

Building a Geospatial Data Clearinghouse for Discovery and Access

Douglas Nebert
ddnebert@usgs.gov

Mark Reichardt
International Activities
10 - 13 April 2000
mreichardt@usgs.gov

Agenda

- U.S. National Spatial Data Infrastructure
- Clearinghouse Defined
- Clearinghouse Growth Nationally
- Towards a Global Implementation
- Issues and Opportunities
- Next Steps

The NSDI

- Established by Presidential Executive Order on 11 April 1994

- NSDI Vision:

Current and accurate geospatial data are readily available locally, nationally, and globally to contribute to economic growth, environmental quality, and social progress

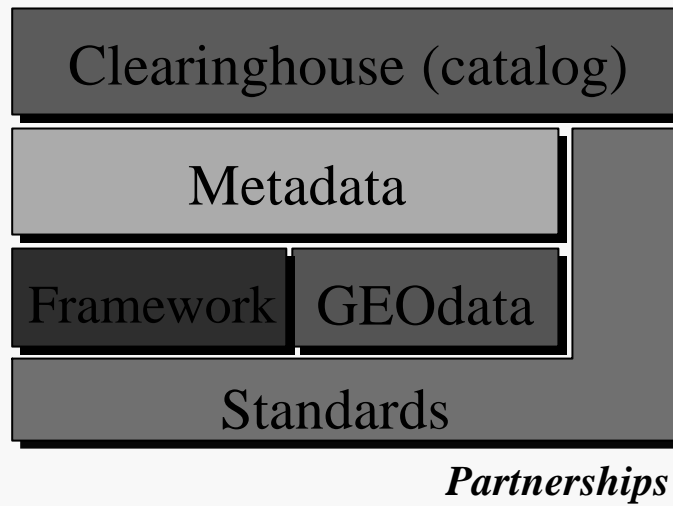


Federal Geographic Data Committee

- Created by the US Office of Management and Budget in 1990 to coordinate geospatial activities across government
- Representation:
 - 17 cabinet and executive level agencies
 - State, local, tribal, academic, industry stakeholders
 - Chaired by Secretary Babbitt
 - FGDC staff supports operations, administered through the U.S. Geological Survey



Core Components of the NSDI



Who Is Involved?



National League of Cities National Association of Counties
Intertribal GIS Council University Consortium on GI Sciences
Open GIS Consortium FGDC
National States Geographic Information Council
International City / County Managers Association

FGDC Standards...

- Created by thematic subcommittees, representing consensus view of data theme
- Submitted for 90-day public review
- Reviewed across disciplines for uniformity
- Published as US Federal Standards
- *Standards by ISO and other national standards are used FIRST!*
- *Standardization makes the Clearinghouse work*

Framework supports...

- Community development of sets of spatial primitives, feature representation, and attribution to a lowest common denominator
- Participant collecting or converting information to common Framework feature specifications
- Multiple representations of real-world features at different scales and times by feature identifier and generalization

Clearinghouse Defined

- A distributed network that includes:
 - a registry of servers
 - several WWW-to-Z39.50 gateways
 - many Z39.50 (GEO) profile servers
- Allows a single query of all or a portion of world-wide servers in a single session



Clearinghouse Defined

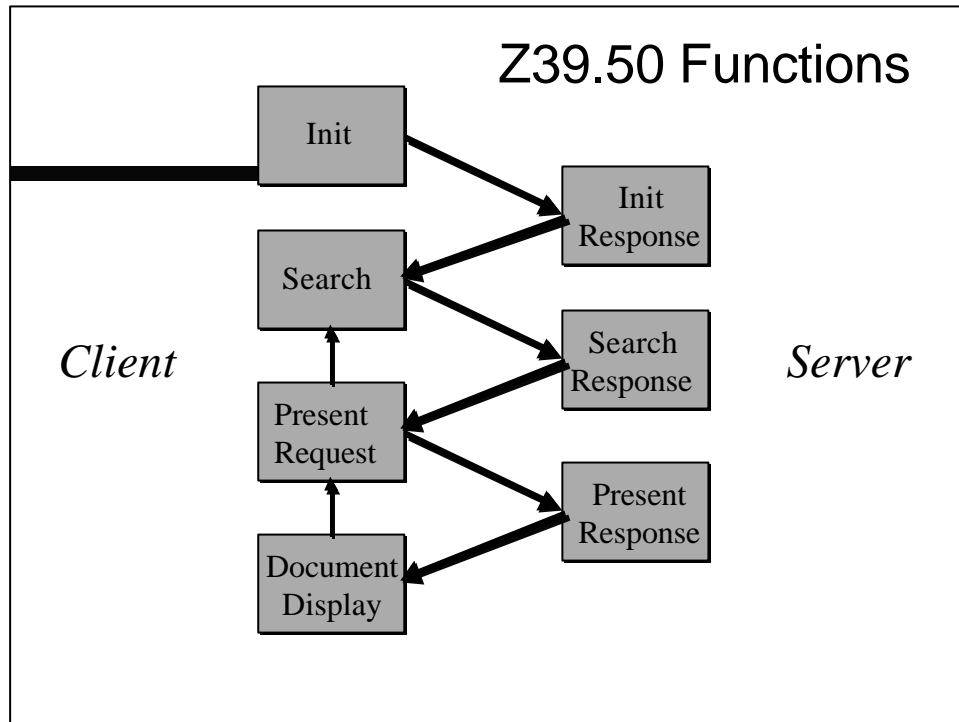
- Links through to full data access where desired / available
- Will not operate without Metadata

ANSI Z39.50/ISO 23950

- Application protocol layered upon OSI or TCP/IP (now HTTP)
- Developed by the library community initially, adopted by museums, chemical abstracts, information locator, now geomatics
- Provides Search and Retrieval services in a platform-independent fashion

Z39.50 Details

- Stateful session-based connection (usually)
- Query expressed using registered integer codes to reference terms, operators, and formats
- Query usually expressed using RPN-style syntax
- Request can return well-known groups of fields Element Sets (what)
- Request also packaged as text, XML, GRS-1, MARC, etc. (how)



What is GEO?

- Implementing guidelines for defining the content and operations for FGDC geospatial metadata in a compliant Z39.50 software system
- Extended set of bibliographic fields to exactly search geospatial holdings
 - Geo coordinates
 - Temporal conditions
 - Free text (i.e. anywhere in metadata)

Z39.50 Profiles are registered

- Identify queryable fields or constructs, operators, structures for re-use
- Specify query syntax to be used
- Define standard groupings of elements to be returned (e.g. Brief, Summary, Full)
- Define expected delivery formats (preferred syntax)

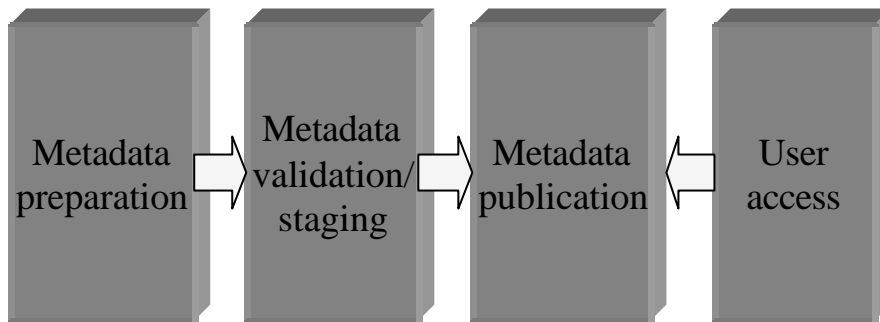
Metadata

Metadata

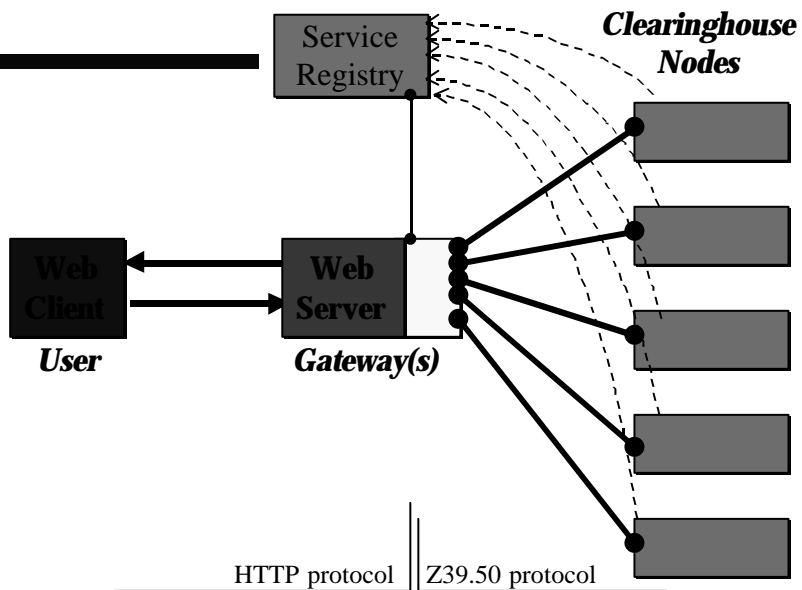


- FGDC Metadata Standard in wide use at national, state, local levels
- Metadata publishing tools available
- ISO TC-211 International Metadata Standard (second draft is in review)
 - FGDC endorses ISO metadata standard
 - FGDC will develop a profile of the ISO standard, and will support nationwide implementation

Clearinghouse Method



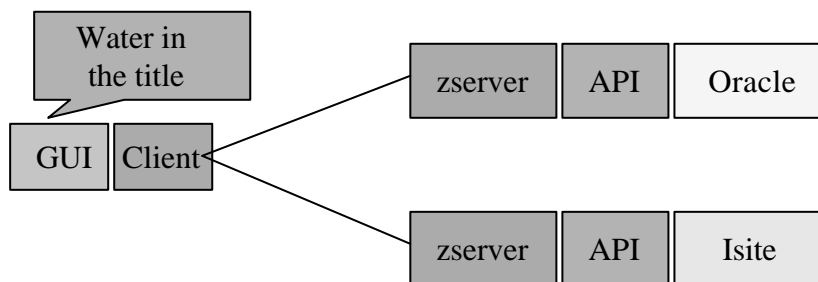
Discovery in Clearinghouse



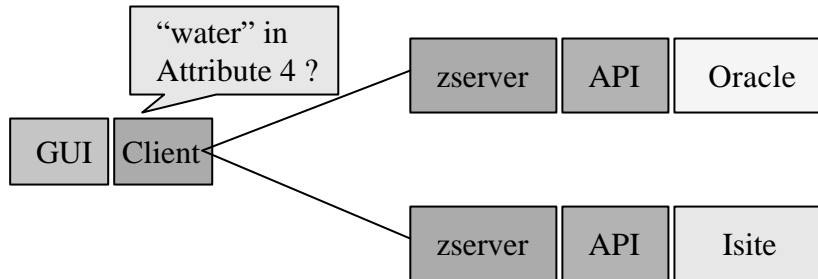
Abstraction

- Z39.50 allows us to separate the implementation of a local database and its data element name from an external or public set of fields and format
- Abstraction allows metadata providers to select from various search engines or databases
- Supports semantic translation between discipline, language, and other communities through server-side “mappings” of public to local fields

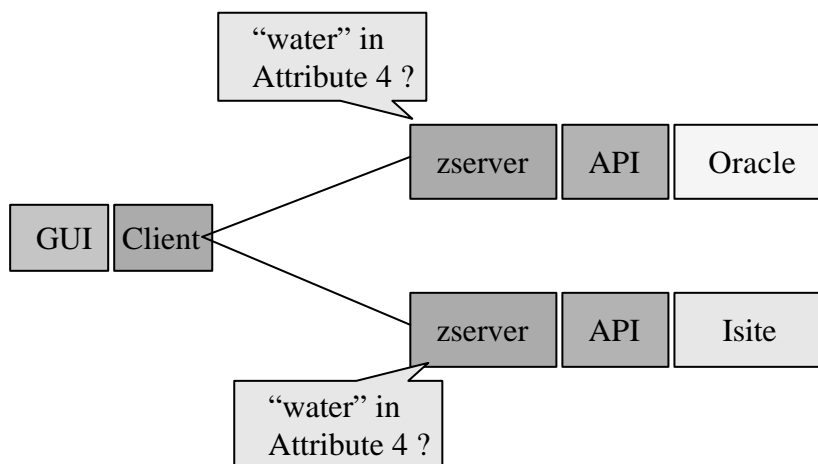
Abstraction of Query



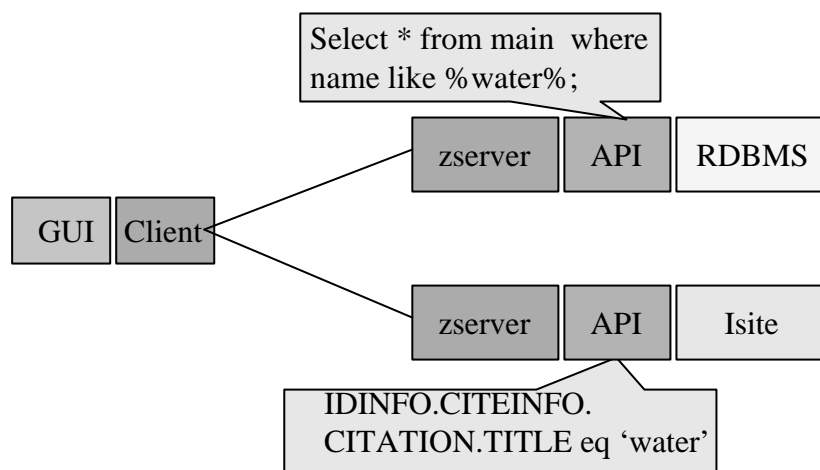
Abstraction of Query



Abstraction of Query



Abstraction of Query



Clearinghouse Interfaces

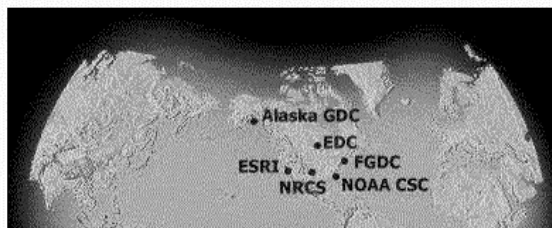
- Access to Clearinghouse may be via:
 - formal Gateway search forms hosted at FGDC and participating organizations
 - Custom “local” access web forms
 - Custom applications that support standard search using ISO 23950/Z39.50
- Next slides show one customized search

Search for Geographic Data
FGDC
Federal Geographic Data Committee
National Geospatial Data Clearinghouse

status
web links
help

The Geospatial Data Clearinghouse is a collection of over 100 spatial data servers, that have digital geographic data primarily for use in Geographic Information Systems (GIS), image processing systems, and other modelling software. These data collections can be searched through a single interface based on their descriptions, or "metadata."

Click on the name of a Clearinghouse Gateway on the map below that is closest to you and you will be presented with a selection of query forms. All entry points or gateways shown have exactly the same lists of servers.



Generate a search form to include the following properties:

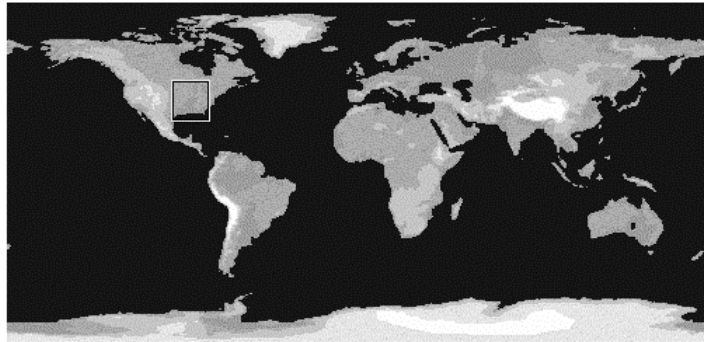
| | |
|----------------------------------|---|
| Geographic Location | <input checked="" type="checkbox"/> Let me pick a place name from a list <input checked="" type="checkbox"/> Search worldwide <input checked="" type="checkbox"/> Let me draw a search area on a map <input checked="" type="checkbox"/> Let me search by latitude and longitude |
| Time Period of Content | <input checked="" type="checkbox"/> Find data for all time periods <input checked="" type="checkbox"/> Let me define a specific date for search <input checked="" type="checkbox"/> Let me define a date range to search within |
| Full Text or Field Search | <input checked="" type="checkbox"/> Let me run a full-text search <input checked="" type="checkbox"/> Let me search against specific fields |
| Available Data Sources | <input checked="" type="checkbox"/> Let me search all available servers <input checked="" type="checkbox"/> Let me pick which servers I want to query |

Go make a custom search form

Reset Form

clearinghouse search

Define the Geographic Area of Interest

[Help](#)Cylindric (Atlantic) 

Search by Field

[Help](#)Search for the word: anywhere in the metadata

| | | |
|--|-------------------|----|
| Arizona Clearinghouse Node for Spatial Data | Search Successful | 0 |
| <u>United Nations Environment Programme - GRID - Sioux Falls</u> | Search Successful | 3 |
| Africa Data Dissemination Service | Search Successful | 0 |
| <u>Inter-American Geospatial Data Network</u> | Search Successful | 2 |
| <u>Illinois Natural Resources Geospatial Data Clearinghouse Node</u> | Search Successful | 15 |
| Helena National Forest, Montana | Search Successful | 0 |
| Greater Yellowstone Area Data Clearinghouse | Search Successful | 0 |
| Global Environmental Information Locator Service | Search Successful | 0 |
| Georgia GIS Data Clearinghouse | Search Successful | 0 |
| Geological Survey of Alabama Geospatial Data Clearinghouse Node | Search Successful | 0 |
| FEMA Flood Insurance Rate Maps, Q3 Flood Data Sampler | Search Successful | 0 |

clearinghouse results

The total number of metadata entries found matching the query is: 1578

The total number of entries to be returned by this query is: 2

Returned 2 of 2 entries.

Returned entries 1 to 2 of 2.

Metadata entry titles and links:

Soils Map of Mexico 1:1000000

[View the full record.](#)

U.S. Federal Region III Land Cover Data Set

[View the full record.](#)

clearinghouse results

Full FGDC Metadata Record

Soils Map of Mexico 1:1000000

Metadata:

- [Identification Information](#)
 - [Data Quality Information](#)
 - [Spatial Data Organization Information](#)
 - [Spatial Reference Information](#)
 - [Entity and Attribute Information](#)
 - [Distribution Information](#)
 - [Metadata Reference Information](#)
-

Identification_Information:

Citation:

Citation_Information:

Originator:

Instituto Nacional de Estadística Geografía e Informática (INEGI)

Publication_Date: 19970922

Title: Soils Map of Mexico 1:1000000

Edition: 1.0

Geospatial_Data_Presentation_Form: map

Publication_Information:

Publication_Place: Kingston, Jamaica

Publisher: The Caribbean Environment Programme (CEP)

Other_Citation_Details:

The Regional Programme on Information Systems Management of Marine and Coastal Resources (CEPNET) is providing the Soils Map of Mexico with the cooperation of Mexico's National Institute of Statistics, Geography and Informatics (INEGI).

Online_Linkage: <URL:<http://www.cep.unep.org/data/north/north.html>>

Description:

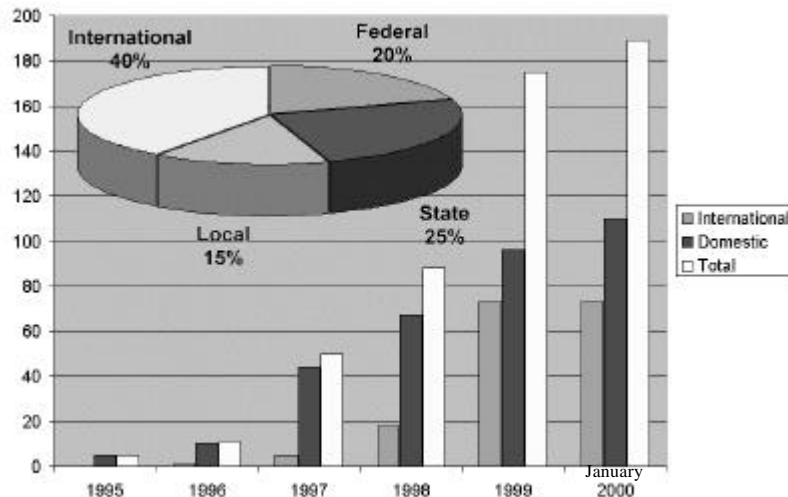
Abstract:

Mexico's National Institute of Statistics, Geography and Informatics (INEGI) was created by presidential decree on January 25, 1983. INEGI generates, incorporates,

How do servers cooperate?

- Global registry of Clearinghouse servers is maintained by FGDC on behalf of the Global Spatial Data Infrastructure activity
- Each gateway communicates with the server registry to update search forms and parameters
- Service Registry will provide for intelligent selection of servers based on server metadata (Q2 FY 2000)

NSDI Clearinghouse Growth 1994 – 1999



Towards a Global Catalog of Geospatial Data and Services

• Many Z39.50 Gateway and Server resources available on the Internet today:

- FGDC National Clearinghouse
- CEONet Canada
- Australian Spatial Data Directory
- South African National Spatial Information Framework

Z39.50 Geo / Metadata Compliant Sites

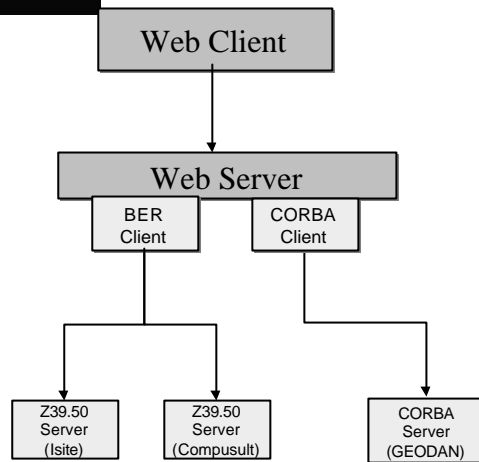
Clearinghouse Nodes Around the World



Gateway Design

- Gateway acts as a WWW (HTTP) to Z39.50 translator or protocol bridge
- Enables search in parallel to multiple servers
- What's Out There?
 - Isite zgate (UNIX) is reference gateway functionality (serial search only)
 - Blue Angel MetaStar Gateway is commercial solution

Blue Angel MetaStar^(tm) Gateway



Server (Node) in More Detail



Available Server Solutions

- Isite is XML-database solution using open source as a reference implementation of a searchable metadata service
- YAZ Toolkit (www.indexdata.dk) is a development system for building service solutions

Commercial Clearinghouse Server Solutions

- CompuSult MetaManager Toolkit to build an FGDC-compliant server on top of a RDBMS (Oracle, MS SQL-Server, Access) <http://www.fgdctoolkit.com>
- Blue Angel Technologies has a MetaStar Repository for support of many types of metadata (GILS, DC, MARC, GEO, etc.) <http://www.blueangeltech.com>
- Isite OpenSource: free use and download from <http://clearinghouse4.fgdc.gov/ftp>

Isite solution

- Isite software suite includes Z39.50 server and XML-based (1.0) document database
- Runs on Windows 9X/NT and UNIX
- Efficient indexes for thousands to hundreds of thousands of entries
- Includes "script" API to interface to DB
- Version 2.07 to have mySQL/Oracle backend examples
- Isite OpenSource: free use and download from <http://clearinghouse4.fgdc.gov/ftp>

Contents of Isite Installation



• index • Izclient • zserver.ini • *bibl.fgdcmap*
• Isearch • Iutil • sapi.ini • *gils.fgdcmap*
• zclient • zserver • *fgdc.fields* • *geo.fgdcmap*



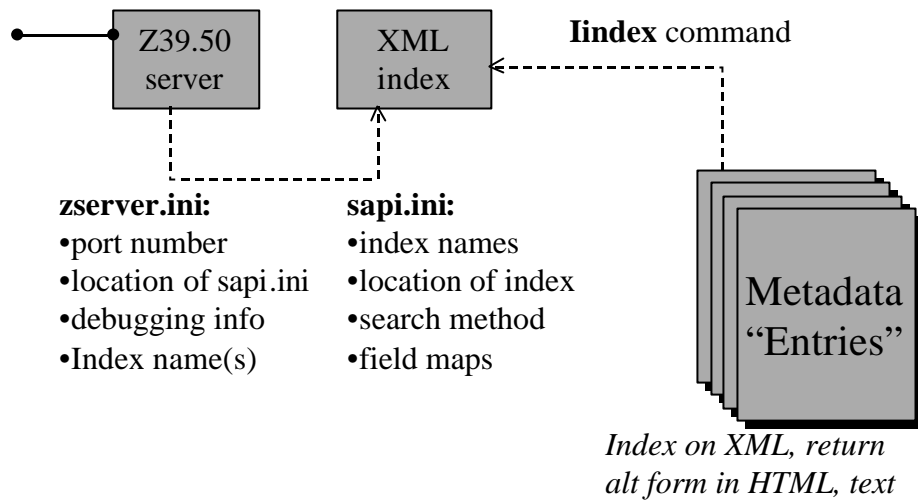
Place XML, HTML, and TEXT here



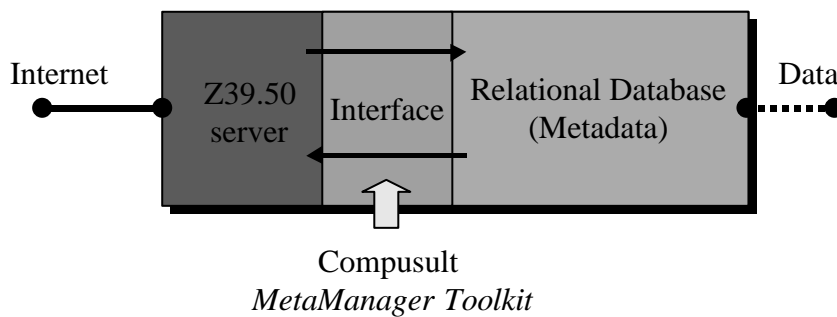
Direct Index to place the index files here

```
Index -d \index\oregon -t fgdc  
-o fieldtype=fgdc.fields \metadata\*.sgml
```

Isite Architecture



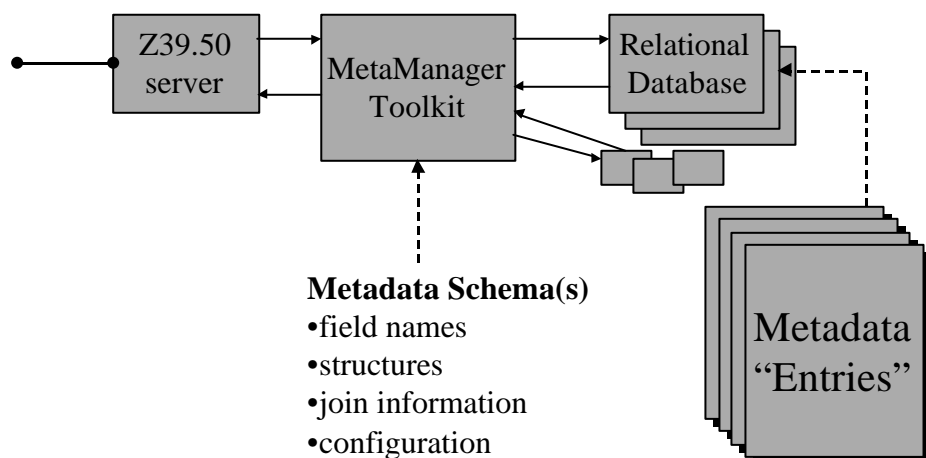
Database Option



Compusult MetaManager

- Allows FGDC metadata to be stored in many tables (ISO within the year)
- Preserves hierarchy/collection metadata
- Manages multiple collections at a site
- Allows re-usable fields as static value, function, field value(s), or combination
- Includes web search interface
- Available for Access (\$1995) and Oracle (\$5000)

MetaManager^(tm) Design



Clearinghouse methods and protocol usage were offered as the "WWW Profile" of OpenGIS Catalog Services Request for Proposals approved in August 1999.

Catalog Services and the OpenGIS Consortium


- OpenGIS Consortium enlisted member organizations to develop an open Catalog Services Standard
- A team comprised of international government, industry
- OGC Standard recently approved
- Publicly available at www.opengis.org

Web-based Mapping Defined

- **Users easily and rapidly search, retrieve, access & exploit geospatial information from multiple locations on the Internet**



DISTRIBUTED



MULTI-
VENDOR

Open Technology specifications that enable Internet-wide exploitation of geodata and sharing of processing services

Better address the vast opportunities to utilize NSDI resources via the Internet

Web Mapping Testbed

OpenGIS Consortium (OGC) members conducted Web Mapping Testbed Phase I over 4 months to:

- define interfaces for web mapping services based on existing mature technology
- demonstrate integration of multiple OGC services together with different vendor products
- Publish open technology standards
- Redefine the environment for web based spatial applications

3. *LandSAT with Roads & Hospitals*



Ionic Software
Viewer Client

MIT
LandSAT
image

Laser-Scan
Roads

Object/FX
Hospitals

3. *Road Restrictions to Emergency Vehicles*



Ionic Software
Viewer Client

MIT
LandSAT image

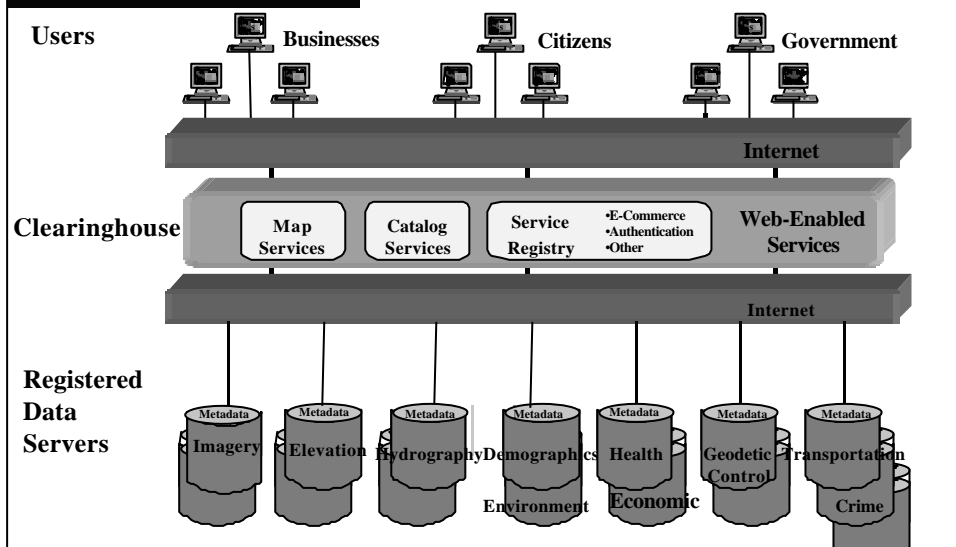
Laser-Scan
Roads
Hurricane
Flooding

Object/FX
Hospitals

Future Directions

- Global Service Catalog of geospatial services under the GSDI initiative (2000)
- Metadata management tools integrated with vendor products – metadata as part of the information management process
- Semantic mapping – consistent thesaurus, multilingual searches across global data bases.
- Customized web-based geospatial applications reduce cost, increase access, and better support decision-making

Tomorrow: A Global Infrastructure Enabled Through Partnerships, Standards, Technology



Additional Information

- www.fgdc.gov Federal Geographic Data Committee
- www.opengis.org OpenGIS Consortium
- www.gsdi.org Global Spatial Data Infrastructure