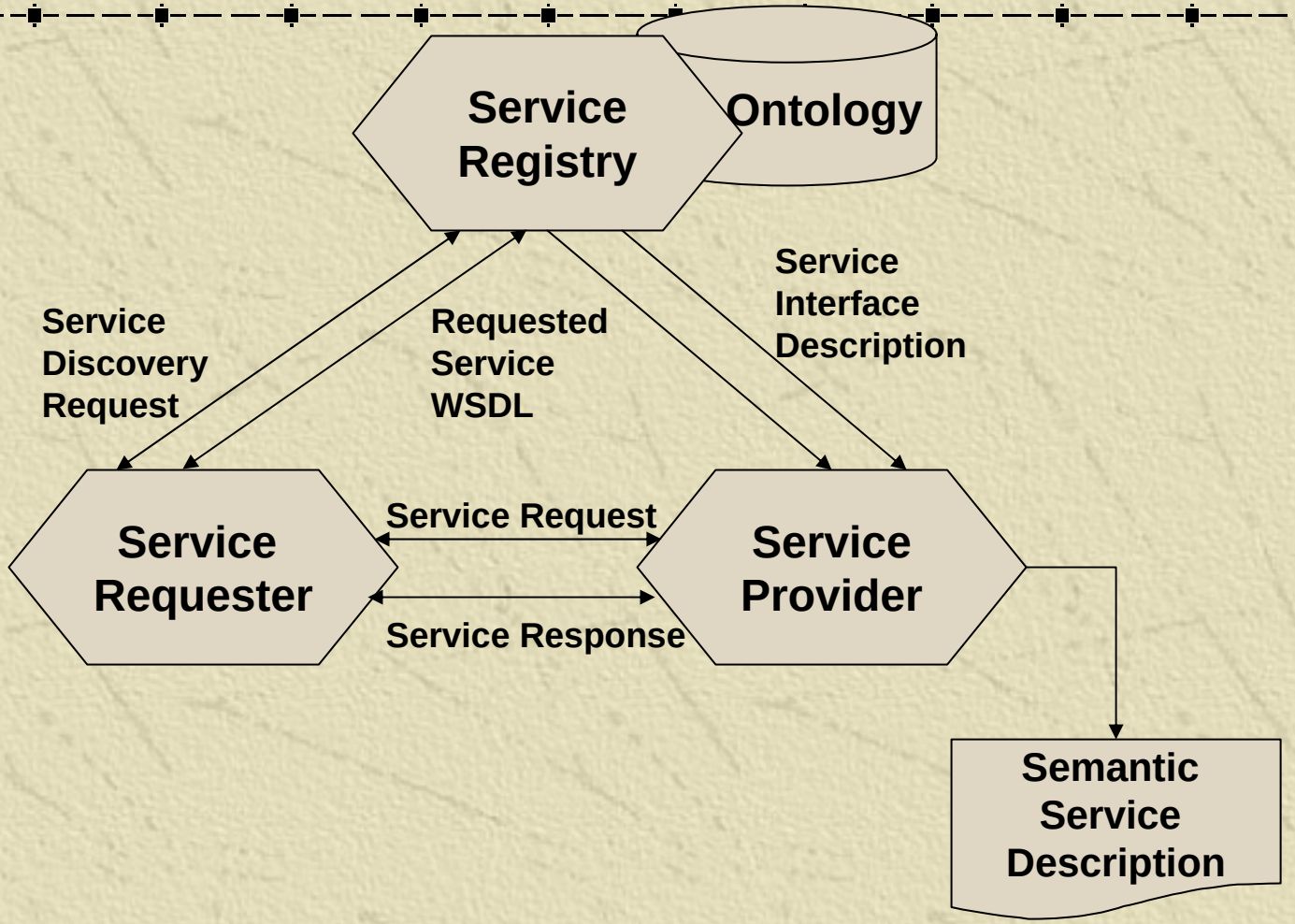
UDDI

- Registry for services where a service requester can browse through available services
- Service advertisement in the form of XML-Schema
  - Inputs
  - Outputs
  - Operation

UDDI Issues in GI Services

- UDDI introduces keyword-based retrieval mechanism. It does not allow advanced metadata-oriented query capabilities on the registry.
- UDDI does not take into account the volatile behavior of services and supports only quasi-static service registries.
- Since UDDI is domain-independent, it does not provide domain-specific query capabilities in particular for GIS domain such as spatial queries.

Enhanced SOA



OGC CS: Semantic Issues

- Presently, capabilities of GI web services are made available in OGC catalogues using standardized metadata templates
- Employs Text-based search method for the discovery of service
  - Problems can occur if requesters and providers of web services use different terms for referring to the same or similar concepts or if they use the same term for different concepts
  - Semantic heterogeneity problem
- Keyword-based service discovery approaches cannot completely capture the semantics of the user's query

Semantic Issues: GI Ontology

- **An envisioned approach to overcome these limitations is to use ontology-based service discovery**
- **Ontologies are used for classification of the services based on their properties**
- **This enables retrieval based on service types rather than keywords**

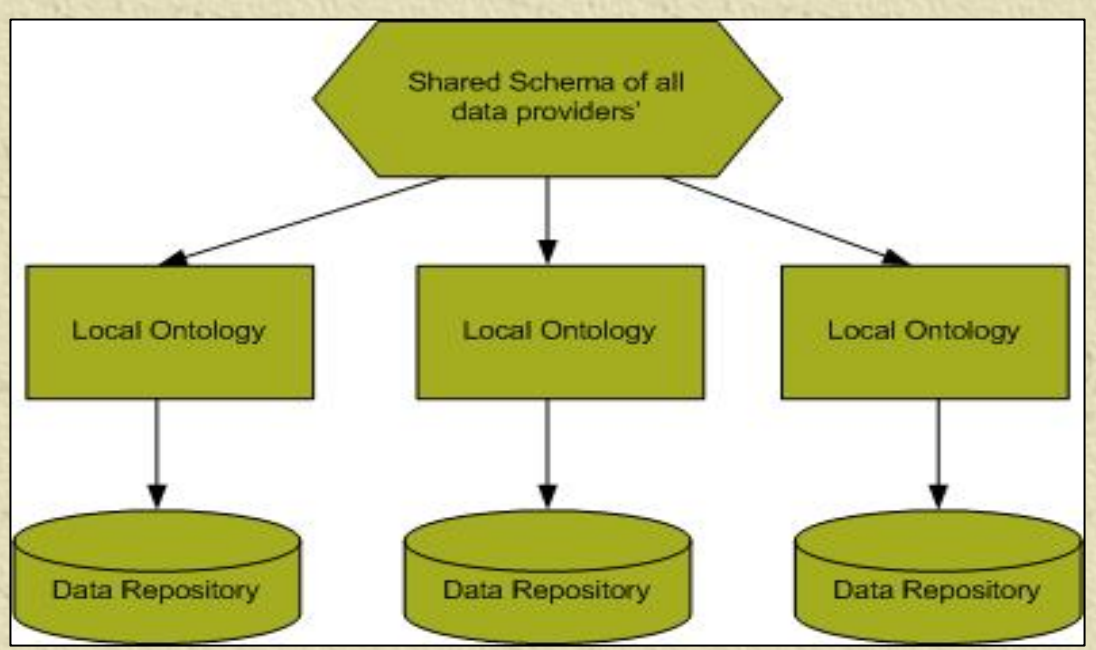
GI Ontology: Structure

Hybrid ontology description

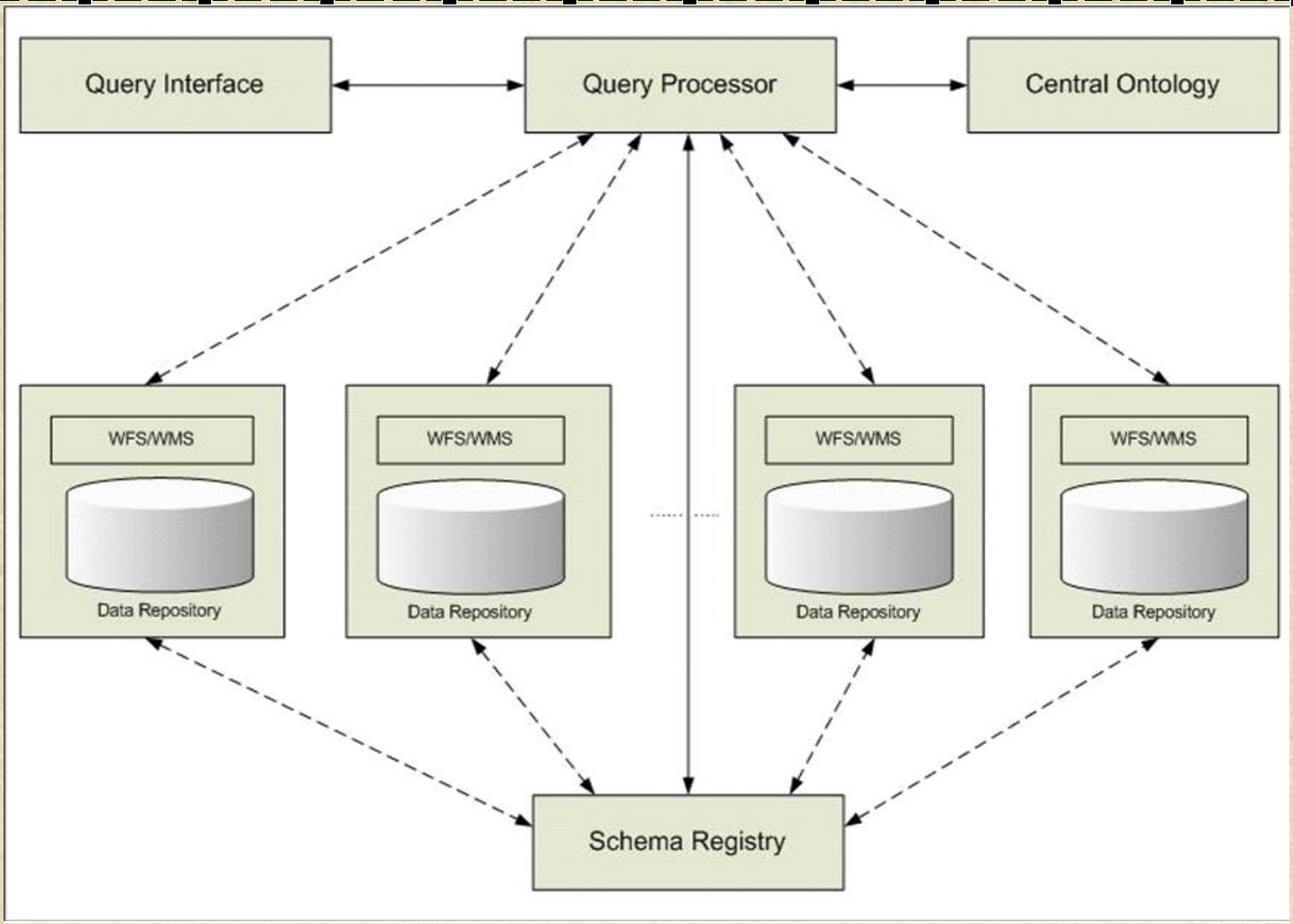
- Application Ontology: The semantics of each source is described by its own ontology
- Domain Ontology: Global shared vocabulary for the domain of concepts

Ontology: Structure

Hybrid Ontology Structure



System Architecture



GI Service Discovery

- Service requests are matched with the service advertisements in Service Registry
- Instead of Direct Key-word search, semantic search is employed
- The matching process goes through two steps
  - Domain Dependent Matching
  - Service Specific Matching

Domain Dependent Matching

- Uses Domain Ontology for Matchmaking
- Categorizes service request based on the request parameters
- Doesn’t involve any particular service advertisements

Domain Dependent Matching: Example

- User Requested Service:

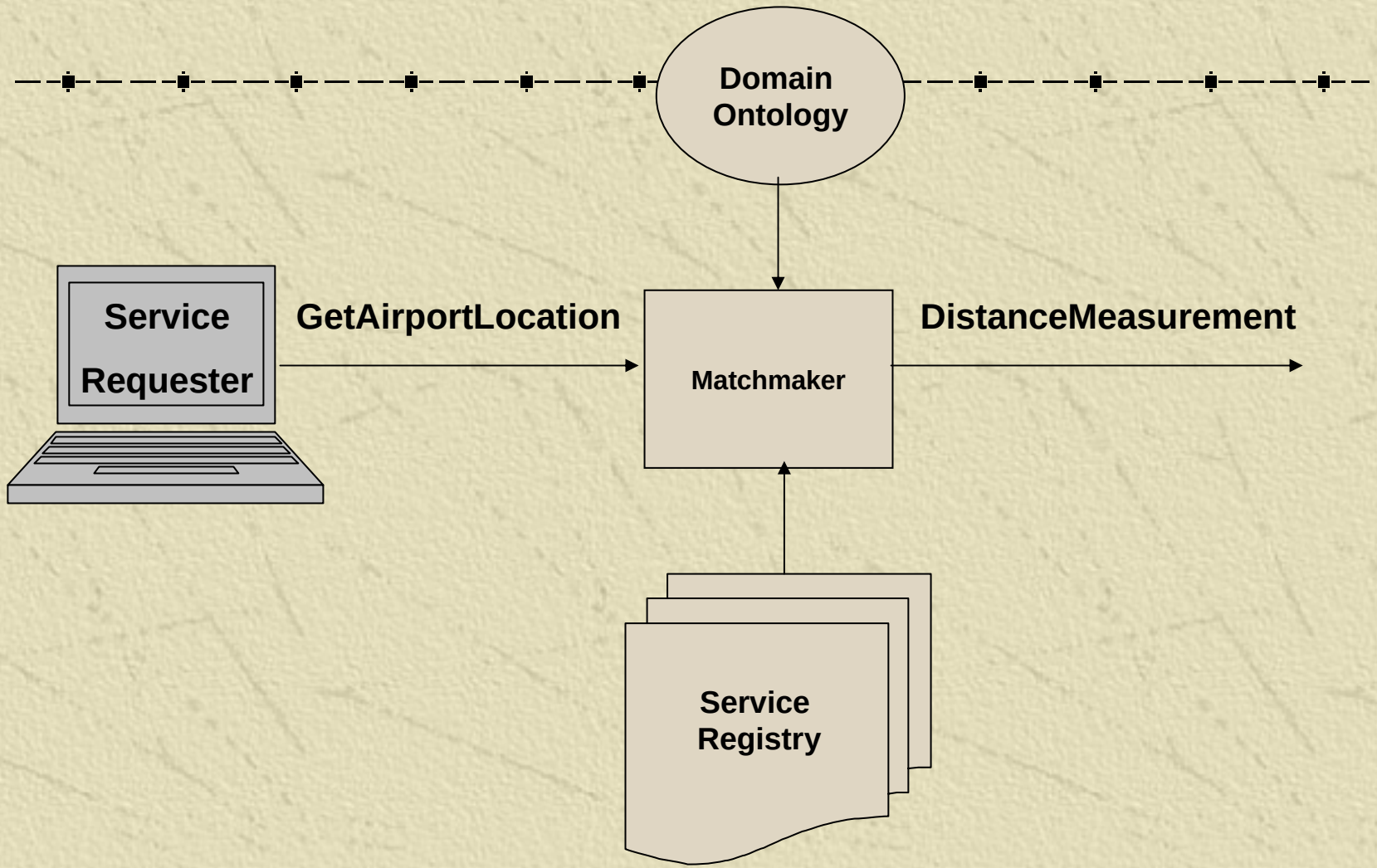
  GetAirportLocation

- Catalog based searching employs searching all the service advertisement
  - Service Discovery gets slowed due to large number of available services
  - Low Recall/Precision

- Ontology based searching categorizes it to be of category

  DistanceMeasurement

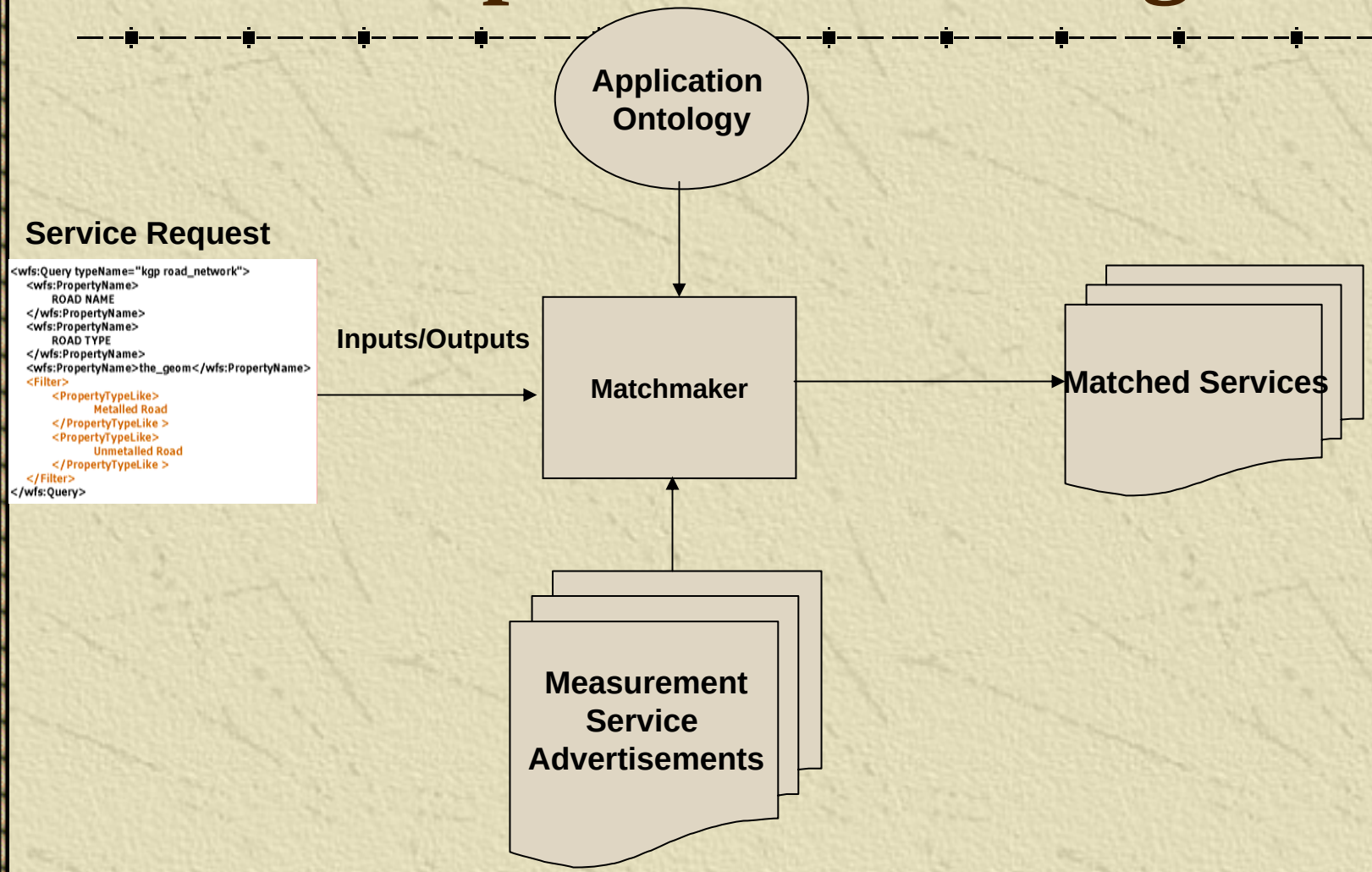
Domain Dependent Matching: Example



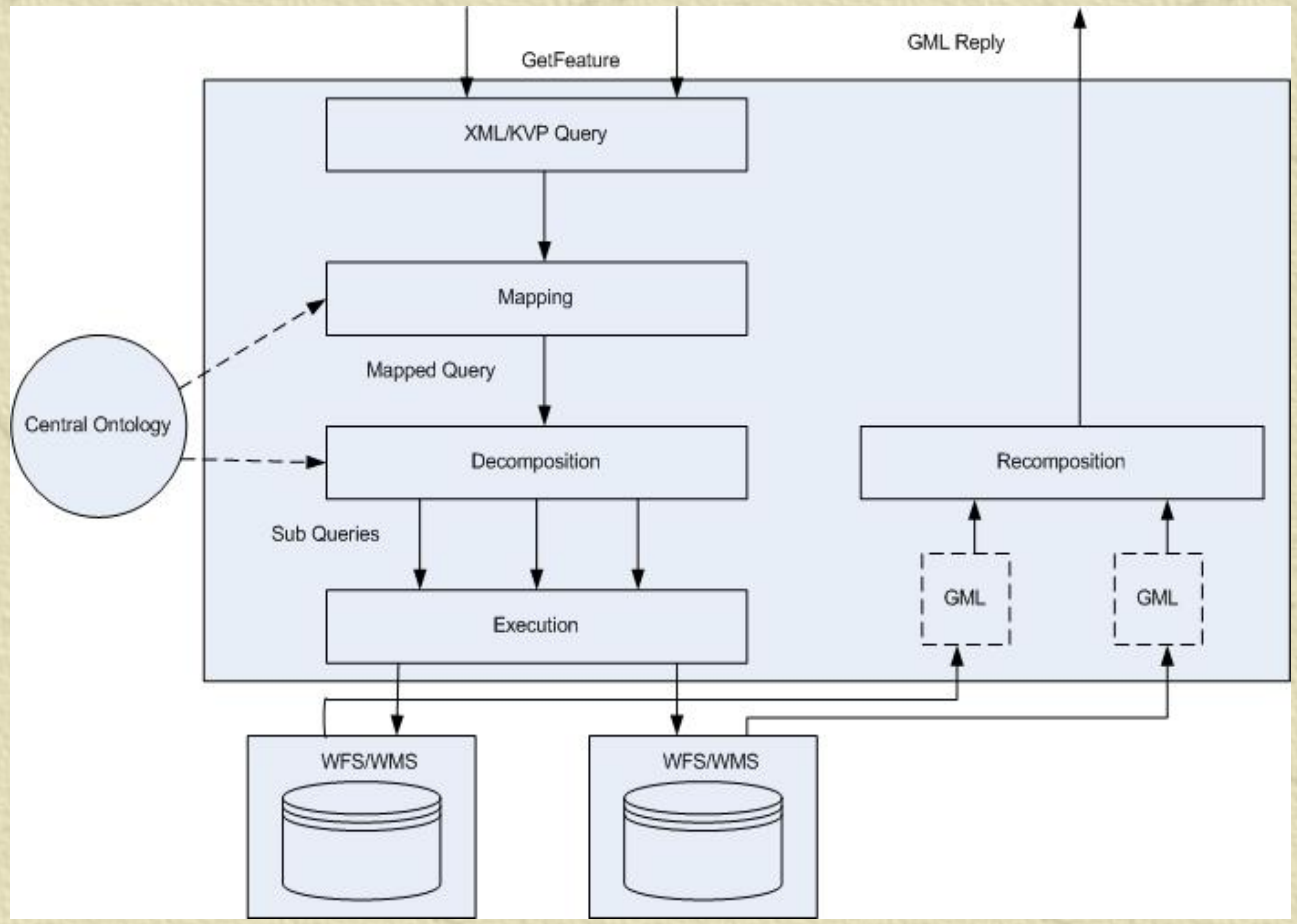
GI Service Specific Matching

- Uses service specific Application ontology for further mapping
- Service advertisements are matched for
  - Inputs: *AirportName, CityLocation*
  - Outputs: *Distance*
- The matching utilizes Application Ontology for semantic matching
  - Increased Recall/Precision

Service Specific Matching



Query Processor



Conclusion

- An Integration Platform for Spatial data through
  - Providing spatial information as services
  - Service Oriented Architecture
  - Adhering to OGC standards
- Heterogeneity issues in spatial information has been addressed through
  - Ontological description for the domain
  - Semantic based service discovery
- Standard Query processing methodology

Thank You !!!

skg@sit.iitkgp.ernet.in
skg@iitkgp.ac.in