Theme 4. SPATIAL DATA CATALOGUES

4.1. General provisions on Spatial Data Catalogues

1. Spatial data that are stored for use in local databases can often be used in external applications once they are published.

2. Spatial Data Catalogues are presented as a means to publish descriptions of spatial data holdings in a standard way to permit search across multiple servers.

3. **Spatial Data Catalogues** are discovery and access systems that use metadata as the target for query on raster, vector, and tabular geospatial information.

4. The principles described in this theme can be interpreted and applied in a range of information management conditions from non-digital collections of map information, through small digital catalogues, to integrated repositories of data and metadata.

6. Support of a discovery and access service for spatial information is known variously within the geospatial community as:

1) 'Catalogue services' (OpenGIS Consortium);

2) 'Spatial Data Directory' (Australian Spatial Data Infrastructure);

3) 'Clearinghouse' and the 'Geospatial One-Stop Portal' (U.S. FGDC).

7. Although they have different names, the goals of discovering spatial data through the metadata properties they report are the same.

8. For the purpose of consistency within this theme, these services will be referred to as '**Catalogue Services**'.

9. Further integration of these services with web mapping,

Notes.

1. **Clearinghouse:**

1) A distributed network of spatial data producers, managers, and users linked electronically;

2) Incorporates the data discovery and distribution components of a spatial data infrastructure.

2. **Gateway** – one of the Web devices/services.

4.2. Distributed Spatial Data Catalogue concept

1. The Catalogue Gateway and its user interface allow a user to query distributed collections of spatial information through their metadata descriptions.

2. This information may take the form of "data" or of services available to interact with spatial data, described with complementary forms of metadata.

3. *Figure 4.1* shows the basic interactions of various individuals or organizations involved in the advertising and discovery of spatial data.

4. A user interested in locating spatial information uses a search user interface, fills out a search form, specifying

5. The search request is passed to the Catalogue Gateway and poses the query of one or more registered catalogue servers.

6. Each catalogue server manages a collection of metadata entries.

7. Within the metadata entries there are instructions on how to access the spatial data being described.

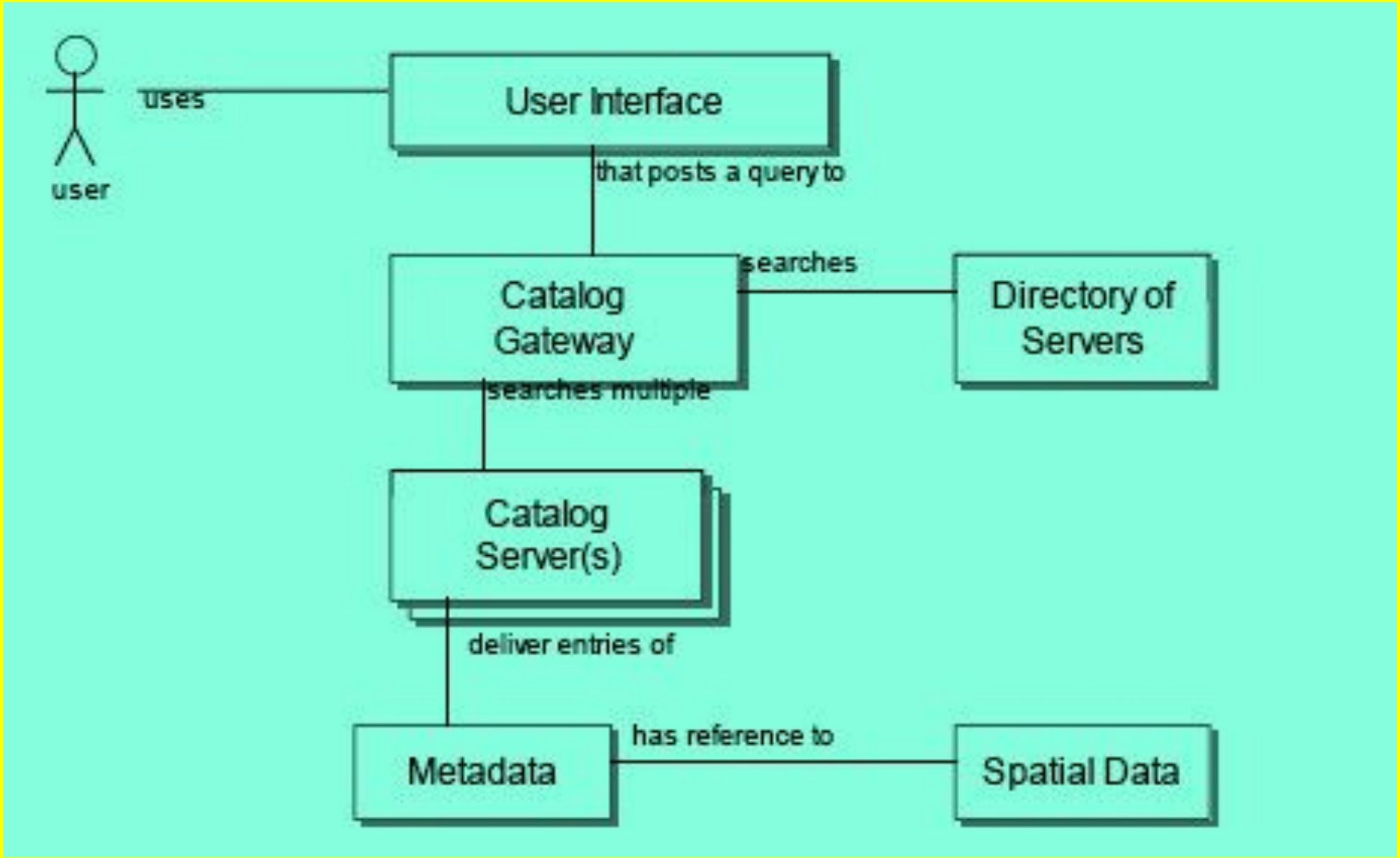
8. There are a variety of user interfaces available in this type of Catalogue search in various national and regional SDIs around the world.

9. Interoperable search across international Catalogues can be achieved through use of:

1) A common descriptive vocabulary (metadata);

2) A common search and retrieval protocol;

Fig.4.1 – Interaction diagram showing basic usage of Distributed Catalog Services and related SDI elements from a user point of view



10. The Distributed Catalogue environment is more than just a catalogue of locator records.

11. The Distributed Catalogue includes reference and/or access to data, ordering mechanisms, map graphics for data browsing, and other detailed use information that are provided through the Metadata Entries.

12. This metadata acts in three roles:

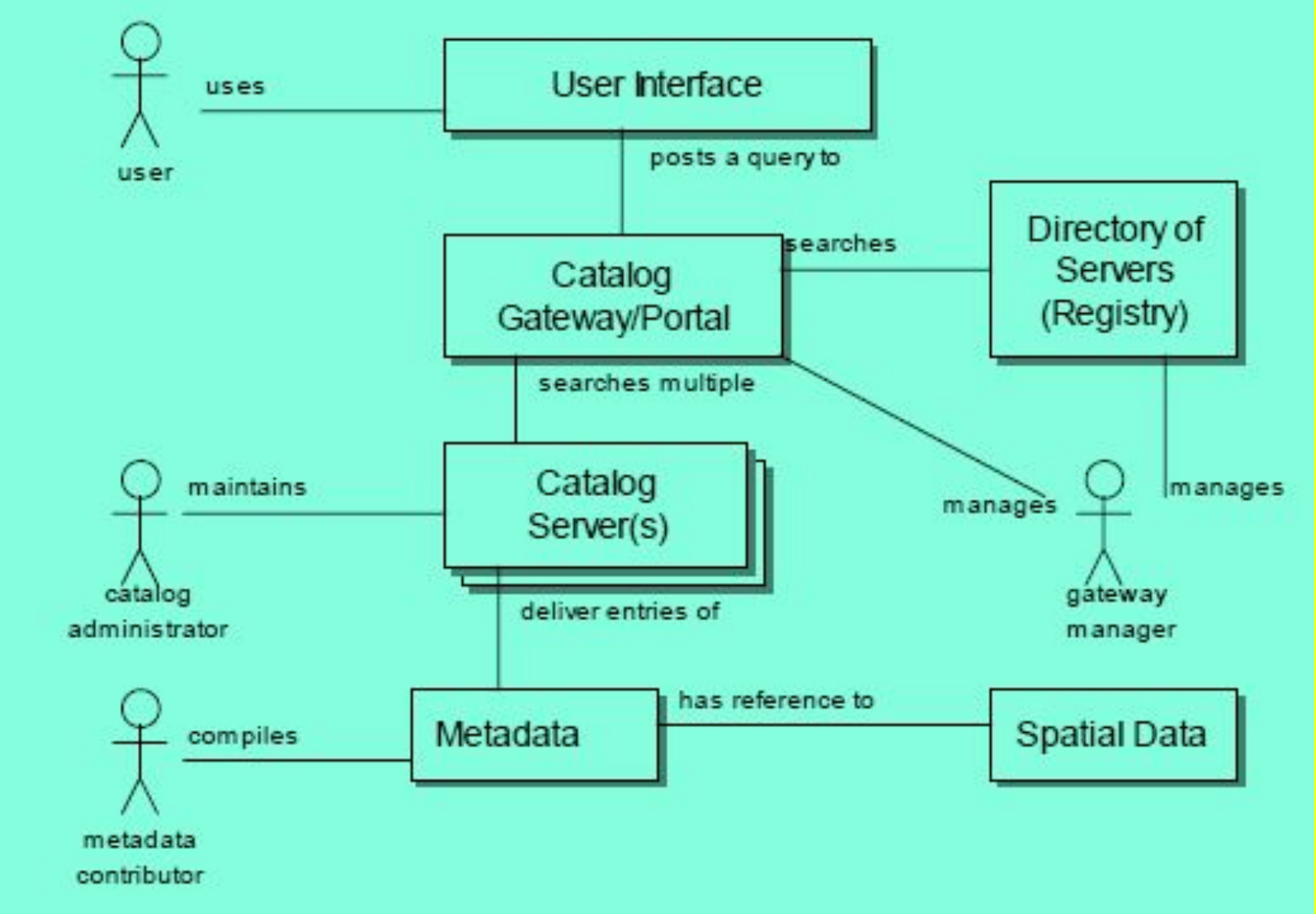
1) Documenting the location of the information;

2) Documenting the content and structures of the information;

3) Providing the end-user with detailed information on its

4.3. Organizational approach to Distributed Spatial Data Catalogue (Fig.4.2)

Fig.4.2 – Interaction diagram showing basic usage of Catalog Services and related



4.3.1. Terminology of Distributed Spatial Data Catalogue architecture

1. **Data Set** – a specific packaging of spatial information provided by a data producer or software, also known as a feature collection, image, or coverage.

2. **Metadata** – a formalized set of descriptive properties that is shared by a community to include guidance on expected structures, definitions, repeatability, and conditionality of elements.

3. **Metadata Entry** – a set of metadata that pertains specifically to a Data Set.

4. **Catalogue** – a single collection of Metadata Entries that is managed together.

6. **Catalogue entry** – a single Metadata Entry made accessible through a Catalogue Service or stored in a Catalogue.

7. **Service entry** – the metadata for an evocable service or operation, also known as operation or service metadata.

8. **Portal (Web Portal) –** a Web resource that provides access to a broad array of related resources and services (*Fig.4.3*). It uses *portlets* to allow many different programs to operate within the same Web page.

9. **Portlet**:

1) A standard Web portal component that processes requests and generates dynamic content;

2) Portlets are used in portals as pluggable user interfaces to add specialized content, such as weather information,

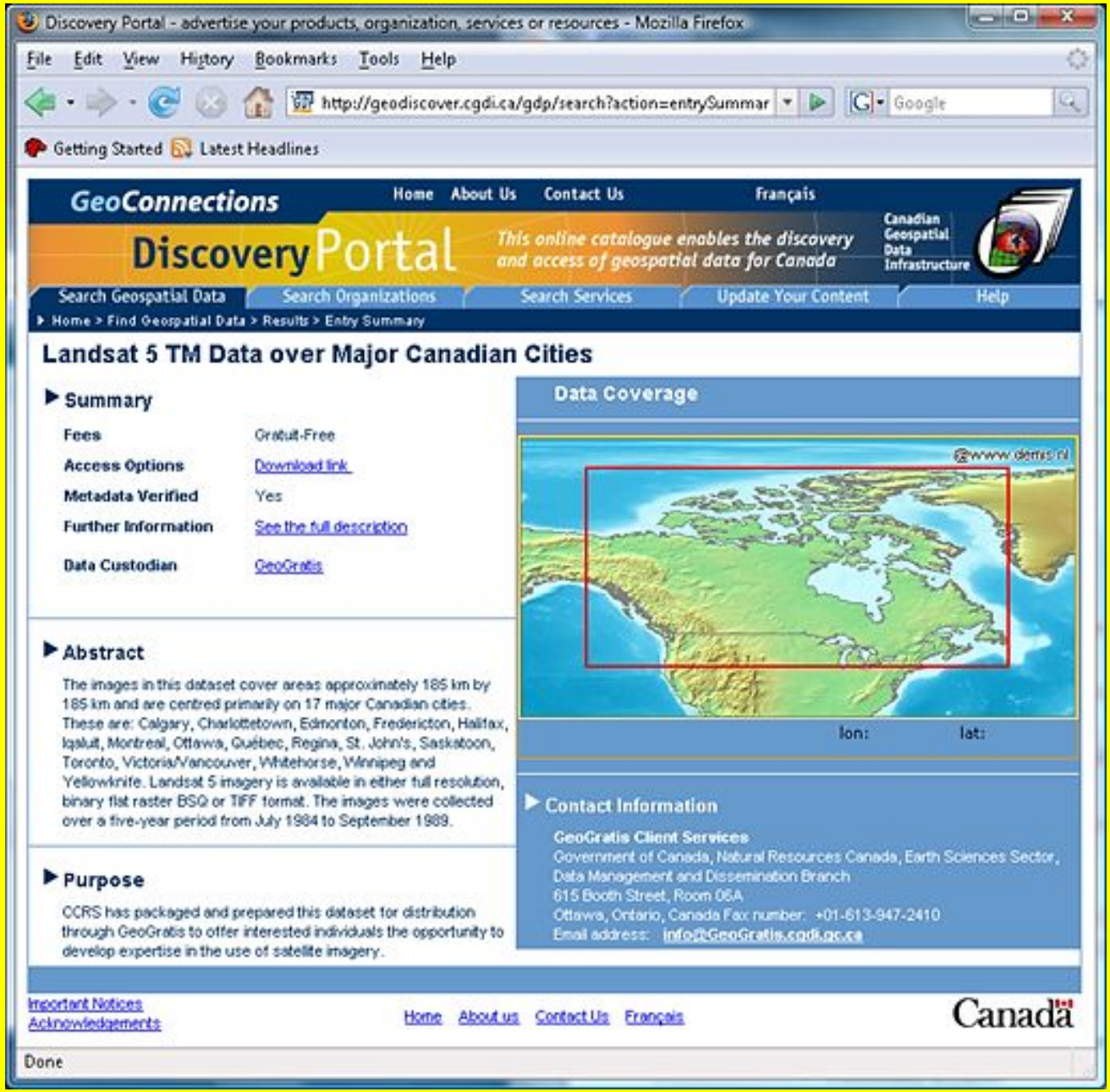


Fig.4.3 – Canadian Geospatial Data Infrastructure: GeoConnections Discovery Portal (*http://geodiscover.cgdi.ca*)

4.3.2. Actors and their functions in Distributed Spatial Data Catalogue architecture

1. **Originator of the Metadata Entry**:

– Has to generate conformant metadata elements packaged so they accurately reflect the contents of the information being described.

2. **Contributor to the Catalogue**:

– Has to provide one or more conformant Metadata Entries to a Catalogue;

3. **Catalogue Administrator**:

4. **Catalogue User**:

– Has to define criteria by which geographically related information could be located and used through:

a) Use of Browse categories;

b) Posing a fielded or full-text query.

5. **Gateway Manager**:

– Has to develop, host, and maintain the distributed search capabilities within the user community;

– Has to manage a contribution to a directory of servers (registry) that participate in the national or regional SDI.

4.3.3. Catalogue Server/Service organizational development

1. The construction of a Catalogue Service capability for spatial information is built upon on the commitment to collect and manage some level of spatial metadata within an organization.

2. The following ***use case scenario for the publishing of a Metadata Entry***:

1) A **Contributor of Metadata** receives the description of a new spatial data set developed by other professional staff;

2) This metadata is generated in a transferable encoding format to allow exchange of the metadata without loss of context or information content;

4) The **Catalogue Administrator** applies any acceptance criteria on the quality of the metadata as required by the organization;

5) If the metadata are acceptable it is inserted into the catalogue;

6) The **Catalogue Administrator** then updates the catalogue to reflect the new entry as available for public access;

7) This Data Set is now considered advertised because its metadata provide a searchable and browse-able record of:

a) Its background;

b) Its temporal and spatial extent;

3. There are three ***principal models for Catalogue Server/Service installation within or among organizations***:

1) **Consortium Model**:

– Is one where a single metadata catalogue:

a) Is built and operated at one location;

b) Is shared by multiple organizations with a common discipline or geographic context;

2) **Corporate Model**:

– Assumes that all metadata are forwarded within an organization to a single service at which time corporate issues of quality, publication, style, and content may be evaluated;

3) **Workgroup Model**:

– Assumes that a service would be established at each place

4.3.4. Catalogue Gateway and access interface organizational development

1. Problem can be divided into **two related parts** that must interrelate:

   1) A User Interface (*Search/Browse Interface, fig 4.2*);

   2) A query distributor (*Catalogue/Gateway Portal, fig 4.2*).

2. *Figure 4.4* shows **the possible configurations of a Catalogue Gateway and the User Interface**:

   1) Client A accesses a User Interface that is downloaded (as forms or an applet) from a host on the Internet that is also managing multiple connections to servers;

   2) Client B is accessing a User Interface from a location that is different from that of the Gateway supporting the construction of customized user interfaces for a community;

3. **Two styles of interaction** are known to exist in Web search interfaces that are equally well **applied to Distributed Catalogue access**:

1) The first style is **query** in which the user specifies search criteria for search using simple to advanced interfaces;

2) The second style is a **browse interface** in which the user is presented with categories of information and selects paths or groupings, often in hierarchical form, to traverse.

Notes.

1. The challenge of constructing and supporting browse mechanism across a global collection of servers is the work required in building and supporting a universal vocabulary for classification and its hierarchy or word space, known as ontology.

2. **Ontology** – a controlled, hierarchical vocabulary for describing a knowledge system.

4.3.5. Organizational registering of Catalogue Servers

1. The nature of Distributed Catalogues requires that the knowledge of the existence and properties of any given catalogue participating in a community be known to the community.

2. The **Directory of Servers' concept** allows an individual catalogue operator to construct and register service metadata with a central authority.

3. National listings of compatible catalogue servers have already been built.

4. The operation of a global network of Catalogue Servers within GSDI will require that a common Directory of Servers be built and managed to assure current content, distributed

5. The **features of the Directory of Servers** may include:

1) One descriptive entry per service collection (server metadata);

2) Ability for a donor to contribute or update a record in the directory;

3) Ability to validate access to a server, as advertised;

4) User browse access of online server metadata;

5) Software search access of server metadata;

6) Management of active/inactive records, accessibility statistics.

6. Several national Distributed Catalogue activities support management services for server-level metadata and contain

7. The GSDI now sponsors a global directory of catalogue servers for all countries to utilize:

1) With delegation of authority made to participating countries to manage and validate host information for their servers (*http://registry.gsdi.org/registry*);

2) But it does not provide for the cataloguing of all service types at this time.

8. The UDDI (*http://www.uddi.org*) offers the potential of a public, replicated "universal business registry" hosted by IBM, Microsoft, and SAP, that could be used by SDI publishers to advertise the existence of their services.

9. Research into the use of the UDDI as a service directory

Key standardization efforts in access to catalogues are found in the:

1) ISO 23950 Search and Retrieve Protocol:

2) The OpenGIS Consortium Catalogue Services Specification Version 1.0;

3) Relevant standards or "recommendations" of the World Wide Web Consortium (W3C).

4.4. Implementation approach to Distributed Spatial Data Catalogue

1. The development of operational Distributed Catalogue Services has been taking place in a number of countries including the United States, Canada, Mexico, Australia, and South Africa as primary examples.

2. The software systems used to implement the ISO 23950 and Web based services has been developed largely through governmental support, resulting in both open source and commercial software solutions.

3. The evolution of protocols and industry practices are difficult to predict, but this theme provides a review of

4. Let's review a ***technical use case scenario for access to a Distributed Catalogue***:

1) A User uses client software to discover that a Distributed Catalogue search service exists. This may be done through:

a) A search of Web resources;

b) A saved bookmark, reference from a referring page;

c) Word-of-mouth referral;

2) User opens the User Interface and assembles the parameters required to narrow down a search of available information;

3) The search request is passed to one or more servers based on user requirements through a Gateway service:

4) Results are returned from each server and are collated and presented to the User. Types of response styles may include:

a) A list of "hits" in title and link format;

b) A brief formatting of information;

c) A full presentation of metadata;

d) Display of Data Set locations on a map, thematic groupings, or temporal extent.

5) User selects:

a) The relevant Metadata Entry by name or reference;

b) The presentation content (brief, full, other) and the format (HTML, XML, Text, other) for further review;

6) User decides whether to acquire the Data Set through

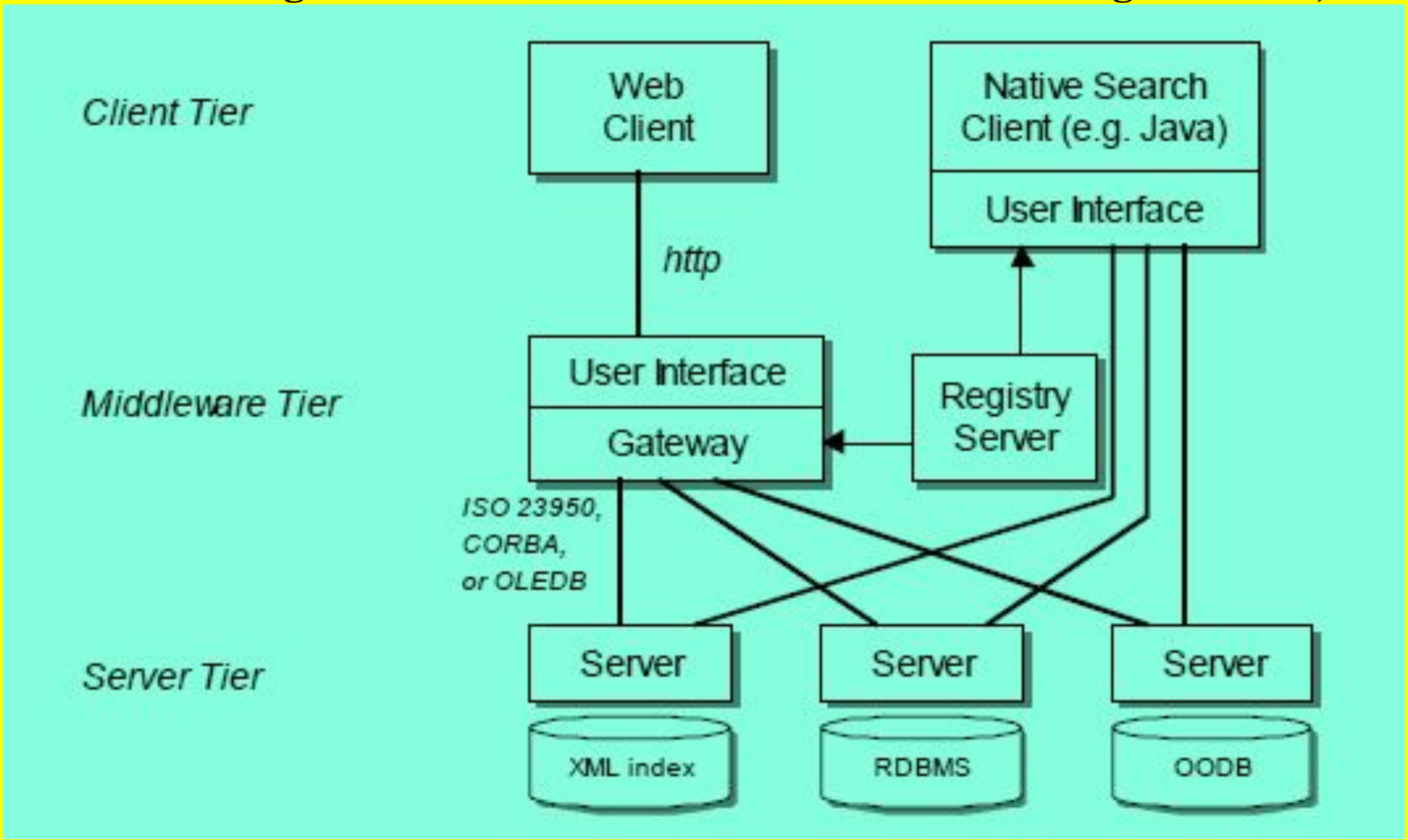
5. The Distributed Catalogue is implemented using a **multi-tier software architecture** that includes (Figure 4.5):

1) A Client Tier;

2) A Middleware or “Gateway” Tier;

3) A Server Tier.

Fig.4.5 – Implementation view of Distributed Catalog Services (CORBA – Common Object Request Broker Architecture; OLE DB – Microsoft's strategic low-level interface to data across an organization)



4.4.1. Catalogue Server/Service implementation development

1. To encourage widespread participation in the Clearinghouse, Catalogue Service software has been developed under direction of the FGDC and other coordination organizations around the world.

2. Reference implementations of software exist to provide a free or low-cost example of metadata management and Distributed Catalogue service that can be quickly implemented.

3. The software can also be used as reference by commercial developers to test anticipated functionality and

4. A Catalogue Service that participates in a Distributed Catalogue should fulfill the following requirements:

1) Support of a standard protocol (ISO 23950 preferred) for search and retrieval on an Internet-accessible server;

2) Linkage to an indexed metadata management system that:

a) Supports multi-field queries on text, numeric, and extended data types;

b) Can return entries in a structured form that are or can be converted into a requested report in HTML, XML, and text. This may be:

– A relational database;
– An object-relational database;
– An XML database;

3) Ability to translate public fields/attribute structures into names and structures used in the metadata management system using a national or international vocabulary (ISO 19115, when available);

4) Ability to add, update or delete Metadata Entries in the metadata management system.

4.4.2. Catalogue Gateway and access interface implementation development

1. As depicted in Figures 4.4 and 4.5, there is often a need for an intermediary to provide application integration for an end user.

2. Known as **"Application servers" or Middleware**, these hosts allow for the storage, construction, and download of user interfaces to end users and communicate with multiple Catalogue Servers simultaneously – a feat not supported by many web browsers due to security settings.

3. Software systems, such as Application Servers, that integrate catalogue search and other GIS and mapping functions benefit from the community development of software development kits (SDKs) based on standards.

4. SDKs can provide client and server libraries for catalogue search and other services based on standard interfaces.

5. Through component architecture, these SDKs expedite development of advanced software by combining appropriate pieces of software together as needed, reducing the need for a programmer to learn the intricacies of a given service.

4.4.3. Implementation registering of Catalogue Servers

1. The operation of a growing network of Distributed Catalogue Servers requires the management of server-level information in a central location.

2. This registry server, shown in Figure 4.5, essentially houses server or collection-level metadata for search and retrieval and use in distributed query.

3. In this way a search may be first made of the Registry of Servers to identify candidate servers to target the query:

– And as a broker, the registry returns the list of likely targets based on criteria such as geographic and temporal

4. A registry facility greatly improves the scalability of a national, regional or global network of Catalogues.

5. In the context of the GSDI, a coordinated registry of catalogue (and other) services is needed.

6. If all Catalogues were registered into a common and distributed registry, resolution of appropriate hosts of spatial information globally will be enabled.

7. A coordinated registry between the U.S. and Canada is proposed through an interagency agreement between the FGDC/GSDI Secretariat and Geomatics Canada:

– As a model for other countries to follow in managing and coordinating their own national catalogue entries with the