Meta Manager

White Paper

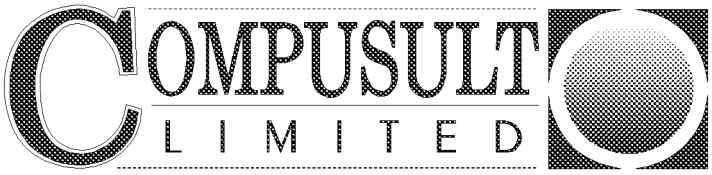
Revision 2.0

Prepared by:

Compusult Limited
40 Bannister Street
Mount Pearl, Newfoundland, Canada
A1N 3C9

Contact: Mr. Barry O'Rourke, President

Toll Free: (888) 307-7707
Tel: (709) 745-7914
Fax: (709) 745-7927
E-mail: management@compusult.nf.ca
URL: http://www.compusult.nf.ca



Document No. P901-005

August 23, 2000

Table of Contents

1.0 INTRODUCTION 1

2.0 METADATA 3
- 2.1 The Value of Metadata 3
- 2.2 What are Metadata? 3
- 2.3 Why Bother with Metadata? 3
- 2.4 How can Metadata be Produced? 4
- 2.5 Why Use a Standard? 4
- 2.6 What Standard Should be Used? 4
- 2.7 Why Use Metadata? 5

3.0 CLEARINGHOUSE 6
- 3.1 What is Clearinghouse? 6
- 3.2 Why Promote a Clearinghouse Activity? 6
- 3.3 Why Not Just Use Internet Indexes or Catalogs on CD-ROM? 7
- 3.4 Who Should Participate in Clearinghouse? 7
- 3.5 What are the Requirements for Being a Clearinghouse Provider and User? 8
- 3.6 What Data are Accessible Through Clearinghouse? 8
- 3.7 How Does Clearinghouse Work? 9

4.0 METADATA MANAGEMENT MADE EASY 10
- 4.1 Product Overview 10
  - 4.1.1 Meta Manager Capabilities and Features 11
  - 4.1.2 Meta Manager Components 14
  - 4.1.3 Systems and Network Infrastructure 15
- 4.2 Addressing the Problems 18

5.0 TWO EXAMPLES 20
- 5.1 Canadian Earth Observation Network (CEONet) 20
- 5.2 U.S. National Wetlands Inventory 21

6.0 WEB ENTERPRISE SUITE 23

1.0 INTRODUCTION

This white paper describes Meta Manager, a toolkit that provides metadata management and clearinghouse connectivity. Meta Manager functionality includes:

- the ability to use and work with different metadata standards (i.e. GILS, FGDC, ISO TC/211);
- the ability to define, map and manage metadata content and to present this content in a number of formats;
- processes and procedures that have been specifically developed for RDBMS metadata management and access;
- server components that enable RDBMSs to process Z39.50 search and retrieval requests (i.e., Clearinghouse connectivity);
- the ability to reverse engineer existing database schemas; and,
- integration into other Compusult Web Enterprise Suite components for providing end-to-end web-based solutions.

The Web Enterprise Suite contains Internet-based applications and toolkits that support web mapping, information cataloging, discovery, retrieval and delivery services. These applications provide the ability to integrate into existing catalog and enterprise management (including e-commerce) facilities.

For additional information on Meta Manager, please visit the Meta Manager website at **www.fgdctoolkit.com**. For additional information on Web Enterprise Suite, please visit the Web Enterprise Suite website at **www.webenterprisesuite.com.**

Meta Manager supplies the tools and linkages that enable organizations to cost-effectively enter into the next generation of Internet searching and discovery that is based on the Z39.50 search and retrieval protocol.

Meta Manager's features include:

- out-of-the-box, Standards-based Commercial Off-The-Shelf (SCOTS) operation and management;
- support for many metadata standards;
- multi-tiered architecture supporting deployment in diverse secure networked environments;
- metadata mapping and inclusion of fields that are not included in the target. database;
- support for direct integration into existing Oracle, ESRI SDE, Microsoft Access and Microsoft SQL Server inventory and metadata databases without requiring modifications or changes to the data structures;

- a comprehensive install script that includes options for installing and configuring the server software; populating relational databases and installing the Meta Manager Administrative interface;
- operation on UNIX and Windows 95/98/NT/2000 platforms;
- ODBC driver support providing connectivity to many relational database engines;
- the Percipio PHTML engine for fast SQL-access querying and retrieval;
- metadata mapping and inclusion of fields that are not included in the target database;
- keyword look-up for selecting and identifying keywords associated with collections in the target database;
- a master collection definition facility for defining metadata fields, functions and defaults that can be applied to all collections of information available for searching and retrieval;
- the flexibility to render metadata using ASCII, HTML, SGML and XML display formats;
- enhanced SQL syntax checking for maintaining the integrity of the targeted databases;
- creation and population of metadata databases by selecting the metadata standard;
- customized result sets for presentation of extended metadata databases;
- enhanced interfaces with GIS Tools (e.g., ESRI);
- metadata loading interfaces;
- increased functionality for the Global Locator Information Service (GILS);
- enhanced text searching capability;
- customizable Clearinghouse client that supports Clearinghouse search and server registration; and,
- Internet and FGDC Clearinghouse connectivity server software.

2.0 METADATA

The following description of metadata was originally obtained from the Federal Geographic Data Committee (FGDC) WWW site, **http://www.fgdc.gov/Communications/Metadata/Metabroc.html**.

2.1 The Value of Metadata

Two very similar paintings of circus performers by Picasso from 1904 are put on the auction block; one brings tens of millions of dollars, the other hundreds of thousands. What is the difference? In one case, the ownership of the painting can be traced through sales slips and auction house records back to the estate of Picasso's dealer. The other painting appeared suddenly on the art market. It looks almost identical, but lacking documentation, how can one be sure it's authentic?

Just as a work of art can change hands many times, so can geospatial data. Once created, data can travel almost instantaneously through a network and be used for any number of different kinds of spatial analysis. Thus transformed, these data can be retransmitted to another user. Change is the essence of geospatial data in a networked environment. The word metadata shares the same Greek root as the word metamorphosis. Meta means change and metadata, or "data about data" describe the origins of and track the changes to geospatial data.

Metadata can help the city planner, the graduate student in geography or the forest manager find and use geospatial data, but they also benefit the primary creator of the data by maintaining the value of the data and assuring their continued use over a span of years.

2.2 What are Metadata?

The concept of metadata is familiar to most people who deal with spatial issues. A map legend is pure metadata. The legend contains information about the publisher of the map, the publication date, the type of map, a description of the map, spatial references, the map's scale and its accuracy, among many other things. Metadata are simply that type of descriptive information applied to a digital geospatial file. They're a common set of terms and definitions to use when documenting geospatial data. Most digital geospatial files now have some associated metadata.

2.3 Why Bother with Metadata?

Metadata helps people who use geospatial data find the data they need and determine how best to use it. Metadata benefit the data producing organization as well. As personnel change in an organization, undocumented data may lose their value. Later workers may have little understanding of the contents and uses for a digital data base and may find they can't trust results

generated from these data. Lack of knowledge about other organizations' data can lead to duplication of effort. It may seem burdensome to add the cost of generating metadata to the cost of data collection, but in the long run it's worth it.

2.4 How can Metadata be Produced?

The information needed to create metadata is often readily available when the data are collected. A small amount of time invested at the beginning of a project may save money in the future. Data producers and users cannot afford to be without documented data. The initial expense of documenting data clearly outweighs the potential costs of duplicated or redundant data generation. A recently developed metadata standard provides a systematic way to collect metadata.

2.5 Why Use a Standard?

When producing a map, the cartographer must organize all the descriptive information that goes into the map legend in a particular format. Titles are put in a specific place, tic marks are made a certain way, meters may be used instead of feet, and so forth. A metadata standard is simply a common set of terms and definitions that describe geospatial data.

2.6 What Standard Should be Used?

The Federal Geographic Data Committee (FGDC) recently adopted a content standard for metadata. According to an Executive order signed by President Clinton on April 11, 1994, all Federal agencies will begin to use this standard to document newly created geospatial data as of January, 1995. This standard provides a consistent approach and format for the description of data characteristics. The standard was developed over a two-year period, with extensive review by professionals at all levels of government. The standard provides a way for data users to know:

- what data are available
- whether the data meet their specific needs
- where to find the data
- how to access the data.

Because these standards are now in place, and large amounts of Federal data will be available in these standards, data managers from State and local governments and private industry will have an incentive to adopt these standards to document their own data. The FGDC is also sponsoring the creation of a National Geospatial Data Clearinghouse which will point users toward the spatial data that are best for their particular project. The intent is not to centralize all geographic data in one location, but to provide links through the Internet to distributed sites where data are produced or

maintained. Managers who document their data using the metadata standards will provide these metadata to the National Geospatial Data Clearinghouse so that users can easily find data. Easier access to data will mean that a company's customers or an agency's cooperators could be increased.

2.7 Why Use Metadata?

Thirty-one years ago, humans landed on the Moon. Data from that era are still being used today, and it is reasonable to assume that today's geospatial data could still be used in the year 2020 and beyond to study climate change, ecosystems, and other natural processes. Metadata standards will increase the value of such data by facilitating data sharing through time and space.

The value of Picasso's painting did not depend solely on his having signed the work, a signature that could easily have been forged. Information about the painting, where it came from and where it had been, increased its value. So when a manager launches a new project, investing a small amount of time and resources at the beginning will pay dividends in the future.

3.0 CLEARINGHOUSE

The following definition of a clearinghouse was obtained from the Federal Geographic Data Committee (FGDC) WWW site, **http://www.fgdc.gov/Clearinghouse/background.html**.

3.1 What is Clearinghouse?

The Clearinghouse Activity, sponsored by the FGDC, is a decentralized system of servers located on the Internet which contain field-level descriptions of available digital spatial data. This descriptive information, known as metadata, are collected in a standard format to facilitate query and consistent presentation across multiple participating sites.

Clearinghouse uses readily available Web technology for the client side and uses the ANSI standard Z39.50 for the query, search, and presentation of search results to the Web client.

A fundamental goal of Clearinghouse is to provide access to digital spatial data through metadata. The Clearinghouse functions as a detailed catalog service with support for links to spatial data and browse graphics. Clearinghouse sites are encouraged to provide hypertext linkages within their metadata entries that enable users to directly download the digital data set in one or more formats. Where digital data are too large to be made available through the Internet or the data products are made available for sale, linkage to an order form can be provided in lieu of a data set. Through this model, Clearinghouse metadata provides low-cost advertising for providers of spatial data, both non-commercial and commercial, to potential customers via the Internet.

Clearinghouse allows individual agencies, consortia, or geographically-defined communities to band together and promote their available digital spatial data. Servers may be installed at local, regional, or central offices, dictated by the organizational and logistical efficiencies of each organization. All Clearinghouse servers are considered “peers” within the Clearinghouse activity -- there is no hierarchy among the servers -- permitting direct query by any user on the Internet with minimum transactional processing.

3.2 Why Promote a Clearinghouse Activity?

The development of a Clearinghouse among U.S. Federal agencies was motivated by a desire to minimize duplication of effort in the collection of expensive digital spatial data and foster cooperative digital data collection activities. By promoting the availability, quality, and requirements for digital data through a searchable on-line system a Clearinghouse facility would greatly assist in coordination of data collection and research activities. Clearinghouse also provides a primary data dissemination mechanism to traditional and non-traditional spatial data users.

For those in need of a mandate, federal participation in the Clearinghouse is directed by Executive Order 12906 through its official creation of the National Spatial Data Infrastructure. Compliance with this Order and its notions of data sharing has gained Cabinet-level interest.

3.3 Why Not Just Use Internet Indexes or Catalogs on CD-ROM?

Digital spatial data and metadata are stored in many forms and systems which make their discovery on the Internet difficult. Indexing of text-based catalogs of available information only on CD-ROM provides a limited and time-dependent view of the status of predominantly network-based resources. Use of current web indexing technology offers literal text search and matching for metadata which happen to be stored in HTML, but do not generally provide the indexing required for search of coordinates, dates and times, and other numeric values. In addition, an increasing amount of information is being stored within dynamic databases behind Web servers, further frustrating and making questionable the value of, external indexes. Clearinghouse provides standard methods for spatial data discovery that do not invalidate existing systems yet provide a federated search capability among distributed sites.

CD-ROM and network-based search capabilities provide complementary services to that user community interested in finding and accessing digital spatial data. Synopses of existing metadata entries from the distributed Clearinghouse can periodically be captured to CD-ROM for off-line browsing and review. The on-line resources they reference, however, will become accessible only in an on-line setting, restricting the utility of a standalone CD product. The general trend toward connectivity of spatial data producers, vendors, and users on the Internet coupled with the development of the Open Geodata Interoperability Specification indicate a long-term public commitment to not only on-line data discovery but direct data access by client processes across internal and public networks.

Clearinghouse provides one testbed solution to catalog interoperability on the Internet today.

3.4 Who Should Participate in Clearinghouse?

Although initially targeted at federal agencies, participation in Clearinghouse prototypes has included federal, state, university and vendor participants in the United States and abroad. Through and international development grant, the U.S. Geological Survey has assisted in the development and activation of Clearinghouse services in Brazil and Costa Rica, providing access to current digital environmental data of global interest. The Australian Environmental Resource Information Network also implements a searchable Clearinghouse node for federated holdings within the country.

The role of the FGDC in Clearinghouse is to develop prototype software, provide reference implementations, facilitate discussions among Clearinghouse participants, develop and present

training materials, and operate a registry service of conforming spatial data servers. It is not the intent of the FGDC to create a centralized data system or index but to facilitate distributed search against multiple, current stores of spatial metadata and data on the Internet.

3.5 What are the Requirements for Being a Clearinghouse Provider and User?

Implementors must have access to multi-user computers (UNIX or Windows-NT) upon which the server software, interfaces and metadata collections are stored. Server sites are connected to the Internet via dedicated, high-speed data connection of 56KB or greater. It is recommended that Clearinghouse servers be located coincident with spatial data collections to encourage synchronization between the spatial data and the metadata, or descriptions, being served.

Organizations not yet connected to the Internet, or who have firewall or security restrictions on being directly connected, may elect to contract with an existing Internet Service Provider or partner with a local Clearinghouse node in a different organization to provide an off-site host computer for Clearinghouse.

Prospective users of Clearinghouse must have access to a current graphical Web browser on a personal computer or workstation. Client interfaces have been developed in HTML 2.0 and with Java extensions though Java is not required for basic access to Clearinghouse. User connections are anticipated through either local area network dedicated access to the Internet or via low-speed, dial-up modem, available throughout the United States at approximately $20 or less per month.

3.6 What Data are Accessible Through Clearinghouse?

A “digital geospatial data set” is the target unit of description defined within the Clearinghouse activity. The definition of a data set can be adjusted to meet a given agency’s requirements but it generally corresponds to the smallest identifiable data product (e.g., file) for which metadata are customarily collected. This may equate to a specific satellite image or vector data set that is managed by a data producer or distributor. Collections of data sets (e.g., flight lines, satellite “paths”, map or data series) may also have generalized metadata that could be inherited by individual data sets.

Accordingly, some data producers require that metadata be maintained on individual geographic features (road) or spatial primitive (line). A continuum of metadata may exist and be maintained by an organization, but the target requirement for data discovery in Clearinghouse remains at the data set level.

3.7 How Does Clearinghouse Work?

To provide search interoperability among different servers of geospatial metadata, the search and retrieve protocol known as ANSI Z39.50-1995 (ISO 10163-1995) was selected by the FGDC Clearinghouse activity. The Z39.50 protocol includes client and server software that establish a connection, pass a formatted query, return query results, and present identified documents to the client in one of several formats. The Z39.50 protocol was initially developed by the library community to discover bibliographic records using a standard set of attributes that would allow any Z39.50 client to present information from different yet similarly-structured servers. On the host (server) computer, Z39.50 server software typically communicates with an appropriate search engine (data base or indexing software) to process the query and formulate the results. In this way, the Z39.50 protocol can provide an alternative access method to existing geospatial data bases or metadata collections without requiring redesign of existing data systems through use of a single, standards-based protocol.

Z39.50 maintenance agencies have the ability to register specific sets of attributes, operators and rules of implementation as Application Profiles. Once adopted, these profiles are available to the implementor community for incorporation into existing client and server software. The FGDC has developed a profile for geospatial metadata, called “GEO,” which provides guidance on how to implement FGDC metadata elements within a Z39.50 service.

There can be a high degree of collaboration and re-use of data elements among the existing profiles. Elements from the base bib-1 attribute set are used in the Government Information Locator Service (GILS) profile, which, in turn, constitute a subset of the GEO profile. The Committee on Earth Observing Satellites (CEOS) has developed a Catalog Interoperability Profile (CIP) which includes the GEO attributes as well as those used by other spatial catalog systems.

4.0 METADATA MANAGEMENT MADE EASY

No one will argue the value of metadata. The clearinghouse infrastructure is well defined and documented and provides a framework for users and maintainers of metadata to publish, access and deliver metadata. Despite this framework, adoption of these standards and methods have been slow. Some of the problems associated with adopting the methodology are described below:

1. Much of the metadata created by organizations is stored in formats not readily accessible by clearinghouse methods.
2. In many cases, metadata was not created according to the definitions and formats identified by the standards.
3. Many metadata sets that have been created are stored in statically formatted HTML documents that require manual updates when data about the data changes.
4. Metadata that has been created according to an organization's internal methods cannot be altered because other software programs and applications expect the data to be in a pre-described format.
5. The learning curve for understanding the metadata content definitions and clearinghouse methods is quite often, for many organizations, prohibitive.
6. Many of the tools associated with implementing a metadata clearinghouse node are shareware lacking in ease of installation and customer support, and often result in significant resource allocation for implementation.
7. The tools associated with creating and maintaining a clearinghouse node are obtained from different sources which require additional tasks associated with configuring client and server applications.

The Meta Manager software suite was specifically created to address these problems. The system provides a turn-key solution that is fully supported and can be configured to serve users, creators and maintainers of metadata both large and small. The following subsections provide an overview of the product and address the problems identified above.

4.1 Product Overview

Meta Manager allows you to easily enter into the world of information marketing, selection, presentation and delivery using existing and creating new data dictionary definitions. Meta Manager supplies the linkages that enable your organization to cost-effectively enter into the next generation of Internet searching and discovery tools that are based on the Z39.50 search and retrieval protocol.

Meta Manager provides a complete solution for creating and defining new and/or mapping existing geospatial metadata definitions and geospatial datasets according to metadata standards such as the

U.S. Federal Geographic Data Committee's (FGDC) Content Standard for Digital Geospatial Metadata and the Global Information Locator System (GILS) profile. Both of these profiles are based on ANSI/NISO Z39.50, the American National Standard for Information Retrieval Application Service.

The ANSI Z39.50-1992 standard is being widely used within the library community for catalog and document search and retrieval. The Z39.50 standard provides for the use of common attribute sets whose use and operations are well-known to both client and server. The latest version of the standard also allows a server to "explain" its searchable attributes and operators to a client to permit an intelligent query of non-common attributes.

The GEO and GILS profiles have been defined to incorporate data elements including bounding coordinate and footprint fields.

Meta Manager provides for the search and retrieval of geospatial metadata entries and related geospatial data sets accessible by GEO-compliant clients in the Internet environment. This includes local applications of isolated TCP/IP networks, local area networks based on the Ethernet protocol, and wide-area networks using a consistent host addressing scheme.

In the Internet environment, the Server component of the software suite acts primarily as a direct access point to information stored within relational metadata databases. These databases can then be accessed through a GEO or GILS-compliant client without changing the internal dictionaries and definitions of the target database.

4.1.1 Meta Manager Capabilities and Features

The Meta Manager software can interface with target information that is managed by Relational Database Management Systems (RDBMS) which contain the spatial data and attribute information. Both client and server applications utilize ODBC drivers that allow connectivity to the database. These include:

- Oracle;
- Microsoft Access; and,
- Microsoft SQL Server.

In essence, the Meta Manager software enables suppliers with holdings maintained within relational databases to become GEO Servers with direct interoperability to other servers and clients on the Internet.

Once connected to a GEO or GILS server, users supported by appropriate clients that understand the relevant profile may navigate through the server. The Meta Manager servers support searching (i.e., accept a search query and return a search response or diagnostic messages) and support

browsing (i.e., accept a well-known search query and return a list of Entries in brief display format).

Primary features of Meta Manager include:

- Support for direct integration into existing Oracle, ESRI SDE, Microsoft Access and Microsoft SQL Server inventory and metadata databases without requiring modifications or changes to the data structures.
- A comprehensive install script that includes options for installing and configuring the YAZ-Z39.50 server software; populating relational databases and installing the Meta Manager Administrative interface.
- Operation on UNIX and Windows 95/98/NT/2000 platforms.
- Meta Manager LE (Lite Edition) has been created to provide suppliers with limited budgets and resources an entry into the clearinghouse model. The Meta Manager LE product is limited by the number of records that can be included in the product and will only connect to Microsoft Access databases within a Windows NT environment. Data providers with limited resources are able to implement a searchable small inventory collection online, without the need for them to set up an existing database schema in MS Access. This is accomplished using a Meta Manager Wizard facility.
- ODBC driver support providing connectivity to many relational database engines.
- The Percipio PHTML engine for fast SQL-access querying and retrieval.
- Metadata mapping and inclusion of fields that are not included in the target database.
- Keyword look-up for selecting and identifying keywords associated with collections in the target database.
- A master collection definition facility for defining metadata fields, functions and defaults that can be applied to all collections of information available for searching and retrieval. Changing the master collection will cause all related global variables to be updated automatically.
- Provides the flexibility to render metadata entries using ASCII, HTML, SGML and XML display formats.
- Enhanced SQL syntax checking for maintaining the integrity of the targeted databases.
- Creation and population of metadata databases by selecting the metadata standard.
- Internet and FGDC Clearinghouse connectivity server software.

- Enhanced Support for GRS-1 and USMARC. Meta Manager GEO servers may transfer records in up to three record syntaxes:

    - Simple Unstructured Text Record Syntax (SUTRS) -- REQUIRED

    In SUTRS, the formatting of the record contents is handled by the server and the client receives a record which is defined by the user to be in verbose text with long metadata tag names (the default presentation used by the FGDC metadata standard), SGML and HTML. Both SGML and HTML are formatted ASCII text documents which make them amenable to distribution via SUTRS.

    Client-side applications can be used to read the markup in the retrieved document and display it accordingly using SGML or HTML. Most World-Wide Web browsers can display HTML-formatted files and some can launch an SGML document viewer when an SGML mime-type is encountered.

    Negotiation between the client and the server can dictate which format is generated by the server as the user-preferred format for the downloaded document.

    - Generic Record Syntax (GRS-1) -- OPTIONAL

    Alternatively, GRS-1 can be used to encapsulate more complex bundles of information including images, related documents, video, or other ancillary information for further processing by the client. GRS could also be used to encapsulate the SGML along with the reference to -- or even inclusion of -- a DTD.

    - USMARC -- OPTIONAL

    MARC is an acronym for MAchine-Readable Cataloging. The standards for the representation and exchange of bibliographic, authority, classification, holdings, and community information data in machine-readable form in the United States are the five USMARC communication formats: Bibliographic Data, Authority Data, Classification Data, Holdings Data, and Community Information.

    Each separate MARC format provides detailed field descriptions and guidelines for applying the defined content designation and identifies conventions to be used to ensure input consistency.

    The server requirement to support the USMARC record syntax is concerned with communicating the record in USMARC record syntax on request from the client.

4.1.2 Meta Manager Components

Meta Manager is comprised of the following four components:

- **Meta Manager Installation Programs** - Automated installation procedures for installing toolkit components on targeted platforms. The installation programs prompt users for configuration information regarding the targeted platform. Once the information has been entered and verified, the program will proceed to perform the installation of all components.

- **Meta Manager Administrative Interface** - The administrative interface is available for UNIX and Windows 95/98/NT platforms and provides data administrators with the ability to define, map and manage relational metadata content. This utility contains all of the facilities necessary for mapping an SQL database schema to a desired standard (e.g., FGDC GEO Profile, GILS, ISO TC/211, etc.).

- **Meta Manager Server** - A Z39.50 Server for connecting relational databases to clearinghouses. The server side is mapped onto the interface of the specific database management system. The communication taking place between the server and the client application is precisely defined within the Z39.50 protocol.

- **Meta Manager Clearinghouse** - An application that provides organizations with the ability to create their own Clearinghouse. The application includes a Z39.50 client that provides intuitive search specification fields, a clearinghouse server registration module and clearinghouse server searching. The client look and feel and operation is fully customizable and supports XML processing.

In addition, HTML loading scripts are freely available for loading existing static HTML metadata documents into a relational model.

Meta Manager may also be connected to the Web Enterprise Suite component Relational Gateway for secure firewall connectivity.

The interface allows users to configure their existing collections within a one-to-two day time frame.

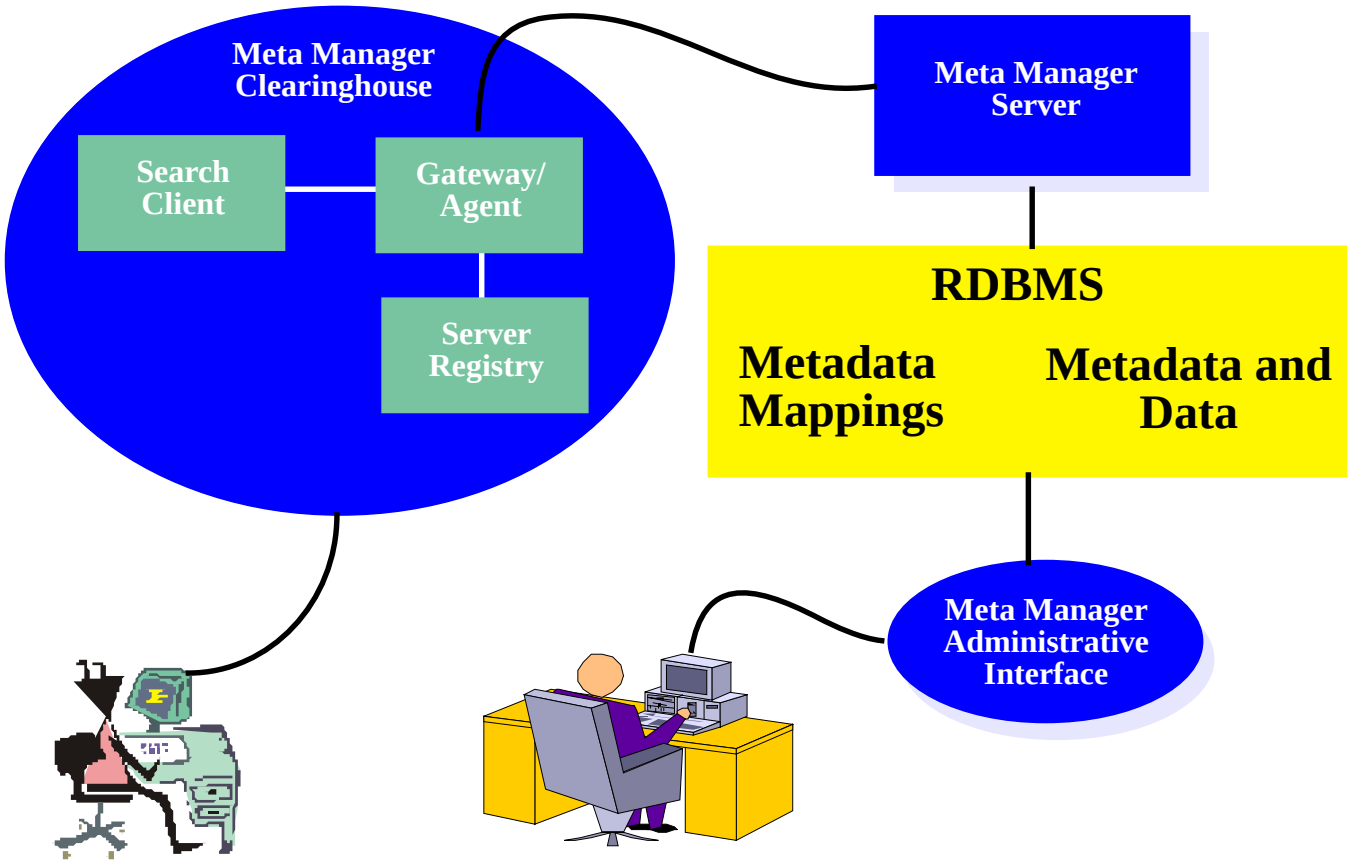


Figure 4-1: Meta Manager Components

4.1.3 Systems and Network Infrastructure

Systems and network infrastructure and security requirements will vary, depending upon the organization using Meta Manager. Meta Manager's multi-tiered architecture provides system administrators with a number of configuration methods for implementing clearinghouse connectivity according to network and security requirements.

Organizations having firewall or security restrictions on Internet connections may choose to select one of the following options:

1. Allow restricted Z39.50 traffic through the firewall to the Meta Manager server.

2. Position a Meta Manager server outside the firewall with a server that contains an un-secure relational database. This database will be updated by the secure database server on a scheduled basis with un-secure metadata content.
3. Position a Meta Manager server outside the firewall and allow SQLNet/ ODBC access through the firewall.
4. Position a Meta Manager server outside the firewall and use the Web Enterprise Suite component Relational Gateway to provide secure http/https access.

Relational Gateway is a middle-tier product that provides Web applications with the ability to connect to relational databases that are maintained inside firewalls.

The product has add-on functionality to support ESRI's Spatial Database Engine (SDE) and provides an interface between SDE-enabled databases and the client-side Web Enterprise Suite components. SDE-enabled databases may be accessed locally or remotely.

The Relational Gateway provides for firewall connectivity and supports a tiered approach to optimum usage of database licenses.

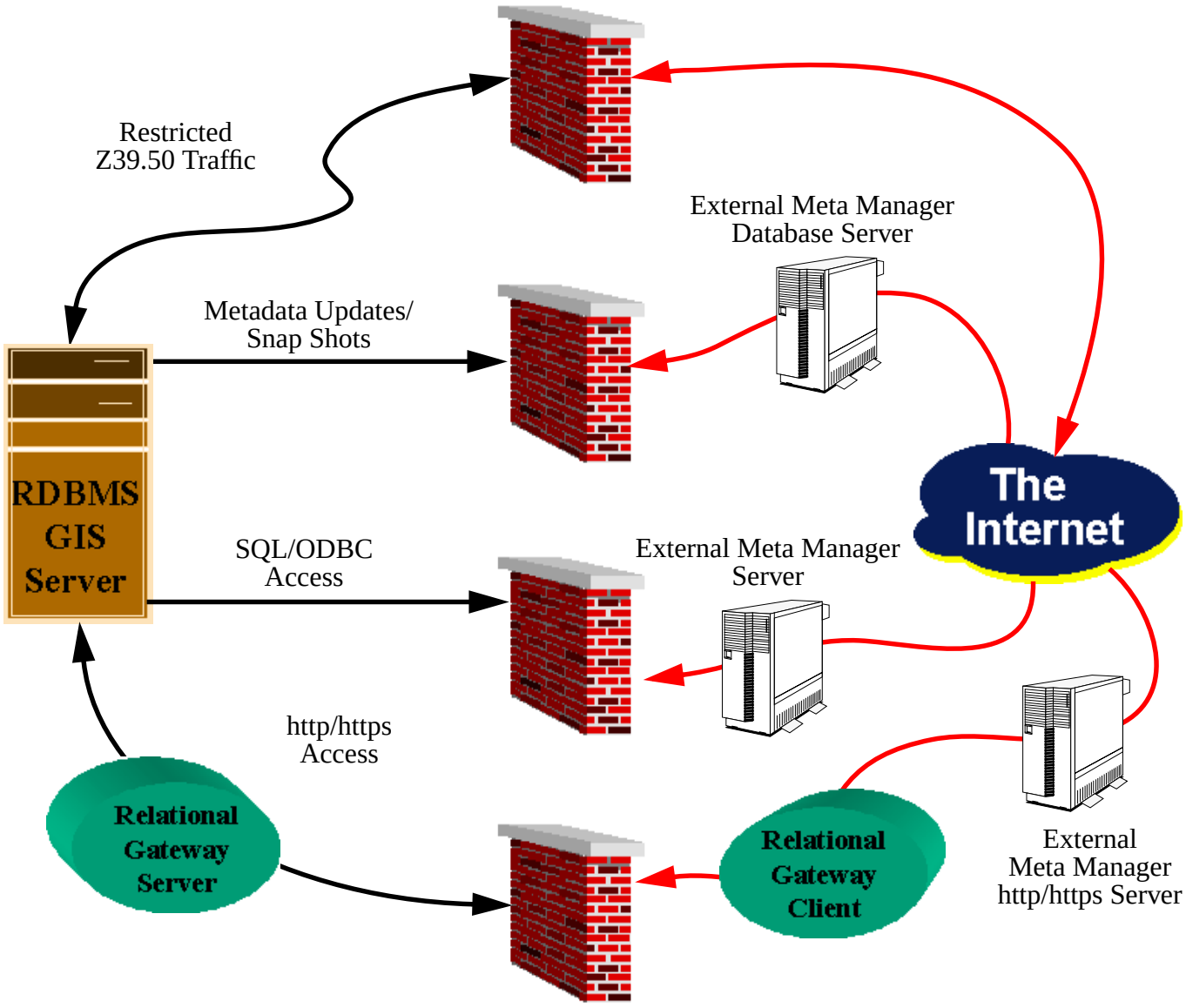


Figure 2: Z39.50 Database Access

Prospective Meta Manager users must have access to a current Z39.50 client on a personal computer or workstation. Alternatively, prospective users might use a WWW-to-Z39.50 gateway such as the one provided by the FGDC Clearinghouse or the Canadian Earth Observation Network (CEONet).

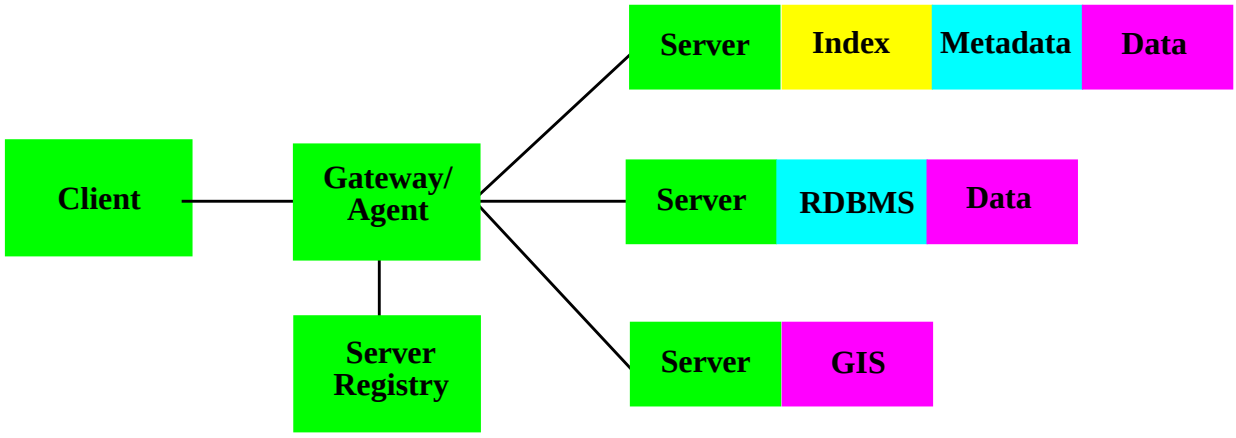


Figure 4-3: GEO Profile-Enabled Clearinghouse

4.2 Addressing the Problems

1. Much of the metadata created by organizations is stored in formats not readily accessible by clearinghouse methods.

Until the introduction of Meta Manager, users storing their metadata in formats not supported by Z39.50 searching engines were required to format the data statically and then index the data by an appropriate Z39.50 search engine. This was a time-consuming process and required “hands-on” knowledge of the Z39.50 tools and applications.

Meta Manager has overcome this restriction by allowing users to utilize their existing relational data definitions. A reverse engineering method allows users to use their existing data definitions. If users store their data in other formats, the tool allows the creation of a metadata template for loading the data into a relational database using readily-available SQL loaders. This process also ensures the metadata and data are synchronized.

2. In many cases, metadata was not created according to the definitions and formats identified by the standards.

Many organizations define metadata according to terminology adopted internally. For example, a date field may have the field name RELEVANT DATE and be stored in a date format that is different than the format required by the standard. Meta Manager overcomes this limitation by allowing users to “map” their metadata into the standard. Users can also choose to apply converters to display their internal data according to the required format.

As well, if users do not maintain metadata according to the mandatory or supplemental fields identified by the standard, Meta Manager provides users with the ability to “fill-in-the-gaps” by using the administrative interface to enter values that can apply to a single collection of metadata or to all collections of metadata across the organization.

3. Many metadata sets that have been created are stored in statically formatted HTML documents that require manual updates when data about the data changes.

Maintaining the metadata in a relational database enables users to easily define and maintain metadata sets and collections dynamically.

4. Metadata that has been created according to an organization’s internal methods cannot be altered because other software programs and applications expect the data to be in a pre-described format.

The Meta Manager mapping facilities allows users to maintain their metadata definitions in formats maintained by the organization. Meta Manager also directly supports ESRI’s Spatial Database Engine (SDE) allowing direct access to geographic searching and attribute definitions with future enhancements for ARC/INFO.

5. The learning curve for understanding the metadata content definitions and clearinghouse methods is quite often, for many organizations, prohibitive.

The installation and use of Meta Manager does not require users to understand the methods and practices of using Internet-based Z39.50 terminology and technology. The installation process easily installs the server-side application that is pre-configured to operate with databases and national and international clearinghouses. All of these interactions occur transparently to the user.

6. Many of the tools associated with implementing a metadata clearinghouse node are shareware lacking in ease of installation and customer support and often result in significant resource allocation for implementation.

Meta Manager is a fully-supported commercial application. It has been endorsed by the FGDC Clearinghouse and Canadian Earth Observation Network (CEONet) initiatives.

7. The tools associated with creating and maintaining a clearinghouse node are obtained from different sources requiring additional tasks associated with configuring client and server applications.

Meta Manager is a complete end-to-end turnkey solution. It contains all of the tools and applications required for metadata management and publishing within Clearinghouse initiatives.

5.0 TWO EXAMPLES

5.1 Canadian Earth Observation Network (CEONet)

The Canadian Earth Observation Network (CEONet) is a comprehensive clearinghouse for suppliers of geospatial data and other related services.

CEONet is an initiative by the Canadian government to create a national infrastructure for providing access to earth observation archives and other complementary spatial databases. This initiative is being driven by the requirements of Canadian users for better access to earth observation data, and by the opportunities for Canadian industry afforded by the rapid growth of the international market for earth observation data, services, and network systems.

The objectives of the Canadian Earth Observation Network (CEONet) are to:

- improve access to Canadian earth observation and geomatics information products and services by national and international clients;
- provide a forum to advertise, and an avenue to distribute, Canadian value-added earth observation and geomatics information products and services; and,
- develop a Canadian industrial capability to exploit the growing international market for earth observation and geomatics data management and distribution systems.

Meta Manager has been designed as a primary toolkit for allowing suppliers to connect spatial databases to CEONet using clearinghouse methods and technology. One of the primary inventories searchable by CEONet is the Canadian Earth Observation Catalog (CEOCat).

CEOCat is a Compusult Percipio application. The operational environment of the CEOCat system includes individuals and organizations from sectors such as education, government and private industry, that require access to remote sensing data. The CEOCat System provides:

- a powerful and easy-to-use WWW search interface that allows users to quickly locate Raw and/or Processed Remote Sensing data cataloged by Canada Centre for Remote Sensing (CCRS);
- a browse interface to allow users to view browse images and metadata associated with search results to determine the appropriateness of the results;
- a simulation of on-line ordering and delivery of full-resolution data made available by CCRS; and,
- ordering information for datasets that are only cataloged by CCRS.

Users access CEOCat through a WWW interface. All users are able to browse the catalog to identify products they wish to order. CEOCat is an autonomous system that provides access to the CCRS archive independently of CEONet.

Metadata regarding Canada's satellite remote sensing archive is stored in an Oracle database. Browse images for the raw remote sensing data are stored on a network-mountable UNIX file system. The metadata descriptions and database configuration was created using remote sensing terminology. In addition, a complex satellite look-up data model was incorporated to allow users to perform cross-satellite/sensor queries and browsing.

The Meta Manager application was installed and configured at CCRS allowing the CEOCat holdings to become clearinghouse-enabled and FGDC-compliant in under 2 days. The server-side software was installed on a SUN Solaris platform and provides gateway access to the system. The Meta Manager administrative interface is operational on both UNIX platforms and Windows 95/98/NT/2000 platforms and provided CEOCat data administrators with the ability to map internal definitions to the FGDC standard. No modifications to the internal storage formats of the metadata were performed.

The application was also used to fill in missing FGDC metadata definitions that were required to make the datasets FGDC-compliant.

5.2 U.S. National Wetlands Inventory

The National Wetlands Inventory (NWI) of the U.S. Fish and Wildlife Service produces information on the characteristics, extent and status of the Nation's wetlands and deepwater habitats. This information is used by Federal, State and local agencies, academic institutions, U.S. Congress and the private sector. The Emergency Wetland Resources Act of 1986 directs the Service to map the wetlands of the United States. The NWI has mapped 89% of the lower 48 states, and 31% of Alaska. The Act also requires the Service to produce a digital wetlands database for the United States. About 39% of the lower 48 states and 11% of Alaska are digitized. Congressional mandates require the NWI to produce status and trends reports to Congress at ten-year intervals. In 1982, the NWI produced the first comprehensive and statistically valid estimate of the status of the Nations wetlands and wetland losses, and in 1990 produced the first update. Future national updates scheduled for 2000, 2010, and 2020. In addition to the status and trends reports, the NWI has produced over 130 publications, including manuals, plant and hydric soils lists, field guides, posters, wall size resource maps, atlases and state reports, and has had numerous articles published by professional journals.

The NWI National Center in St. Petersburg, Florida, includes a state-of-the-art computer operation which is responsible for constructing the wetlands layer of the National Spatial Data Infrastructure. Digitized wetlands data can be integrated with other layers of the NSDI such as natural resources and cultural and physical features, leading to production of selected color and customized maps of the information from wetland maps, and the transfer of digital (computer-readable) data to users

and researchers world-wide. Dozens of organizations, including Federal, State, county agencies, and private sector organizations such as Ducks Unlimited, have supported conversion of wetland maps into digital data for computer use. Statewide databases have been built for 9 States and initiated in 5 other States. Digitized wetland data are also available for portions of 37 other States. Once a digital database is constructed, users can obtain the data at no cost over the Internet, or through the U.S. Geological Survey for the cost of reproduction.

NWI maintains a MAPS database of metadata containing production information, history, and availability of all maps and digital wetlands data produced by NWI. This database is available over the Internet. The six U.S. Geological Survey Earth Science Information Centers (ESIC) have on-line access to the database. The NWI metadata database is maintained in Oracle.

A pilot project was conducted with the FGDC Clearinghouse in Reston to determine if the NWI Oracle metadata database could be made FGDC-compliant and clearinghouse-searchable. The project was completed in under two days. Meta Manager was used to map the existing metadata definitions into the FGDC standard. Again the Meta Manager administrative interface was used to fill in missing FGDC attributes.

6.0 WEB ENTERPRISE SUITE

Meta Manager is a component of the Compusult Web Enterprise Suite of Internet-based applications and toolkits, which support web mapping and information cataloging, discovery, retrieval and delivery services. Meta Manager is also a contributor to the OpenGIS Consortium (OGC) Catalog Services Interoperability initiative.