An Operational Metadata Framework for Searching, Indexing, and Retrieving Distributed Geographic Information Services on the Internet

Ming-Hsiang Tsou

Department of Geography, San Diego State University, CA 92182-4493, USA
http://typhoon.sdsu.edu/tsou/index.html
mtsou@mail.sdsu.edu

[1]**Abstract:** A comprehensive metadata scheme for distributed geographic information services should include multiple types of information services, including geodata objects, software components, and web map services. This paper examines the existing metadata standards and their implementation frameworks and presents an operational, object-oriented, hierarchical metadata architecture as an alternative solution for searching, indexing, and retrieving distributed GIServices on the Internet. An operational metadata framework can facilitate the establishment of self-manageable, self-describable GIS web services, which can be freely combined and used on the Internet. Hierarchical metadata repositories can provide a meaningful metadata archive structure and can improve metadata search mechanisms, where geospatial datasets and services are grouped and organized by their unique features or functions. By collaborating with operational metadata contents and hierarchical metadata repositories, the new metadata framework will help users and systems to access on-line geodata objects, software components, and web map services efficiently and effectively.

1. Introduction

Along with the increasing volume of geospatial data, the storage and management of GIS databases has become a major challenge for scientists and GIS professionals. Distributed geographic information services (GIServices) are one of the possible solutions for the management of very large-size GIS databases. However, it is currently difficult to access distributed GIS datasets and web mapping services remotely due to their heterogeneity. Many research projects, including digital

---

1 Tsou, M.H. (2002) An Operational Metadata Framework for Searching, Indexing, and Retrieving Distributed Geographic Information Services on the Internet. In *Geographic Information Science (GIScience 2002)*, M. Egenhofer and D. Mark (eds.), Lecture Notes in Computer Science Vol. 2478, Springer-Verlag, Berlin, pp. 313-332.

libraries, data clearinghouses, data mediators, are focusing on the management issues of distributed geographic information services.

The main problem for the management of distributed GIServices is the heterogeneity of geospatial data models and formats required in different GIS applications. Currently, one of the popular solutions is to create metadata associated with geospatial data items and services, which can be interpreted by users or metadata search engines. Metadata becomes the key to bridge the heterogeneous environments of distributed GIS databases and services and to provide users with the semantics and syntactic of GIS databases [1]. However, the management of GIS metadata records is problematic under current rational database approach [2].

This paper will examine the existing metadata frameworks developed by Federal Geographic Data Committee (FGDC) and International Organization for Standardisation (ISO) Technical Committee 211 (ISO/TC211). An operational, object-oriented metadata architecture will be introduced to provide an alternative framework for searching, indexing, and retrieving distributed geographic information services via the Internet. By collaborating with operational metadata contents and hierarchical metadata repositories, the new metadata framework can help users and systems to access on-line geodata objects, software components, and web map services efficiently and effectively.

2. The Development of Geospatial Metadata Standards

Metadata are usually defined as data about data. Some researchers describe metadata as the abstraction of representational details or representation of domain knowledge [3], [4], [5]. In this paper, the definition of metadata focuses on the operational meaning for distributed geographic information services as the following – *metadata is the information which can facilitate users or computer systems to access, archive, and manipulate centralized or distributed information services, such as data objects, software components, and web services.* The adoption of operational metadata will make geographic information objects and services self-describable and self-manageable in distributed network environments.

The early development of metadata in GIS applications began at the Federal level with the work of the Spatial Data Transfer Standard (SDTS) committee in 1980's [6]. The goal of SDTS is to provide a common ground for data exchange by defining logical specifications across various data models and structures [7], [8]. Fifteen years later, a content standard for digital geospatial metadata (CSDGM) was approved by FGDC on June 8, 1994. The CSDGM includes seven major components, which are identification, data quality, spatial data organization, spatial reference, entity and attribute, distributed information, and metadata reference information. Hundreds of fields are required to be filled to complete a comprehensive, standardized metadata record [9].

The FGDC released the second version of the content standards in 1998, which modified some production rules for easy implementation of metadata. The new version also added two new functions for the CSDGM: the definition of profiles and User Defined Metadata Extensions. A profile is a subset of the standard metadata elements that describes the application of the FGDC Metadata Standard to a specific user community. For example, the biological research community can define their own profiles for biological/ecological data sets, such as vegetation, land use, and habitats [10]. Profiles are formalized through the FGDC standards process [11]. User Defined Metadata Extensions are extended metadata elements to the original Standard. A specific research discipline can define a set of extended metadata entities for their specific applications. For examples, a remote sensing community can define the metadata extensions for remote sensing research [12]. Although the concepts between metadata profiles and extensions are very similar, the main difference is that the metadata extensions emphasize the new metadata elements outside the original Standards, whereas the metadata profiles focus on the modification of existing standards.

Besides the FGDC's metadata standards, one of the significant international metadata standards is the ISO 19115 Metadata Standard (previous published as ISO15046-15) created by the International Organization for Standardisation (ISO) Technical Committee (TC) 211. The ISO metadata standards proposed a conceptual framework and an implementation approach for geospatial metadata which were developed partially based on the 1994 FGDC standards [13].

The ISO 19115 metadata standard is one of the most comprehensive (but also the most complicated) metadata schemes for distributed GIServices. The framework of geospatial metadata specified by ISO/TC 211 includes three conceptual levels: a data level, an application level, and a meta-model level. Each level highlights different aspects of the metadata model and its relationship to geographic datasets (Figure 1).

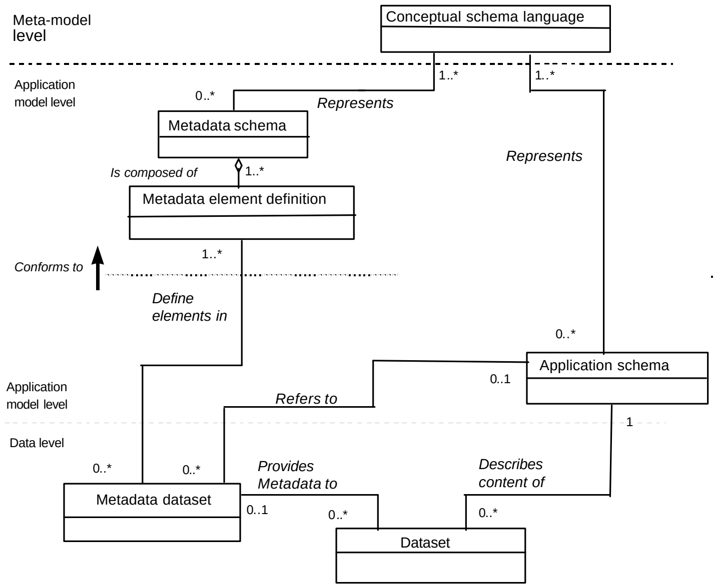


Fig. 1. ISO/TC 211 metadata relationship [14] (p. 22).

The element in the highest level (meta-model level) of ISO metadata standards is the metadata schema language, which is used to describe a conceptual metadata schema and an application schema at the application model level. The metadata schema provides the metadata element definitions for a metadata dataset. A metadata dataset describes the administration, organization, and content of a dataset at the data level [14]. Similar to the FGDC metadata attributes specifications, the design of mandatory, conditional, and optional items in the ISO 19115 metadata standard allows the implementation of metadata standards to become more flexible and dynamic and to be easily adopted in a distributed network environment.

Another feature of the ISO metadata standards is to provide a language-based implementation framework for metadata structure and encoding. ISO/TC 211 suggests that metadata software will support the input and output of metadata entries using the Standard Generalized Markup Language (SGML), as defined by ISO. Each metadata entry will be encoded as a SGML document entity including a SGML declaration, a base Document Type Declaration (DTD), and the start and end of a base document element. That is the same format as the Extensible Markup Language (XML). The XML-based ISO metadata standard will become a major advantage for the future implementation of metadata datasets, especially for web-based applications.

In general, the use of metadata can facilitate the identification, interoperability and auto-transfer functions of distributed GIServices. A comprehensive metadata structure is essential for the future development of open and distributed GIServices [15]. However, complicated metadata standards, such as ISO 19115 and CSDGM, may undermine the widespread use of metadata and their implementation procedures. The construction of metadata should be flexible and have alternative methods for different data types because metadata are both data-oriented and application-oriented.

This paper suggests that a comprehensive metadata scheme for distributed GIServices should include multiple types of information services, such as geodata objects, software components, and web map services. Also, the metadata for distributed GIServices should focus on the operational functions instead of descriptive contents. An operational metadata framework can facilitate the establishment of self-manageable, self-describable GIS web services, which can be freely combined and used across the Internet. The following section will focus on the actual design of operational metadata frameworks for geodata objects, software component, and web map services.

3. The Design of Operational Metadata Frameworks

Traditional metadata schemes developed by FGDC and ISO/TC211 emphasized the establishment of a standardized format and adopted relational database concepts, where each metadata item is represented as an individual record or an XML document. The standardization of metadata formats may be problematic in actual implementation because a single standard may not be appropriate for heterogeneous geospatial datasets. For example, a single metadata standard would be inadequate to simultaneously describe both a vector data model and a raster data model, without many extraneous fields. Likewise, the metadata standards for remote sensing images are often quite different from biological reservation data sets.

Another problem with traditional metadata schemes is that their GIS relational database design detaches metadata from their associated data, which jeopardizes metadata availability when geodata objects are moved or modified [2], [16]. The situation will worsen in distributed network environments where geodata objects and associated metadata records often need to be transferred and copied from one location to another. To prevent the loss of metadata records, one possible solution is to create encapsulated metadata schemes that adopt object-oriented modeling techniques and embed metadata as encapsulated items within the data itself (Figure 2).

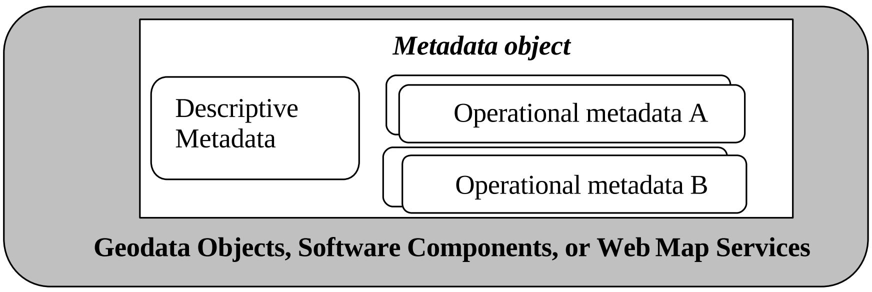


Fig. 2. An object-oriented metadata scheme.

Figure 2 illustrates a possible framework for object-oriented metadata objects, which include two types of metadata components: descriptive metadata and operational metadata. Descriptive metadata components include traditional metadata information that can mainly be interpreted by users, such as data descriptions, distribution information, and metadata reference, etc. Operational metadata components contain machine-readable information which can be applied automatically in specific GIS operations, such as map display, spatial analysis, GIS modeling, etc.

The design of operational metadata can facilitate dynamic interactions and integrations among geodata objects, software components (programs), and web map services. An example of the design of three different metadata objects for GIServices metadata is illustrated in Figure 3.

Geodata object metadata will facilitate the access and distribution of geodata objects among heterogeneous GIS databases. The contents of geodata object metadata can help users and computer systems to index, access and manipulate geodata objects in distributed network environments.

Software component metadata will help automatic interactions between software components, (such as Java applets, Microsoft .NET objects, and CORBA objects) and GIS applications across different computer platforms. The design of software component metadata will focus on cross-platform settings, remote procedure calls, GIS operation requirements, and the registry of component functionality, etc.

Web map service metadata may be used for advertising and broadcasting available web map services. Currently, many web map services are under development, such as ESRI's ArcIMS, INTEGRAPH's GeoMedia Web Map, and AutoDesk's MapGuide map services. However, current GIS metadata frameworks are not providing a needed solution for indexing and cataloging web map services. This paper will introduce web map service metadata to facilitate GIS users in publishing/accessing web map services on the Internet.

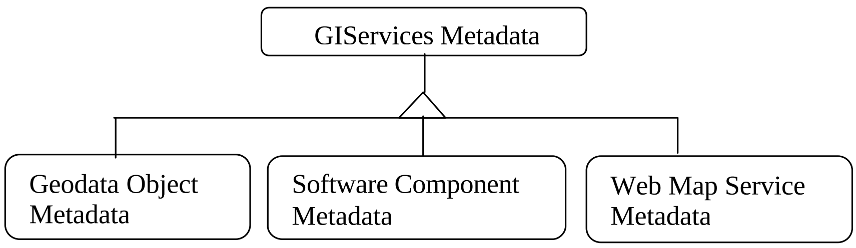


Fig. 3. The hierarchy of GIServices metadata objects.

In general, the three proposed types of GIServices metadata illustrate a task-oriented framework for the identification and management of distributed GIServices. Different GIServices require different kinds of metadata objects to facilitate the access, distribution, and adoption of GIS data objects, software components, and web map services. The following section will introduce the specifications for these metadata objects in detail.

3.1 Operational metadata for geodata objects

The design of operational metadata for geodata objects needs to consider what kinds of operations or tasks are associated with geodata objects. Four representative tasks are proposed here for the specifications of geodata object metadata, which include *map display, spatial query, spatial operation, and data connectivity* (Figure 4). The actual implementation of geodata object metadata could include more tasks or elements based on specific task needs or application requirements.

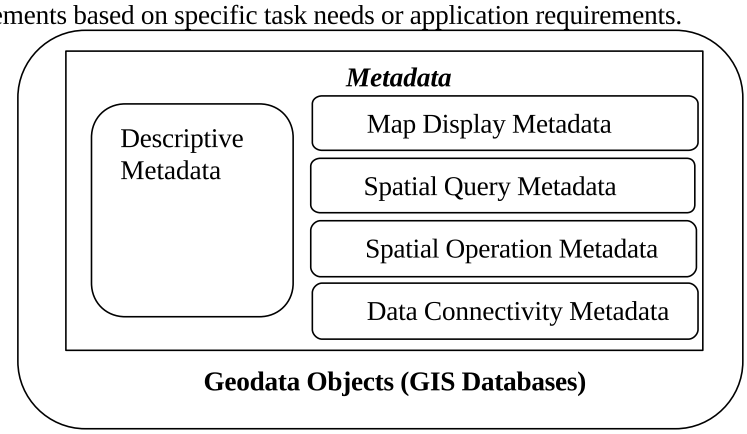


Fig. 4. The design of geodata object metadata.

The design of the map display metadata element is to specify the representation methods of geodata objects on electronic media or computer screens. The contents of the map display metadata may include the following items:

? Feature type (raster/vector, point, line, polygon, or volume)
? Attribute type (nominal, ordinal, interval/ratio, or multiple attributes)
? Map symbols (attribute lookup table, symbol size, symbol icons/shapes)
? Color scheme (2 bits-B/W, 8bits-256 color, 32 bits-true color)
? Scale threshold

These metadata contents can be interpreted automatically by mapping software or web map services to apply both a color scheme and map symbols dynamically. With the help of map display metadata, geodata objects become self-describable and self-manageable map layers. Web map users can decide whether they want to manually change color schemes and symbols or just adopt the default settings configured in the map display metadata. One thing to note is that the definition of map symbol should consider the dynamic environment of distributed mapping services with different computer display techniques and screen resolutions. For example, if a line symbol is displayed on a Pocket PC with small screen resolution (300x200), the width of the line symbol will be adjusted automatically according to the size of screen.

The design of spatial query metadata is to describe the GIS query requirements of geodata objects. The contents of spatial query metadata will include the following items:

? Query language (natural language, SQL or other spatial query languages)
? Query syntax
? Query interface (the interface which can provide remote access point)
? Results display (the output format and holdings of spatial query results)

Spatial operation metadata will specify the possible spatial operations associated with geodata objects and their requirements. For example, a “road” data object is usually associated with the “buffer” operation or the “network analysis” operation. The contents of spatial operation metadata will include the following items:

? Associated GIS operations (overlay, buffering, network analysis, etc.)
? Data format (acceptable by GIS programs)
? Operation logs (the history logs of spatial operations)

Data connectivity metadata focus on the mechanisms of remote access and download procedures for geodata objects. The design of data connectivity metadata will specify the interactions between geodata objects and remote machines or databases. The contents of data connectivity metadata will include the following items:

? Local access metho ds (the communication in a single machine)
? Remote access methods (remote database connections)
? Data compressing/uncompressing methods (Wavelets, gzip compression, etc.)
? Registration of data objects

3.2 Operational metadata for GIS software components

The design of operational metadata for GIS software components needs to consider what kinds of operations or tasks are associated with GIS software components. Four representative tasks are illustrated here for the specification of GIS software component metadata contents, which include *GIS data input requirements, GIS data output specifications, run-time system requirement, and component registration* (Figure 5).

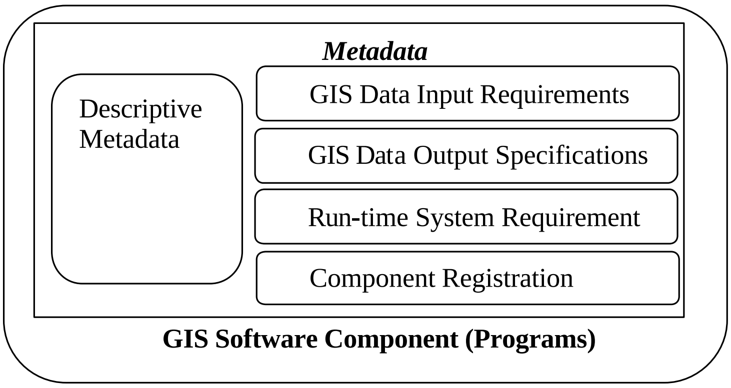


Fig. 5. The design of GIS software component metadata.

GIS data input requirement metadata will specify the requirements of GIS input data for GIS operations. For example, the map display operation will require projections and coordinate systems in order to display multiple map layers properly. The contents of GIS data input requirement metadata will include the following items:

? Data input format (DLG, Shapefiles, or SDF)
? Data uncertainty threshold (the feasible range of data accuracy for this operation)
? Prerequisite (coordinate systems, projections, topology, etc.)
? Component category (overlay, network analysis, or hydrological modeling)

GIS data output specification metadata focus on the specifications of output data objects generated by GIS software components. These information items can be used by subsequent GIS operations. The principal reason for creating data input and data output metadata is to facilitate the combination of GIS operations and procedures. For example, users can combine "buffer" and "reselect" operations together if the "buffer" output specification meets the requirement of the "reselect" input requirement. The contents of GIS data output specification metadata will include the following items:

? Data output format (DLG, Shapefiles, or SDF)
? Data uncertainty threshold (the range of data accuracy after this operation)
? Operation effects (the change of data characteristics after this operation)

The design of run-time system requirement is to facilitate cross-platform GIS components and applications. Different GIS components may require a unique run-time environment, such as the local disk size, CPU speeds, and the size of RAM. These information items can help distributed users or systems to make sure that distributed GIS software components are working properly in remote machines. The contents of run-time system requirement will include the following items:

? Hardware requirements (CPU speed, temporary disk size, RAM, etc)
? Virtual machine requirements (Java or Microsoft Virtual machine)
? Component profiles (component size, type, and run-time efficiency)

Component registration metadata will be used to register GIS components via universal registry services or web service registration frameworks on the Internet. The registration of GIS components will allow Internet GIS users to know where to find the component or programs they need. The contents of component registration metadata will include the following items:

? Unique Component ID (for registration)
? Functionality classification (hydrological modeling, map overlays, etc.)
? Possible GIS applications (urban planning, natural resource management, etc.)

3.3 Operational metadata for web map services

The design of operational metadata for web map services is to facilitate the access/distribution of web map services and the dynamic integration of multiple web map services. This paper introduces four representative tasks for the specifications of web map services metadata: *access method, display type, map capabilities, and map theme* (Figure 6).

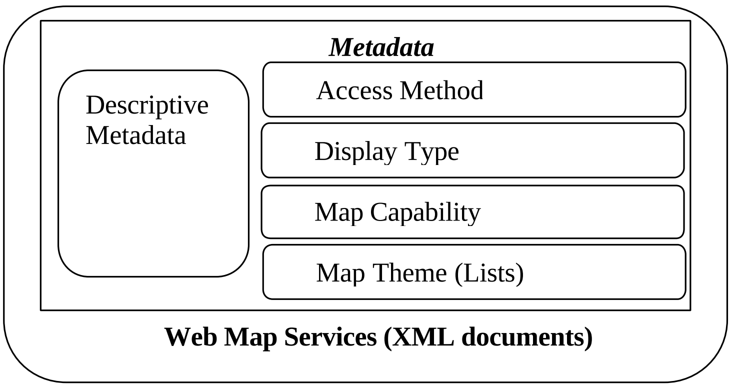


Fig. 6. The design of web map service metadata.

Access method metadata indicates the requirement of remote access methods to web map services. Web map users and computer systems can access the metadata information to create a link between their local GIS applications and remote web map services. For example, the county of San Diego can create a dynamic mapping service which combines the real-time weather map services from National Oceanic and Atmospheric Administration (NOAA) and local San Diego highway maps to provide the safe driving information for local commuters. The contents of access method metadata will include the following items:

? Web-based map browser requirements
? Communication protocols (TCP/IP or SOAP)
? Communication language (XML or HTML)
? Bandwidth requirements (10K or 10MB per second)

Display type metadata will specify the display mechanism and map rendering functions for web map services. These information items can be used on a remote computer to check if the targeted GIS client is capable of displaying web maps or requires pre-process operations. The contents of display type metadata will include the following items:

? Map Format (vector data or raster images and their rendering algorithms)
? Compression methods (wavelets or gzip)
? Display requirement (Minimum/Maximum screen resolution, color depths)

Map capability metadata will specify the available functions of web map services, such as spatial query, on-line buffering, or network analysis. Different web map services may possess different display functions and GIS operations. The contents of map capability metadata will include the following items:

? Map function list (Zoom, Pan, Query, Table display, Print, etc.)
? Required user interfaces (keyboards, mouse, new devices, 3D glasses)
? Additional information (Help file, tutorials, etc.)

Map theme metadata will describe available map themes provided by web map services. For example, a USGS Internet map server may provide roads, hydrological features, and contour lines in a single web map service. Each map theme metadata will include the map extent of individual layers that can be used for a spatial search. The contents of map theme metadata will include the following items:

? Map theme lists (multiple items associated with individual map extents)
? Web map extent
? Data objects links (to access multiple data objects in a single web map service)

In summary, this paper introduced an operational metadata framework which emphasizes three important concepts for the use of metadata. First, the new design changes the traditional functions of metadata from descriptive information into task-oriented, operational, machine readable metadata contents. The GIS processing-oriented metadata scheme will facilitate distributed GIS processing, accurate map display, and automatic data conversion across the networks. Second, the encapsulation of metadata into data objects will protect the metadata from being lost in the network environment and prevent accidental intervention in critical metadata content. Third, an operational metadata scheme can be applied to different

GIServices, including geodata objects, GIS software components, and web map services. The next section will discuss the actual implementation of metadata frameworks for distributed GIServices.

4. The Implementation of GIService Metadata Frameworks

Under a traditional relational database framework, there are two types of metadata implementation approaches. One is to create a centralized metadata database or catalog. The other is to establish distributed metadata repositories that can be accessed via an information gateway server. The goal of both approaches is to help users to index, archive, and search distributed GIServices. This section will first discuss the advantages and disadvantages of both approaches under current metadata frameworks. A hierarchical metadata repository framework will be introduced next as an alternative method in implementing metadata for geodata objects, GIS software components, and web map services.

4.1 Digital libraries and data clearinghouses

The first approach might be called the "digital libraries" solutions. This approach creates a centralized metadata database or catalog containing millions of metadata records within standardized metadata formats. Each metadata entity is represented as a record in relational database engines, such as Microsoft Access or IBM's DB2. Some digital libraries also include the extended functions of gazetteers to facilitate place name queries and retrievals. The user interfaces of on-line digital libraries are usually web-based in order to facilitate the remote access of centralized metadata records. The process of metadata query and retrievals are processed on the server-side computers. Figure 7 illustrates an interface example from Alexandria/California Digital Library.

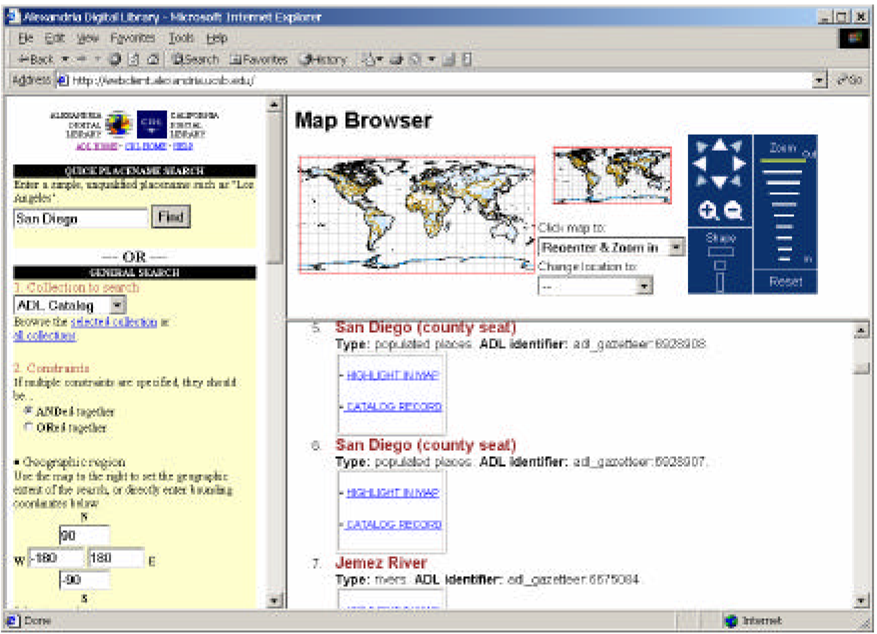


Fig. 7. Alexandria Digital Library with a centralized metadata catalog.

The second approach for metadata implementation is the "data clearinghouse" approach that utilizes the Z39.50 protocol to index and access multiple metadata repositories remotely. At present, FGDC's National Spatial Data Infrastructure (NSDI) and associated data clearinghouse nodes adopt this approach. The full name of the Z39.50 protocol is *"ANSI Z39.50-1995*, *Information Retrieval Application Service Definition and* Protocol Specification" [17]. Z39.50 is a US national standard defining a protocol for a client/server information retrieval. Z39.50 was first approved by American National Standards Institute (ANSI) in 1988. It was then extensively revised in 1992 and 1995. The protocol was originally proposed for use with bibliographic information from the library information community. The protocol specifies formats and procedures governing the exchange of messages (a request or a response) between a client and a server. The client can send a request to the server to search a database and identify records that meet specified criteria, and to retrieve those identified records [18].

In general, the approach of a distributed metadata repository (data clearinghouse nodes) can provide more flexible and scaleable frameworks for geospatial metadata compared to digital library solutions. The Z39.50 protocol can provide a more abstract view of remote databases compared to rigid structured query languages (SQL) used by relational databases. Z39.50 only deals with logical entities stored in the remote metadata files instead of specific database implementation [17]. Figure 8 illustrates the web-based interface of NSDI's clearinghouse search form.

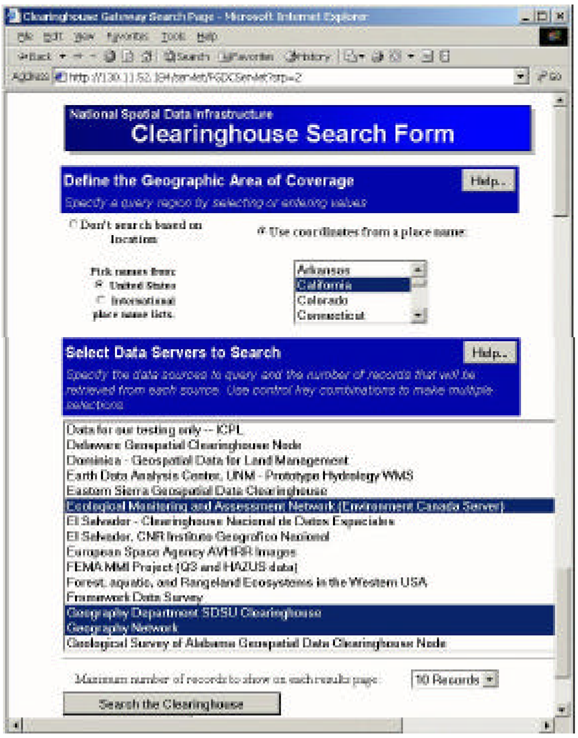


Fig. 8. The FGDC Geospatial Data Clearinghouse Search Form.

Besides the flexibility of metadata frameworks, one unique feature of the data clearinghouse approach is its capability to query multiple metadata repositories at the same time via Z39.50 protocol. Distributed data clearinghouse nodes can provide this unique function by installing a software package, called ISITE, which is highly recommended by FGDC for the management of metadata indexing and query. ISITE is a software package developed by the Center for Networked Information Discovery and Retrieval (CNIDR) (http://www.cnidr.org). ISITE has a built-in search engine (Isearch) for the indexing of metadata files. It includes the Z39.50 communication applications under the Transmission Control Protocol/Internet Protocol (TCP/IP) level [19]. Figure 9 illustrates the mechanism of querying multiple metadata repositories via the Z39.50 protocol.

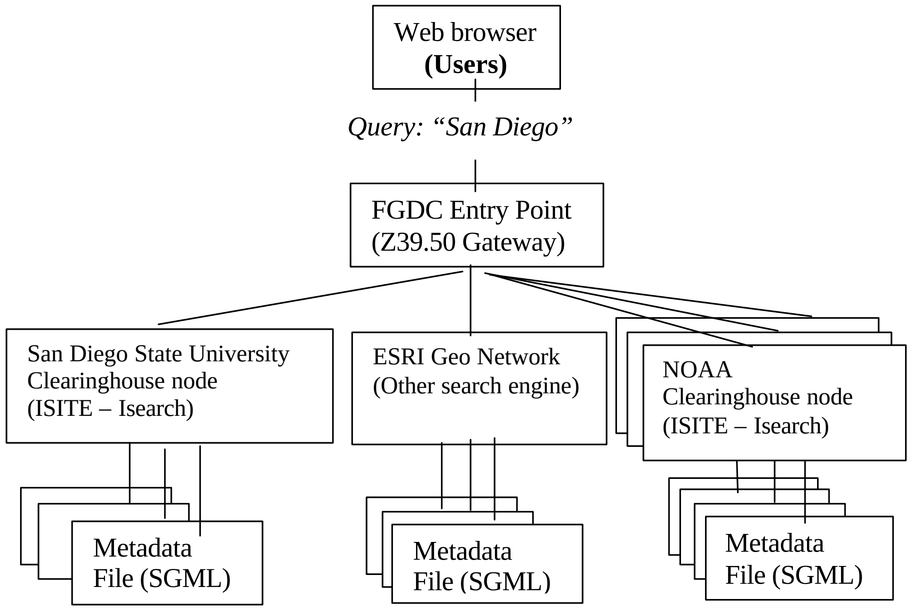
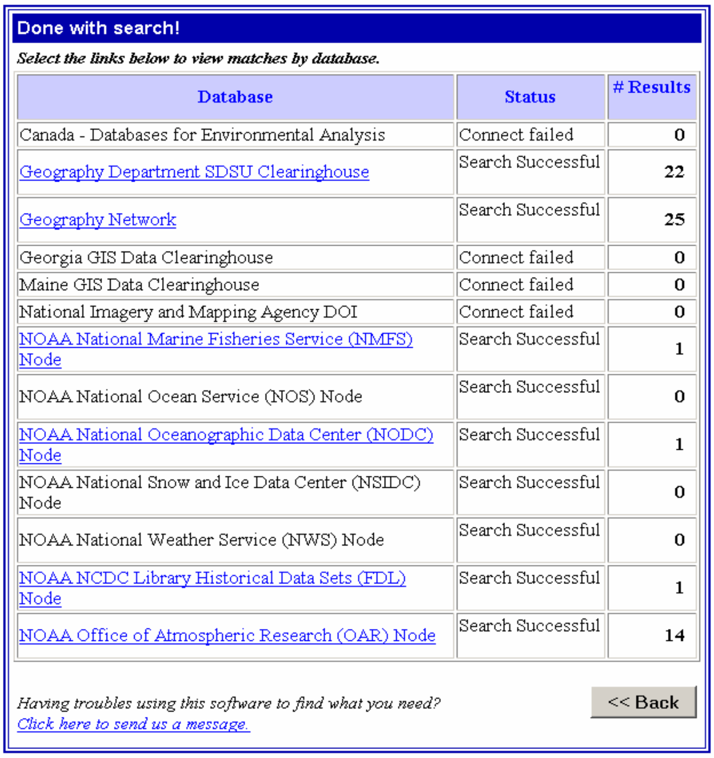


Fig. 9. The mechanism of querying multiple FGDC's clearinghouse nodes.

The top textbox illustrates that one metadata query (search for "San Diego") was initiated from a user's web browser. The web browser then accessed a FDGC's entry point node that has a web server with Z39.50 gateway functions. The entry point server then distributed the user's query to multiple clearinghouse nodes simultaneously, including San Diego State University clearinghouse node, ESRI's Geography Network, and NOAA's clearinghouse node. The ISITE software in each local clearinghouse node has already indexed their metadata records on a regular basis. When each clearinghouse node received the request from the FGDC's Z-gateway server, their local ISITE software uses their local Isearch program to search their metadata index records and then send the results back to the FGDC's entry point. The FGDC's entry point then combined the query results and displayed them on the user's web browser (Figure 10).

In this example, a query for "San Diego" text was sent out via the FGDC's entry point and the result indicate that 22 metadata files contained "San Diego" in San Diego State University's clearinghouse and a further 25 metadata files in ESRI's Geographic Network.



Done with search!

Select the links below to view matches by database.

Database	Status	# Results
Canada - Databases for Environmental Analysis	Connect failed	0
Geography Department SDSU Clearinghouse	Search Successful	22
Geography Network	Search Successful	25
Georgia GIS Data Clearinghouse	Connect failed	0
Maine GIS Data Clearinghouse	Connect failed	0
National Imagery and Mapping Agency DOI	Connect failed	0
NOAA National Marine Fisheries Service (NMFS) Node	Search Successful	1
NOAA National Ocean Service (NOS) Node	Search Successful	0
NOAA National Oceanographic Data Center (NODC) Node	Search Successful	1
NOAA National Snow and Ice Data Center (NSIDC) Node	Search Successful	0
NOAA National Weather Service (NWS) Node	Search Successful	0
NOAA NCDC Library Historical Data Sets (FDL) Node	Search Successful	1
NOAA Office of Atmospheric Research (OAR) Node	Search Successful	14

Having troubles using this software to find what you need?
Click here to send us a message.

<< Back

Fig. 10. NSDI Search Results from multiple clearinghouse nodes.

The FGDC's data clearinghouse approach provides a more flexible and efficient query mechanism for multiple metadata repositories compared to traditional digital library solutions. However, current implementation frameworks suggested by FGDC may have several potential problems for distributed GIServices in the future.

First, FGDC places all distributed clearinghouse nodes on the same level (under the entry point gateway server) without any classification. GIS users will have difficulty in deciding which clearinghouse nodes may contain metadata they seek. If the number of clearinghouse nodes exceeds a user's perception, the user may either give up using data clearinghouse nodes or try to apply all of the nodes available. The query results could contain hundreds of nodes which a user may not be able to process or evaluate. One possible solution is to create a hierarchical metadata repository mechanism which groups similar clearinghouse nodes or metadata together on multiple levels. The hierarchical metadata repository structure is introduced in next section.

A second problem with FGDC clearinghouse nodes is that the results of a metadata query are stored by individual clearinghouse servers. Due to the limitation of ISITE software, it is difficult to create an integrated list of records from multiple databases or to collect them as a new metadata catalog. This limitation will prevent subsequent query functions.

A third problem is that each data clearinghouse needs to register their nodes via the FGDC's gateway server. However, many GIS projects and institutes may not be able to register their data clearinghouse due to the incompatibility of their metadata engines or web servers. FGDC's clearinghouse mechanisms only emphasize the concepts of information retrieval rather than resource discovery. Without proper registration of data clearinghouse nodes, many on-line GIS datasets and resources are invisible for GIS users under current frameworks. One possible solution is to utilize software agent technologies to automatically search, index, and create metadata records and repositories across the networks. Such mechanisms are called "web robots" or "information spiders", when applied in Web search engines.

These potential problems of metadata implementation frameworks will require the reconsideration of fundamental metadata model design and index services architecture. This paper introduces an operational metadata framework and a hierarchical metadata repository architecture with potential to provide a better solution. The next section introduces a design for hierarchical metadata repositories.

4.2 Hierarchical metadata repositories

The registration framework of current FGDC's data clearinghouses is horizontal. As hundreds of data clearinghouse nodes are registered at the same level, GIS users have difficulty when specifying required nodes from the hundreds of possible selections. One possible solution is to create a hierarchical framework for metadata repositories (Figure 11). Geospatial datasets can be grouped or organized by their themes or spatial locations under this framework. By adopting a hierarchical metadata repository structure, GIS users can easily search, index, and distribute geodata objects, software components, and web map services on the Internet.

Figure 11 illustrates a hierarchical metadata repository structure for geospatial datasets. In this example, two data objects, [San Diego Land use] and [San Diego Roads] associated with their metadata objects are registered in [San Diego Atlas Metadata Repository]. The [San Diego Atlas Metadata Repository] is registered under the [California Atlas Metadata Repository].

The advantage of such a hierarchical structure for metadata repositories is that it can provide a meaningful metadata archive structure and can improve metadata search mechanisms for GIS users or systems. Each parent metadata repository can relay users' requests to its child level nodes and the search results can be integrated on the parent level and sent back to users. Another advantage is that one data object or map service can be registered under multiple repositories at the same time. Multiple metadata registrations allow flexible access channels for GIS users. For example, a [San Diego Road] data object can be registered under both the [San Diego Atlas Metadata Repository] or [California Transportation System Metadata Repository].

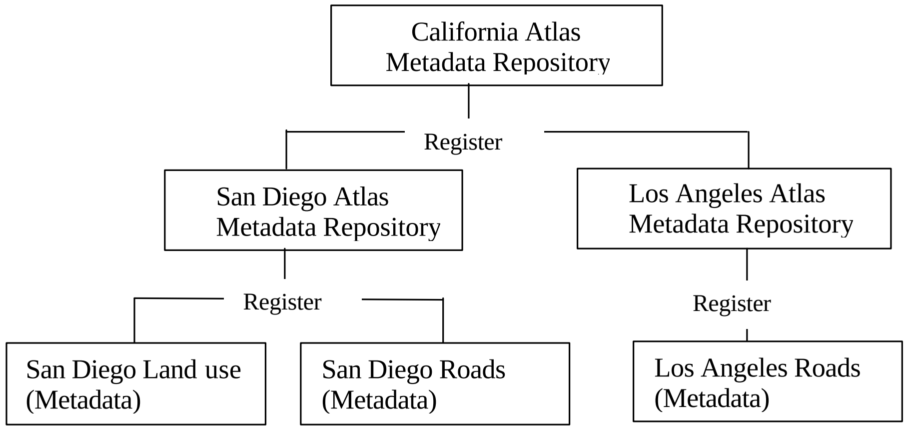


Fig. 11. Hierarchical metadata repository framework for geospatial datasets.

The hierarchical metadata repository framework can also be applied on GIS components and web map services. Figure 12 illustrates a hierarchical structure for GIS component metadata. In this example, a [TIN Model] component can be registered in a [Hydrological Modeling Tool Metadata Repository], which is the child node of [Spatial Analysis Tool Box Metadata Repository]. This framework also allows multiple registrations. For example, a [Buffering] software component can be registered under both [Location Analysis Tool Repository] and [Network Analysis Tool Repository].

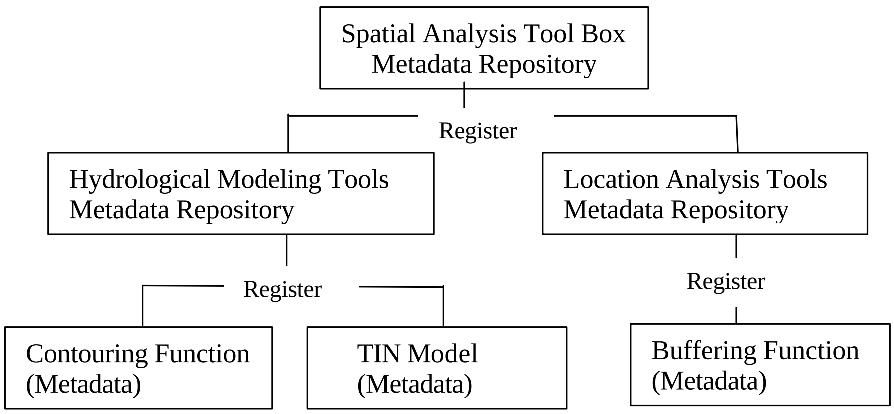


Fig. 12. The hierarchy of GIS component metadata repositories.

The hierarchical metadata structure for distributed GIServices will facilitate the distribution/access of geodata objects, software components and web map services.

The data owners or GIS software programmers can register their products in multiple metadata repositories based on unique data features or GIS functions. GIS users and applications can utilize the hierarchical metadata repositories to search for the data/programs they need under specific categories rather than search through thousands of items from unorganized data clearinghouses.

4.3 A hypothetical GIS operation example

In order to demonstrate the capability of operational metadata and hierarchical registration framework, this section introduces a hypothetical GIS example. A GIS spatial analyst, Jack, wants to locate a new grocery store in Boulder, Colorado. He needs to obtain related map information and perform a GIS overlay analysis for this task. The following criteria must guide the grocery store site selection:

1. The land use must be in a residential urban area.
2. The site must lie above the 500 year flood plain.

To accomplish this GIS task, Jack needs to gather [land use, Boulder] and [flood, Boulder] data objects. Jack searched the hierarchical data clearing house nodes under the Colorado Geodata Repository and found the two requested Boulder data objects. After downloading the data objects, the next step is to perform the "union" overlay operation. Jack accessed the on-line overlay operation toolbox and downloaded a Java applet called "union". With the help of operational metadata, the union applet used the "GIS data input requirement" metadata to verify the qualification of the [Land use] and [flood] objects (Figure 13). Since the data formats of the two objects were acceptable for the union applet, the overlay operation was executed and a new data object, [Criterion-A] was generated. The new [Criterion-A] object inherited the metadata attributes from its parent data objects (land use and flood). For example, the associated operation metadata would include overlay, buffer, and hydrological modeling. The same inheritance mechanism applies to the coordinate systems and map extents. A new operation log was saved in the new spatial operation metadata of [Criterion-A] automatically (Figure 13).

5. Discussion

This paper examined the development of metadata standards for distributed GIServices and introduced an object-oriented, operational metadata framework for distributed geodata objects, software components and web map services. It is suggested that an object-oriented metadata framework be implemented under a hierarchical metadata repository structure to facilitate efficient search/retrieval of metadata. The hierarchical structure provides a logical archive framework and improved access method for distributed GIServices. It is believed that the adoption of operational metadata frameworks and hierarchical repositories will provide GIS users more flexible and more efficient ways to access, index, and search metadata and distributed GIServices.

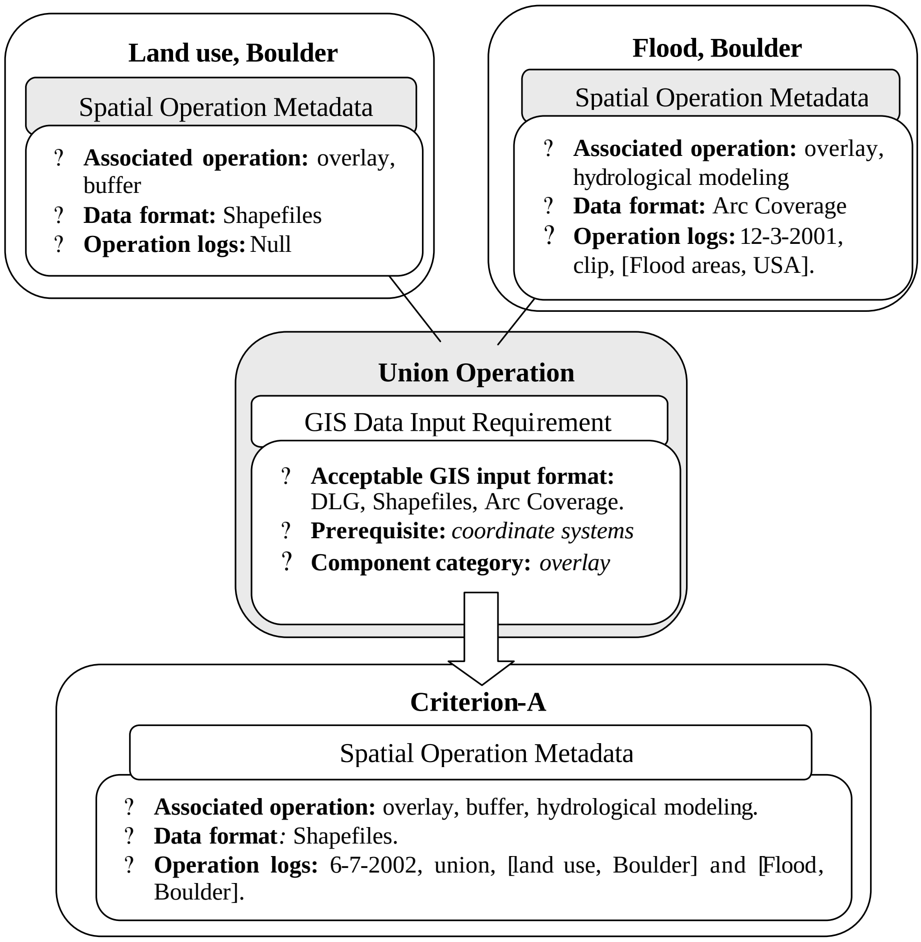


Fig. 13. A hypothetical GIS operation example.

There remains several unsolved problems in the metadata framework of distributed GIServices. First, current search mechanisms only focus on the contents of metadata with text-based keywords search engines. There is no semantic search mechanism behind these types of search. For example, if a user searches for “San Diego Roads”. Most metadata search engines will use “San Diego” as the keyword to query metadata files. However, it is very likely that some “California State Roads” may contain the roads in San Diego even though the metadata of “California State Roads” may not include any text related to “San Diego”. Therefore, a method of semantic search of distributed GIServices will be a further challenge for geospatial scientists. One possible solution is to create a “metadata ontology” or “metadata knowledge base” for semantic metadata search [5]. In the previous example, *“San Diego is part of the*

State of California" would become a knowledge rule, that can be applied by the metadata search engine.

The second problem with the existing metadata model is related to the differences between data attribute search and metadata search. For example, if a user needs to download a web map that includes the geometry features of the San Diego International Airport, the search will need to focus on the actual attribute data content query rather than the metadata content. Current metadata frameworks have difficulty combining both types of search at the same time due to the heterogeneous databases and software engines. The development of intelligent software agents may be able to solve this problem by connecting database engines with data objects dynamically and perform both attribute content search and metadata search simultaneously.

In summary, these problems all relate to the fundamental design of the metadata model for distributed GIServices. The GIS community needs to consider these metadata issues when developing their metadata frameworks. This paper only provides a conceptual design solution for metadata models and implementation frameworks. The actual implementation of operational metadata and hierarchical repository architecture will require collaboration among GIS professionals, federal governments and institutes, and private sectors to realize the full potential of an enhanced metadata model.

Acknowledgement

This paper forms a portion of the "Software Agent Communication Mechanisms" project, sponsored by San Diego State University and the "Web-based Geospatial Information Services and Analytic Tools for Natural Habitat Conservation and Management" project supported by NASA's Affiliated Research Center at San Diego State University. Funding by the NASA ARC program and matching funds from the San Diego State University Foundation are acknowledged and greatly appreciated.

References

1. Plewe, B.S., Johnson, S.R.: Automated Metadata Interpretation to assist in the use of unfamiliar GIS data source. In: Goodchild M.F., Egenhofer, M.J., Fegeas, R., Koffman, C.A. (eds.): *Interoperating Geographic Information Systems.* Kluwer Academic Publishers, Boston (1999) 203-214
2. Beard, K.: A Structure for Organizing Metadata Collection. In: *Proceedings of the Third International Conference on Integrating GIS and Environmental Modeling,* Santa Fe, New Mexico, CD-ROM. (1996) http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/santa_fe.html
3. Gardel, K.: A (Meta-) Schema for Spatial Meta-data. In *Proceedings of Information Exchange Forum on Spatial Metadata*, Reston, Virginia (1992) 83-98

4. Vckovski, A.: *Interoperable and Distributed Processing in GIS*. Taylor & Francis, London (1998)
5. Sheth, A.: Changing Focus on Interoperability in Information Systems: from System, Syntax, Structure to Semantics. In: Goodchild M.F., Egenhofer, M.J., Fegeas, R., Koffman, C.A. (eds.): *Interoperating Geographic Information Systems*. Kluwer Academic Publishers, Boston (1999) 5-29
6. Moellering, H.: Opportunities for Use of the Spatial Data Transfer Standard at the State and Local Levels. *Cartography & Geographic Information Systems*, Special Issue, **19**(5) (1992) 332-334
7. Fegeas, R.G., Cascio, J.L., Lazar, R.A.: An Overview of FIPS 173, The Spatial Data Transfer Standard. *Cartography and Geographic Information Systems*, **19**(5) (1992) 278-293
8. Wu, C.V.: *Object-Based Queries of Spatial Metadata*. Unpublished doctoral dissertation. Buffalo, New York: State University of New York at Buffalo, Department of Geography (1993)
9. Federal Geographic Data Committee (FGDC): *Content Standards for Digital Geospatial Metadata Workbook*, Version 1.0. FGDC/USGS, Reston, Virginia (1995)
10. Federal Geographic Data Committee (FGDC): *Content Standards for Digital Geospatial Metadata Part 1: Biological Data Profile*. FGDC-STD-001.1-1999. FGDC/USGS, Reston, Virginia (1999)
11. Federal Geographic Data Committee (FGDC): *Content Standards for Digital Geospatial Metadata*. (revised June 1998). FGDC, Washington, D.C. FGDC-STD-001-1998. (1998)
12. Federal Geographic Data Committee (FGDC): *Content Standards for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata (Public Review Draft).*: FGDC/USGS, Reston, Virginia (2000)
13. ISO/TC 211/WG 3: *Geographic Information – Part 15: Metadata*. ISO/TC 211-N538, ISO/TC 211-N538, ISO/CD 15046-15 (1998)
14. ISO/TC 211/WG 1: *Geographic Information – Part 1: Reference Model*. ISO/TC 211-N623, ISO/CD 15046-1.2 (1998)
15. Gardner, S.R.: The Quest to Standardize Metadata, *BYTE*, November 1997, **22**(11) 47-48 (1997)
16. Tsou, M.H., Buttenfield, B.P.: Client/Server Components and Metadata Objects for Distributed Geographic Information Services. In *Proceedings of the GIS/LIS' 98*, Fort Worth, Texas (1998) 590-599
17. Lynch, C.A..: The Z39.50 Information Retrieval Standard. Part I: A Strategic View of Its Past, Present and Future. *D-Lib Magazine*, April 1997. (1997) http://www.dlib.org/dlib/april97/04lynch.html
18. American National Standards Institute (ANSI). *ANSI/NISO Z39.50-1995: Information Retrieval (Z39.50): Application Service Definition and Protocol Specification.* Z39.50 Maintenance Agency, July 1995 (1995) http://lcweb.loc.gov/z3950/agency
19. The Clearinghouse for Networked Information Discovery and Retrieval (CNIDR). *The Isite information system version 1.04*, CNDIR, Research Triangle Park, North Carolina (1994) http://www.cnidr.org/software/Isite/guide.html