Use of spatial metadata in a LDAP/CORBA architecture

Steve Ramroop & Richard Pascoe

Department of Information Science, University of Otago, Dunedin, New Zealand
Phone: +64 3 479 5608 Fax: +64 3 479 8311
Email: sramroop, rpascoe @infoscience.otago.ac.nz

Presented at SIRC 2000 – The 12th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
December 10-13th 2000

ABSTRACT

Organizations associated with the use of Geographic Information Systems (GIS) have collected a large volume of spatial data. Descriptions of these data sets, called metadata, are seen as an important component to facilitate their access (Shah & Sheth, 1998). Most organizations have utilized some form of local or international metadata standard. To represent the diversity of the available metadata standards, a generic metadata standard is used to identify the common features from a sample of three metadata standards. The three standards are: Australia New Zealand Land Information Committee (ANZLIC) metadata standard; Content Standard for Digital Geospatial Metadata (CSDGM); and Dublin Core (DC) metadata standard. These standards are presented using hierarchical Directory Information Trees (DITs) associated with the Lightweight Directory Access Protocol (LDAP). LDAP DITs are accessed using the Common Object Request Broker Architecture (CORBA) which is an approach of particular interest to the Open GIS Consortium (OGC)[1] (Buehler & McKee, 1998).

Keywords and phrases**:** Metadata, LDAP, CORBA, OGC

1.0 INTRODUCTION

Descriptions of spatial data assists in the discovery and selection of data sets that are suitable for the application in which they are required. The information needed to create these descriptions is often readily available when spatial data sets are collected. During this time most organizations make the effort to catalog the descriptions associated with each data set. By cataloging the descriptions, the initial expense to document data sets is less than the cost of having to capture data sets which may have already been captured.

The descriptions for each data set are commonly known as *metadata* which is defined in Section 2. There are a number of metadata standards (Moellering, 1991) containing some descriptions that are common and others that are different. Nevertheless, from a general perspective they are compared to a generic structure which is discussed in Section 3. In this paper, the generic structure is related to three chosen metadata standards: the ANZLIC metadata standard, discussed in Section 3.1; the CSDGM metadata standard, discussed in Section 3.2; and DC metadata standard, discussed in Section 3.3. The comparison between the standards and the generic structure is presented in Section 3.4.

LDAP's hierarchical tree structure is adopted to form directory trees containing the descriptions of the three metadata standards. These trees are called *Directory Information Trees (DITs)* which are used to represent the hierarchical structure that exists in metadata standards. To define DITs, the relevant LDAP concepts are presented in Section 4. CORBA is used to access the LDAP DITs, as described in Section 5. Conclusions from this paper are presented in Section 6.

---

[1] http://www.opengis.org/

2.0 METADATA

Medycky-Scott et al. (1996, pp:67) defined *metadata* as `data about data' or `information that gives meaning to data'. They explained that metadata is a key enabling resource which can help facilitate the integration of geographic data. More specific definitions for metadata tend to reflect on the functional use, and content of metadata. For example:

- `Metadata is the critical information used to document and organize an organization's internal investment in spatial data and to provide information to process and interpret data received through a transfer from an external source' (Rossmeissl, 1998, pp:2)
- `Metadata represent information about the data in individual databases and data repositories' (Shah & Sheth, 1998, pp:2)
- `Metadata is data about the content, quality, condition, and other characteristics of data' (FGDC, 1997, pp:1)
- `Metadata can be viewed as a resource serving four functions: in data discovery; in data transfer; in data management; and in data use' (Medycky-Scott et al., 1996, pp:67)

In this paper, the definition adopted for metadata is a modified version of the `data about data' definition. This version includes the notion of metadata describing the geographic processing capabilities called *geoservices*, and metadata describing geographic data called *geodata*. These two notions are used by the OGC and are collectively known as *georesources*. Therefore, the modified definition for metadata is, `data about georesources'.

Shah & Sheth (1998) acknowledge that metadata is an important component to facilitate access to geodata. This is because metadata: provide descriptions of georesources; assist in the discovery and management of georesources; and assist with geodata transfer. These reasons are influenced by the metadata standard adopted.

3.0 METADATA STANDARDS

In this paper a generic structure is used to compare all metadata standards. This structure is hierarchical and contains information describing georesources which we call *entries*. The generic structure consists of a set of entries that cannot be sub-divided further which we call *atomic entries*; and entries which can be sub-divided which we call *compound entries*. Metadata standards containing one level of compound entries followed by atomic entries in their hierarchy are called *Simple Structures* while metadata standards containing a mixture of compound and atomic entries beyond the first level of the hierarchy are called *Composite Structures*. Composite Structures may also contain atomic or compound entries used by other compound entries of the same metadata standard which makes the Composite Structures even more complex. Such entries we call *cyclic compound entries*.

For a given set of geodata sets the criteria for evaluating metadata standards against the generic metadata structure are:

Number of entries: If the number of entries for a given metadata standard is lower than a given threshold, then the metadata standard is *oversimplified*. If the number of entries is higher than a given threshold, then the metadata standard is *highly complex*. The determination of a suitable threshold is the mean number of entries for all of the metadata standards considered.

Extensibility of metadata standard: This refers to the ease with which users can define extensions to a given standard. Extensions to Simple Structures are readily accommodated because of their simplicity, however, extensions in Composite Structures are complex because their existing entries can have relationships to the new entries of the metadata standard which, depending upon the number of entries in the existing standard can be cumbersome to manage.

Depth of metadata standard hierarchy: This refers to the number of hierarchical levels for any given metadata standard. For Simple Structures the depth of the standard is by definition one while the hierarchy depth for Composite Structures is more than one.

Update of metadata entries: Simple Structures are usually easy to update because of their low hierarchical depth and in most instances requires less storage space when compared to Composite Structures which are difficult to update and usually require more storage space due to the high hierarchical depth.

Search of metadata standard: Search for entries in Simple Structures is faster because of the existence of the small hierarchical depth to traverse. When compared to Composite Structures the search for entries can be slower because of the larger hierarchical depth that define the complexity of the structure.

For the research described here, three metadata standards for describing geodata are chosen based upon the number of metadata elements they contain, and their recognition by the GIS community. These standards are:

ANZLIC metadata standard: ANZLIC is the inter-governmental council responsible for the coordination of land and geographic information in Australia and New Zealand (ANZLIC, 1996, pp:1). The ANZLIC standard is being used in the Australian Spatial Data Directory (ASDD)[2].

CSGM metadata standard: This is a Federal Geographic Data Committee (FGDC)[3] approved metadata standard supporting the collection and processing of geospatial metadata useable by all levels of government and in the private sector. The standard defines the availability, fitness, access, and transfer of geodata.

DC metadata standard: Intended to facilitate the discovery of electronic resources. The DC standard is used to describe the contents in museums, libraries, government agencies, and commercial organizations.

Entries conforming to these standards are discussed in the next 3 Sections.

3.1 ANZLIC standard

ANZLIC entries consists of: compound entries, called *Categories*, which contain groups of related ANZLIC metadata descriptions; and atomic entries, called *Elements*. The main compound entry in the ANZLIC standard is called *Dataset* and contains 10 compound entries and 32 atomic entries for each geodata set. A complete description of the ANZLIC core entry set is provided in ANZLIC (1996). The *Dataset* compound entry defines the root definition of the ANZLIC metadata standard and contains the following second level compound entries:

```
Dataset + Custodian + Description + Currency +
Status + Access + Quality + Contact Information +
Metadata Date + Additional Information
```

The distinction must be made that the second level compound entry *Dataset* is an entry which contains atomic entries while the root definition *Dataset* contains compound entries even though they both are called by the same name.

The ANZLIC standard is a Simple Structure therefore the ease of extending the ANZLIC standard is high. Hence, the ANZLIC standard is easy to update which is reflected by its a hierarchical depth of 2. Searching is relatively fast, however this depends upon the volume of geodata sets described.

3.2 CSDGM standard

The CSDGM standard is hierarchically organized into *sections*, *compound elements*, and *data elements* that define the information content for metadata to document a geodata set. *Sections* and *compound elements* in the CSDGM standard correspond to the compound entry from the generic structure, discussed in Section 3 while *data elements* in CSDGM correspond to the atomic entry in the generic structure.

The root of the CSDGM hierarchy is the main compound entry called *Metadata* which contains the following 10 compound entries:

```
Identification Information + Data Quality Information +
Spatial Data Organization Information + Spatial Reference
Information + Entity and Attribute Information + Distribution
```

---

[2] http://www.environment.gov.au/net/asdd/

[3] http://www.fgdc.gov/index.html

```
Information + Metadata Reference Information
```

These compound entries provides information about the higher level concepts, the types of values that can be provided for the entries, and information about the entries that are mandatory or repeatable (FGDC, 2000, pp:11). All compound entries are defined using either atomic entries or through intermediate compound entries.

A complete description of the CSDGM entry set is provided in (FGDC, 2000). Briefly, of the 10 CSDGM sections, there are 7 sections which define the *Metadata* compound entry, and an additional 3 supporting compound entries, (Citation Information, Time Period Information, Contact Information) which are used in a number of the 7 sections. These supporting sections represent the cyclic compound entries from the generic metadata structure discussed in Section 3. Within the 10 sections there is a total of 336 entries which are used to describe geodata sets.

The number of entries is high in the CSDGM standard, hence more storage space is needed. The extensibility of the CSDGM standard is difficult due to the existence of cyclic compound entries which imply that updating can be cumbersome. The largest depth of the hierarchy is 16 which occur in the *Identification Information* compound entry, which also contains a cyclic compound entry called *Citation Information*. Searching for metadata using the CSDGM structure is slower because of the large hierarchical depth to traverse.

3.3 DC standard

The DC standard consists of 15 compound entries called *elements*. Each compound entry is defined using a set of ten atomic entries called *attributes*. At the root of the hierarchy, the 15 compound entries consists of ten atomic entries, six of which are common to all DC compound entries, and the remaining four atomic entries *(Name, Identifier, Definition, Comment)* are different in their attribute values. The DC standard has a Simple Structure which is extensible, easily searched, and easily updated. The hierarchy depth is 2 and there are no cyclic compound entries. The root of the DC hierarchy is the main compound entry called *Resource* and contains the following compound entries:

```
Title + Creator + Subject + Description + Publisher +
Contributor + Date + Type + Format + Identifier +
Source + Language + Relation + Coverage + Rights
```

A complete description of the DC standard is provided in (DC, 1999).

3.4 Comparison between standards

Table 1 presents a summary of the entries and the depth of the trees associated with the three metadata standards. The computed threshold of 134 is used for categorizing the standards to be oversimplified or highly complex. Therefore, CSDGM is highly complex, while ANZLIC and DC are oversimplified.

When comparing the number of entries among each metadata standard, ANZLIC's approach is less ambitious in its level of descriptions than CSDGM's approach while DC's approach is the least ambitious of them all.

Metadata standard	Compound entries	Atomic entries	Total number of entries	Depth of tree
ANZLIC	10	32	42	2
CSDGM	119	217	336	16
DC	15	10	25	2

Table 1: *Comparison between metadata standards*

CSDGM is not easily extensible and updating can be cumbersome when compared to ANZLIC and DC standards. The depth of the CSDGM hierarchy is the largest and because of this, the search for metadata may take longer than searches using the ANZLIC and DC structures.

ANZLIC (1996, pp:4) stated that, `users need a level of detail, clarity and accuracy in the metadata sufficient for them to judge whether or not to make further inquiries of the contact organisation responsible for a data set'. In addition, maintaining a comprehensive directory imposes a significant burden on organizations that manage geodata sets. ANZLIC (1996) suggested that experience indicate a balance need to be struck between these two considerations which is achieved using ANZLIC.

DC is not a detailed standard like ANZLIC or CSDGM, however, DC is a standard which is being given further consideration as a geographic metadata standard, because of its simplicity and extensibility (Gibb, 1999). Traditionally DC was used in formal resource descriptions such as in library catalogs and in museums. To represent geographic metadata, the standard requires modification to describe geodata and at the same time relate the standard to other existing geodata metadata standards.

To query metadata stored in such organizations, a mechanism is needed to translate queries from one metadata standard to the other. A description of such a mechanism can be found in Ramroop & Pascoe's (1999a) research which proposes a Virtual Centre. Supporting the need in organizations to remain autonomous, Buehler & McKee (1998) expressed OGC's interest for further research into the use of the *Lightweight Directory Access Protocol (LDAP)* to catalog geographic metadata.

4.0 CONCEPTS IN LDAP CATALOGS

LDAP's design is intended to be a global directory service where metadata describing georesources in a directory is organized hierarchically (SDK, 1998). The root of an organization's hierarchy contain the name and location of the georesource organization, while the entries of the metadata standard forms the leaves of the hierarchy. LDAP directories consists of *entries* which is similar to the entry defined for the generic metadata structure discussed in Section 3.

LDAP entries are described using *attributes* which contains *values*. For example, an entry describing a geodata set called *Dataset* may have attributes and corresponding values as shown in Figure 1.

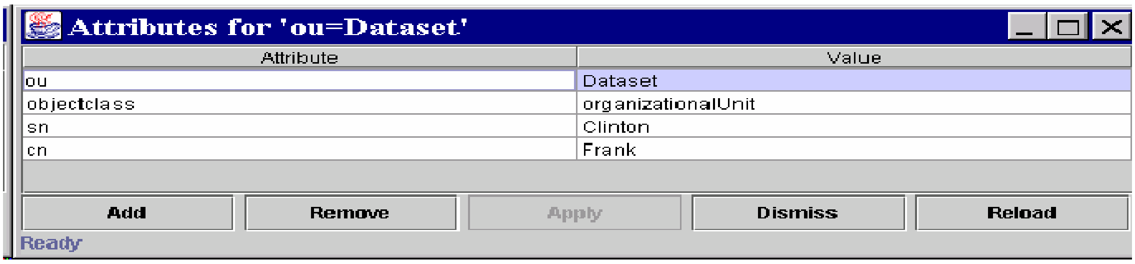

Figure 1: *Example of attributes and values for a single DIT entry*

An attribute can have more than one value. For example, a geodata set can have more than one value for its contact information such as phone numbers, or email addresses.

The Standalone LDAP Daemon (SLAPD) provides an implementation of LDAP[4]. There are specific attributes used in LDAP which are as follows:

- **domain component (dc)** is the LDAP attribute type from a Domain Nameserver System (DNS) such as X.500;
- **common name (cn)** is the name of the organization;
- **surname (sn)** is the surname of the owner/manager of the organization;
- **organizational unit (ou)** represents entries for smaller organizations;
- **unique identifier (uid)** is the identification associated with each entry in the tree;
- **organization (o)** represent the name of the organization.
  - In addition, the following is needed as input at the initial install of SLAPD servers at each organization:
  - **owner:** the person responsible for the organization from which the values of **cn** and **sn** are obtained;
  - **country (c):** indicate the name of the country the organization is located; and

port number: identify the computer port to be used by the SLAPD server for communication.

Netscape Directory Server 3.0 for Java[5] provides the directory service used to manage entries in LDAP directories. To perform this service the Java Naming Directory Interface (JNDI)[6] is used as the Graphical User Interface (GUI).

[4] http://www.aeinc.com/aeslapd/
[5] http://developer.netscape.com/
[6] http://java.sun.com/products/jndi/

Figure 2 shows an example of a hierarchy of entries in a LDAP DIT. The organization's root of the hierarchy is called **Root** while the other entries are **o= Cadastral** which contain branches to the other entries **ou= GIS** and **ou= Registry**. The **ou=GIS** entry is branched into **uid=Roads** and **uid=Soils**. This branching of the DIT creates the hierarchical character of LDAP trees. As each entry is added to the tree, the attributes of the lower root entries of the tree are inherited into the new entries of the tree.

Further sub-divisions from the root entry of the organization are entries representing the chosen metadata standard. These entries are represented by the definition of LDAP classes that adhere to the generic metadata standard. The entries are identified by a *distinguished name (dn)* which uniquely identifies entries at any position of the DIT and the path of names that trace the entries back to the root of the tree. For example, from Figure 2 the **Roads** entry has a dn entry of:

uid=Roads,ou=GIS,o=Cadastral

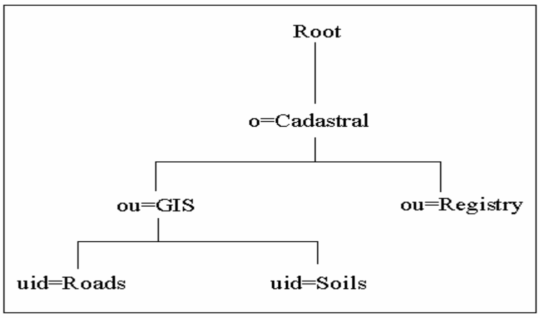

Figure 2: *Example of a LDAP hierarchy of entries*

To search for this entry, the LDAP client establishes a connection with the LDAP server by using a **TCP/IP** port number, then a simple method of authentication is executed before search operations are performed. The LDAP server utilizes the dn to locate an entry and to perform LDAP operations, discussed in the next section.

4.1 Operations of LDAP client/server

By using the LDAP client, the following operations are available which is applied to SLAPD DITs (SDK, 1998, pp:3): searching for and retrieving entries from catalogs; adding new entries to catalogs; updating entries in catalogs; deleting entries from catalogs; and renaming entries in catalogs.

An important feature of LDAP is the recursive capability called *referral*, where a LDAP server is referred to other LDAP servers. An LDAP server may be configured to send a referral to another LDAP server if the client request a *dn* with a suffix that is not in the server's directory tree. For example, if the server directory contains entries under *o=Cadastral* and the client request an entry under *o=Cadastral.com* then a referral is sent to another LDAP server. Figure 3 shows the steps taken during referral.

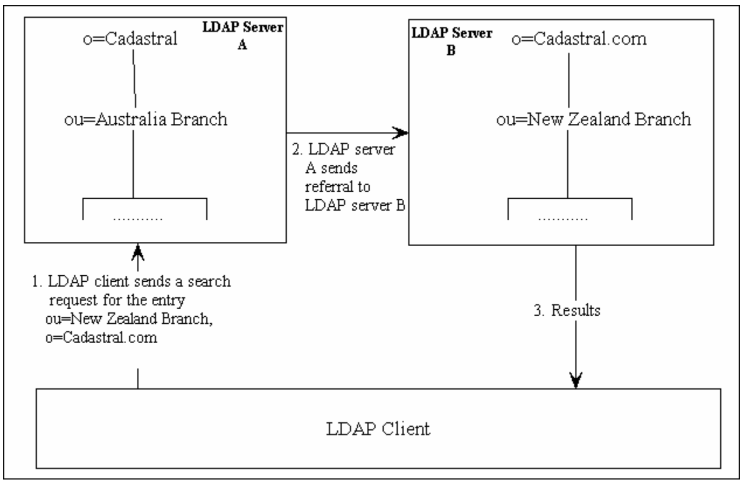

Figure 3: *LDAP referral*

The collaboration diagram in Figure 4 shows the exchange of messages, and objects within a SLAPD server which facilitates the search for metadata. From the Figure, the exchange of messages is achieved by executing the four methods as identified in the following steps:

1. A new **LDAPConnection** object connecting to known LDAP servers is created.
   Connection to the server involves the use of the **connect()** method of the **LDAPConnection** object. The IP address and the port number are used as the connection parameters.

2. Bind and authenticate with LDAP servers.
   An LDAP bind request is executed.

3. Perform the LDAP search operation on directories.
   Searching is achieved by using the LDAP **search()** method which require: definition of the starting point in the catalog, (or the base dn of the entry where to start searching); the scope of the search (identifies the level under the base dn on the DIT to perform the search); a search filter (specifies what to search for); the type of information to be returned; and the search constraints applied to the search.

4. All LDAP connections are closed using the **disconnect()** method.

To be able to use these methods which are accessible by users on the Internet, the *Common Request Broker Architecture (CORBA)* is adopted for reasons discussed in the next section.

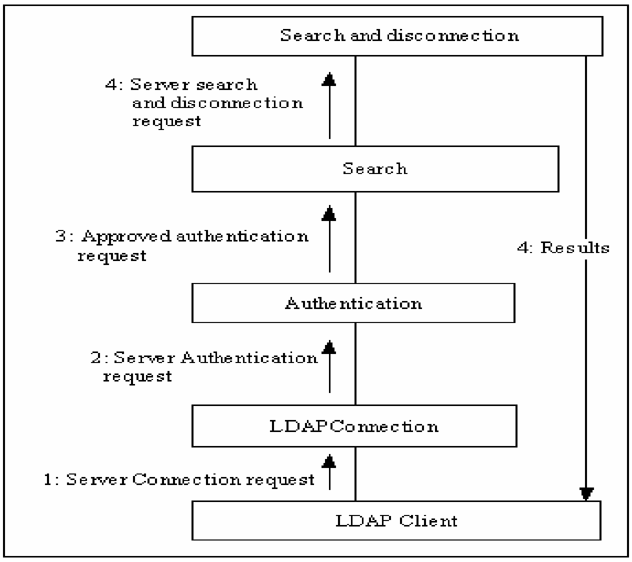

Figure 4: *LDAP collaboration diagram*

5.0 CORBA ARCHITECTURE

CORBA facilitates the interoperability between disparate technologies. The technologies supporting interoperability are: Java, C++, C, Smalltalk, XML, COM/DCOM which is facilitated using CORBA's *Interface Definition Language (IDL)*. `IDL-specified methods can be written in and invoked from any language that provides CORBA bindings' (Orfali & Harkey, 1998, pp:5). Figure 5 shows the bindings of a number of applications built using different computer languages to the CORBA Object Request Broker (ORB).

Buehler & McKee (1998) explained that from OGC's OpenGIS specification[7], CORBA facilitate independent distributed computing platforms. In this research CORBA is adopted because its architecture:

- is endorsed by the OGC's OpenGIS specification by using the object oriented approach;
- supports distributed systems;
- is independent of the computer language, operating system, platform, and vendor; and
- consists of a core implemented by various commercially available ORBs;

[7] http://www.opengis.org/technodev_hp.htm

Using the CORBA architecture and Netscape Directory Server 3.0 for Java, clients are able to access objects containing methods distributed throughout a network. These objects are object implementations which for example are instances of the LDAP directories. CORBA's IDLs facilitate the accessibility to these objects by being a purely declarative language and by allowing the implementation details to be filled with the necessary operations as discussed in Section 4.1. Clients are able to use the implementation details through the ORB which requests and receive responses from other objects locally or remotely.

`CORBA use objects as a unifying metaphor for bringing existing applications to the bus' (Orfali & Harkey, 1998, pp:3). The CORBA bus is used to coordinate the functions of the LDAP directory search service and the client. The CORBA architecture with LDAP object implementation contains a number of components which is shown in Figure 5.

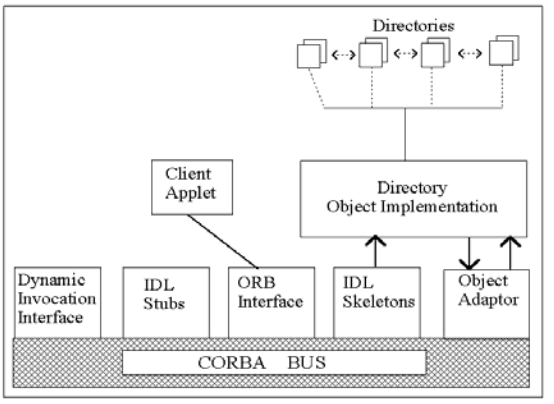

Figure 5: *CORBA Object Management Architecture interface with directories*

The role of the various components in the structure of CORBA object management architecture are as follows (Orfali & Harkey, 1998, pp:11-13):

Client applet - provide the user interface and invoke services on servers;

Dynamic Invocation Interface (DII) - allow the discovery of methods to be invoked using static invocations or at run time;

IDL stubs - provide the static interfaces to object services by encoding operations and parameters into a message that can be sent over the network;

ORB interface - consists of APIs with local services specific to applications;

IDL skeletons - provide a run time binding mechanism for servers that need to handle incoming method calls for components that do not have IDL based compiled skeletons (or stubs);

Object adaptor - utilize the ORB's communication services and accept requests for services on behalf of server's objects; and

Directories - provide the network of LDAP directories.

Figure 6 shows the client applet in action, where the object implementation refer to the instances of directories and the CORBA ORB is the transport medium used for accessing the methods from the LDAP objects.

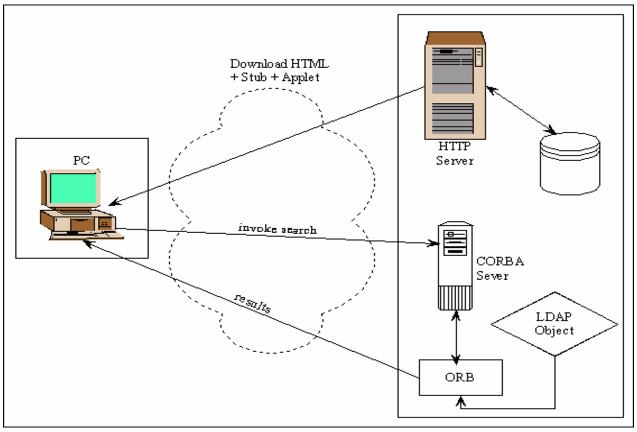

Figure 7: *Client applet in action*

6.0 CONCLUSION

Metadata is an important ingredient in the cataloging of the already existing large volume of geodata. This paper identify a high level of complexity of the CSDGM standard; and a level of simplicity for ANZLIC and DC standards. When comparing the three chosen metadata standards, ANZLIC and DC have a Simple Generic Structure which is extensible, and updating is less cumbersome than the CSDGM standard. The depth of the hierarchies and number of entries is the highest for the CSDGM standard which influences the speed with which search results are reported back to the client.

The use of LDAP to store spatial metadata is demonstrated. Benefits of using LDAP DITs are: referral among servers; DITs are hierarchical and reflects the structure of metadata standards; and dns are the unique ids for entries. The representation of metadata standard structures in LDAP is not constrained by the level of detail, since there are unlimited number of branches in the definition of a network of LDAP catalogs. This indicates that such a network is easily extensible. The use of CORBA, which is endorsed by OGC further extends such a network to include other disparate technologies.

REFERENCES

ANZLIC, ed. (1996), *Core Metadata Elements for Land and Geographic Data Directories in Australia and New Zealand*, Working Group on Metadata Australia New Zealand Land Information Council.

Buehler, K. & McKee, L. (1998), *The OpenGIS Guide. Introduction to Interoperable Geoprocessing and OpenGIS Specification*, third edition, OpenGIS.

DC (1999), *Dublin Core Metadata Element Set, Version 1.1: Reference Description.* http://purl.oclc.org/dc/documents/rec-dces-19990702.htm Accessed on 08/09/2000.

FGDC (1997), Geospatial data clearinghouse activity'. Federal Geographic Data Committee http://www.fgdc.gov/clearinghouse/

FGDC (1998), Content standard for digital geospatial metadata (CSDGM), *Federal Geographic Data Committee (FGDC)* National Spatial Data Infrastructure (NSDI). FGDC-STD-001-1998.

FGDC (2000), Content standard for digital geospatial metadata (CSDGM), *Federal Geographic Data Committee* http://www.fgdc.gov/metadata/contstan.html Accessed on 17/07/00.

Gibb, R. (1999), OIM metadata working party. http://www.massey.landcare.cri.nz/ rgg/mfe/oim/ Accessed on 27/09/99.

Medycky-Scott, D., Cuthbertson, M. & Newman, I. (1996), Discovering environmental data; metadatabases, network information resource tools and the GENIE system, *International Journal or Geographical Information Systems* 10(1), pp: 65-84.

Moellering, H. (1991), *Spatial Data Transfer Standards : Current International Status*, Elsevier Applied Science, London and New York.

Orfali, R. & Harkey, D. (1998), *Client/Server Programming with Java and CORBA*, 2nd edn, John Wiley and Sons Inc. ISBN 0-471-24578-X.

Ramroop, S. & Pascoe, R. (1999a), Implementation architecture for a national data center, *Interoperating Geographic Information Systems. Second International Conference, INTEROP'99, Zurich, Switzerland.* pp. 65-74. Lecture notes in Computer Science (1580), edited by A. Vckovski, K.E. Brassel and Hans-Jorg Schek. Published by Springer-Verlag Berlin Heidelberg.

Rossmeissl, H. J. (1998), Information access; where do we go from here?, *The 26th Annual Conference of AURISA (Australian Urban and Regional Information Systems Association, INC.), Perth, Western Australia*. Editor: AURISA.

SDK (1998), *Netscape Directory SDK 3.0 for Java Programmer's Guide*, version 3.0, Netscape Communications Corporation 1998. 501 East Middlefield Road, Mountain View, CA 94043.

Shah, K. & Sheth, A. (1998), Logical information modeling of web-accessible heterogeneous digital assets, *Proceedings of the IEEE Forum on Research and Technology Advances in Digital Libraries*, *IEEE ADL '98, April 22-24, 1998, Santa Barbara, California, USA. IEEE Computer Society* pp. 266-275. ISBN 0-8186-8464-X.