INVENTION DISCLOSURE ACKNOWLEDGMENT

August 23, 2002

Docket Number Assigned: 2979

To: Wesley C. Patrick

From: Louis Rodriguez, Legal Department

Title of Invention: ***Olympus Data and Information Sharing System™ (Olympus DISS™)***

Action to be Taken:

This Invention Disclosure (ID) was conceived as the result of work on a project for the United States Nuclear Regulatory Commission under which Southwest Research Institute (SwRI) has the right to retain title to any inventions conceived or first reduced to practice during the course of work under the contract. The SwRI Patent Committee will make a determination as to whether or not a request to the United States Government is made to retain title. If SwRI elects to retain title rights to the invention, the U.S. Government would still have a royalty-free license to practice the patent anywhere in the world for Government purposes. After the determination is made by the Patent Committee, the decision will be reported to the sponsor. If the Institute does not request greater rights, the sponsor will then be free to either prosecute a patent on behalf of the U.S. Government or permit the technology to become public knowledge. In either event, we will maintain the original ID in the Institute's permanent files.

Please refer to the above-assigned docket number in all correspondence/transmissions to this office. Thank you.

Enclosure
cc (w/enc): Dorel Marius Necsoiu
Brandi Winfrey
Ron Kalmbach

SOFTWARE REQUIREMENTS DESCRIPTION
Olympus Data and Information Sharing System, Version 1.00 (Olympus DISS)

Prepared by

M. Necsoiu
B. Winfrey

Center for Nuclear Waste Regulatory Analyses
San Antonio, Texas

February 2002

Approved by:

H. Lawrence McKague

H. Lawrence McKague, Element Manager
Geology and Geophysics

2/19/02

Date

CONTENTS

FIGURES

1 SOFTWARE FUNCTION

Geographic Information Systems (GIS) may consist of hundreds or thousands of geographic data files. Each file usually has a large amount of metadata attached to it. Metadata describes and characterizes the geodata files. A system is needed to systematically organize and search the metadata files for data files within a specific geographic location, of a particular type, or from a certain time period. In initially scoping out a design for the Olympus system, several different software codes and configurations were tested. The final configuration met planned goals. This SRD describes the existing configuration and the constituent software that makes the Olympus system. However, changes may occur in the future which will be documented in the Software Change Reporting and Resolution Report.

Olympus was designed to provide the necessary organization and search capabilities. It is an intranet web-based geographical data and information sharing system designed to search and retrieve geographic data at its source. Olympus uses established data standards, reference such as the Federal Geographic Data Committee standard (FGDC), providing a flexible mechanism to build applications upon.

The system software will be centralized, using several software components (Figure 1):

ArcCatalog, a Commercial Off-the-Shelf (COTS) software component produced by ESRI, Inc. (www.esri.com);

mp and Isite information system, two public domain packages produced by USGS (www.usgs.gov) and the Center for Networked Information Discovery and Retrieval (CNIDR) (www.cnidr.org). Isite is a Z39.50 protocol gateway software including a web interface and the Isearch engine;

Harvester will be a software component produced by CNWRA.

Olympus DISS will provide search and retrieval mechanisms for querying a metadata database, containing records associated with each specific geographic dataset (geodata). Each metadata file will contain information that describes geodata in the same way a card in a library card catalog describes a book.

Olympus could create and "ingest" metadata from a variety of geodata formats: ERDAS IMAGINE, TIFF, MrSID, JPEG, ERDAS 7.5 LAN, ERDAS Raw, ESRI GRID Stack File, ESRI Shapefile and ESRI Arc/Info coverage. Once metadata and the associated geodata are created, the user(s) will place it in a designated repository area. Periodically the Olympus system will "harvest" metadata in the repository and automatically build an index and a relational database with metadata information.

Olympus will provide geographic (spatial), keyword, and temporal search and retrieval capabilities for the repository of geographic data. It will do this through a web based graphical user interface. This user interface will consist of a login page (Figure 2), search page (Figure 3), result page (Figure 4), and metadata pages (Figure 5).

Spatial searches will allow the user three methods of entering data for queries. The first method will be to enter geographic coordinates in the provided text fields. The second method will be to simply draw a box around the desired geographic location on a map of either the United States

or the world. The third method will be to select, from a dropdown menu, the state (in the United States) that the user wishes to search on. If a map is used, the select tool will be personalized by selecting both the color and style from two dropdown menus. The selectable style will be either Point, Point (Compressed X and Y), or XY Plane. In all methods, the geographic coordinates will be the values queried for in the database.

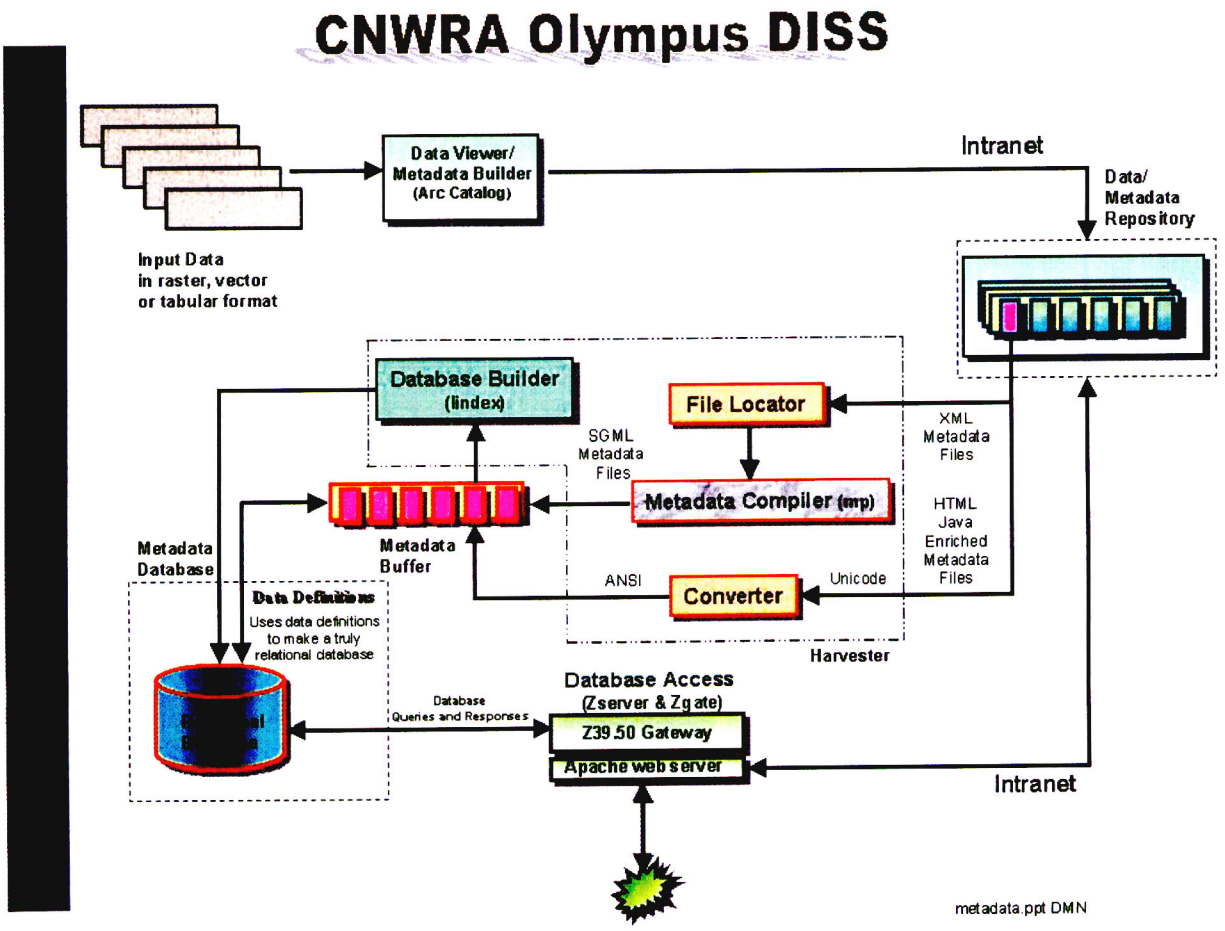

Figure 1. Data Flow

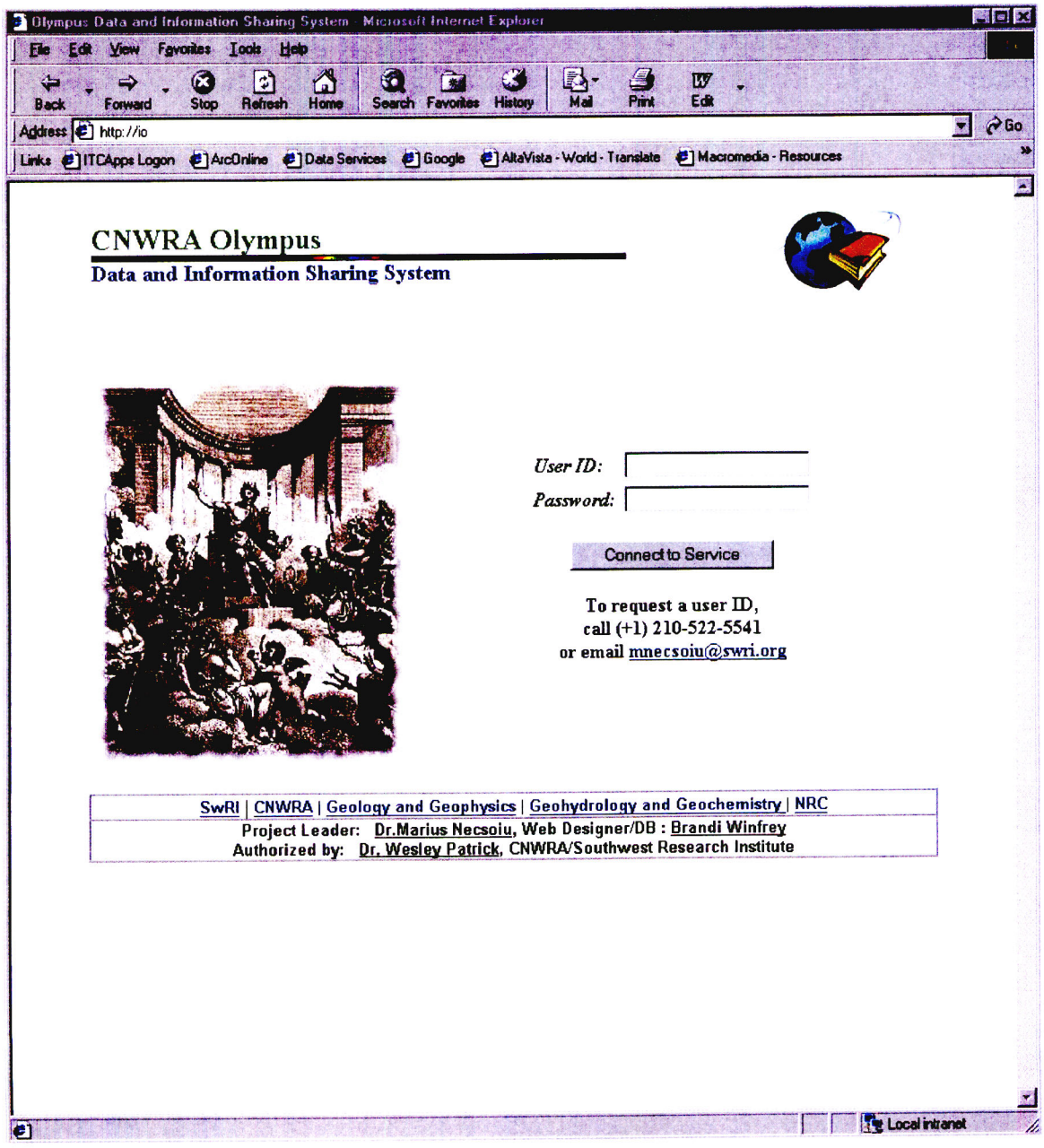

Figure 2. Conceptual User Interface – Login Page

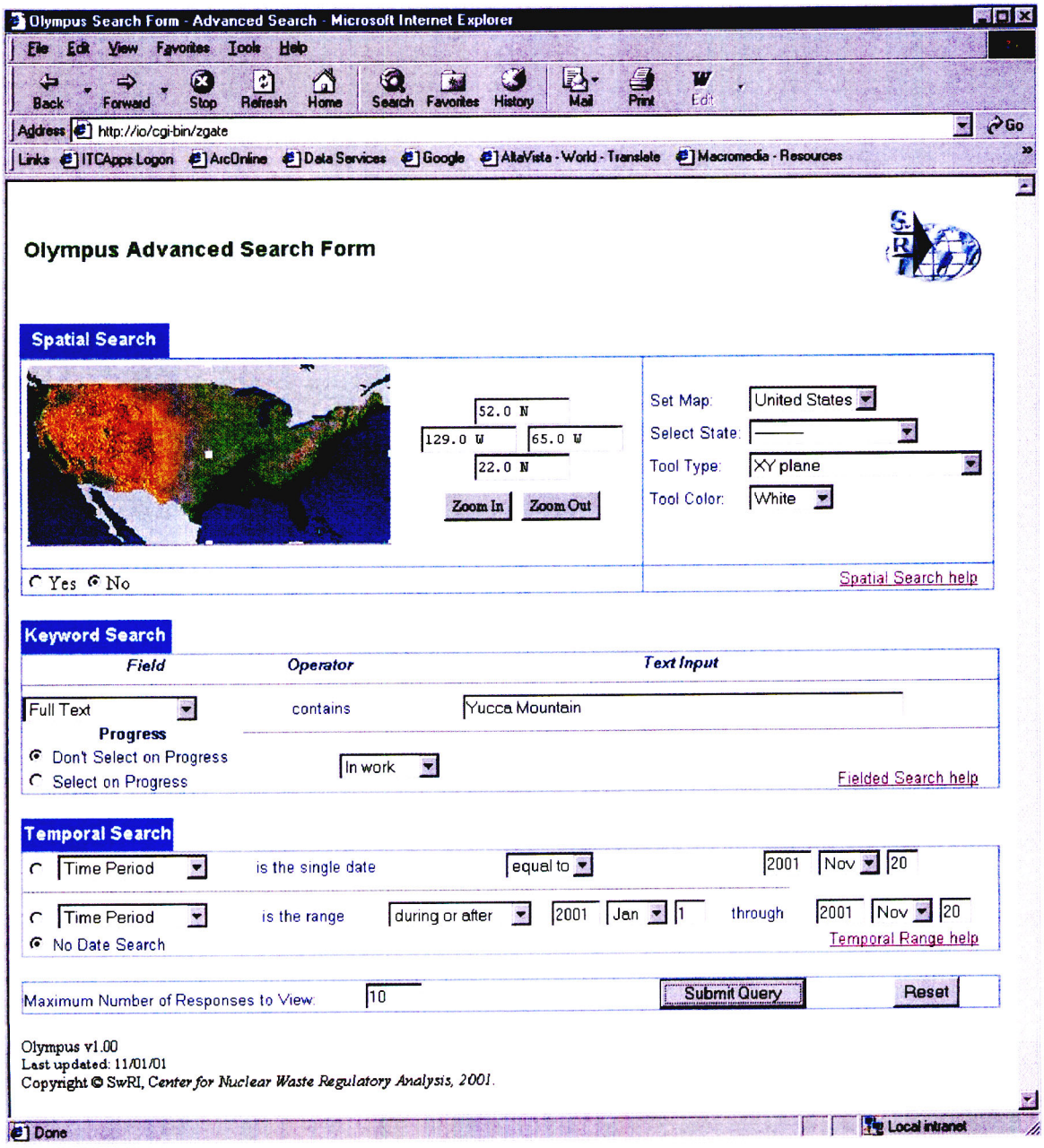

Figure 3. Conceptual User Interface – Search Form

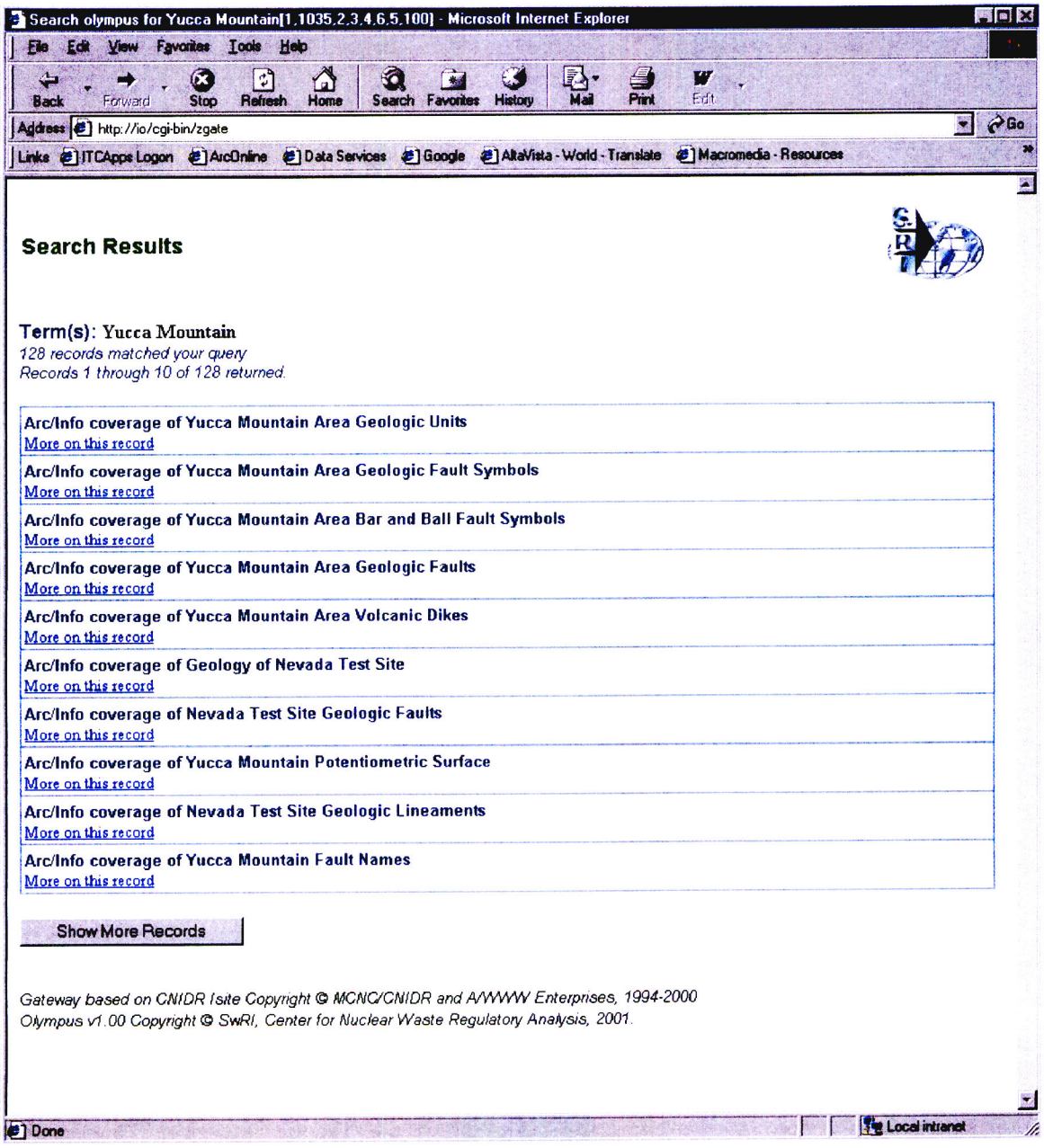

Figure 4. Conceptual User Interface – Search Results

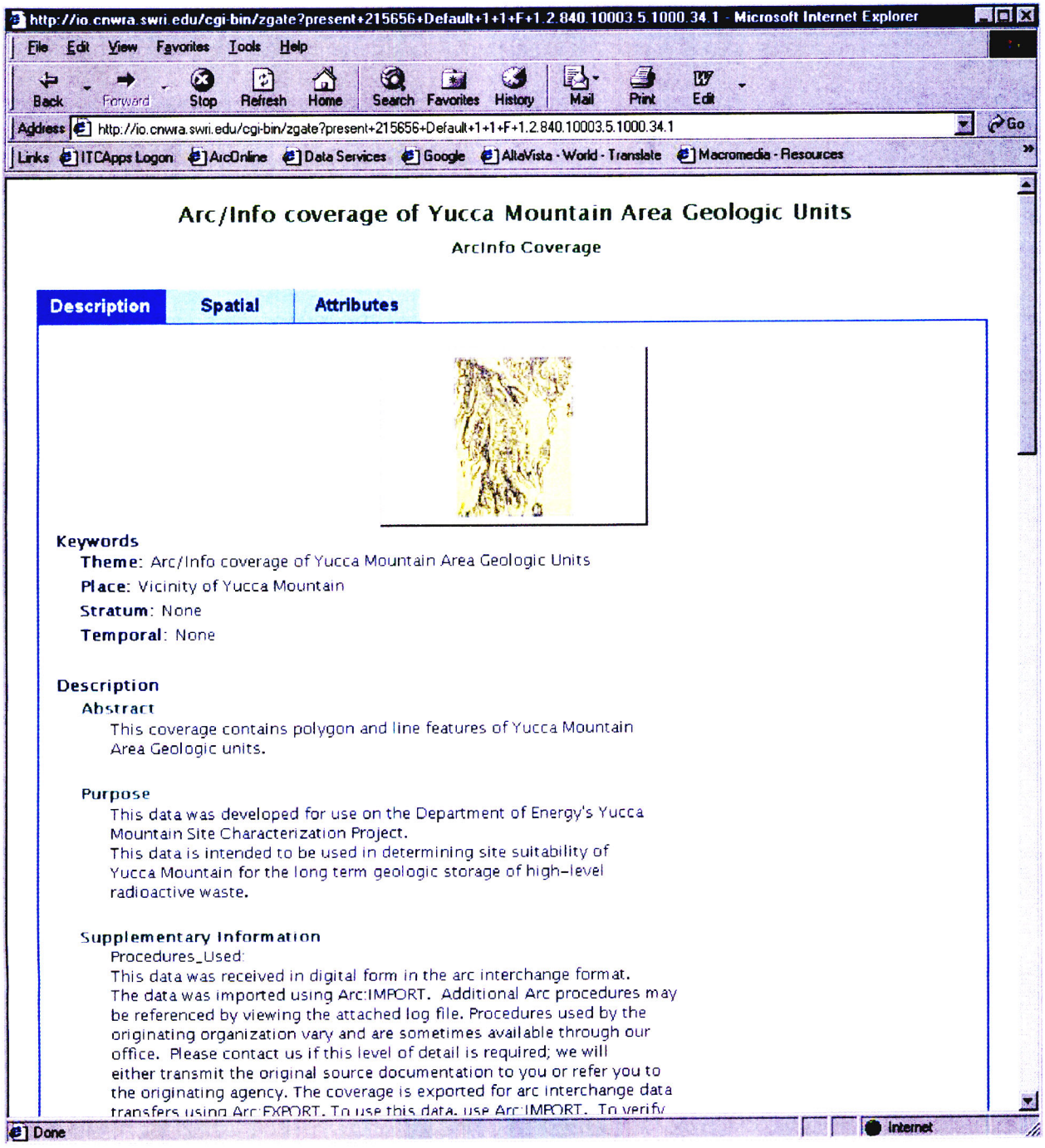

Figure 5. Conceptual User Interface – Metadata

The Keyword Search option will allow the user several ways to search for text values. There will be a dropdown menu of several types of text to search on as well as a text field in which to enter one or more keywords. An asterisk (*) could be used to represent wild card characters. There will also be an option to search for data in different stages of development (i.e., planned, in work, and complete).

The Temporal Search option will have two methods of searching for dates. The first will search for a single date which can be equal to, before, or after the single date listed. The second method will search for a range of dates equal to, before, or after the two dates listed.

All three options (i.e., Spatial, Keyword, and Temporal) may be selected independently or in conjunction with each other to better define the user's search criteria. If an error is made in the selection process, the user will be able to Reset the page back to its original default values and begin the criteria selection again. Finally after completing the desired selections, the user will have the option to list the maximum number of responses to view on the result page before submitting the query.

The results page (Figure 4) will list the term(s) queried for, the number of matching records found, the number of records currently being viewed, and the titles and links to Metadata describing the data available. From this page, the user will either click on a link to the desired Metadata page (Figure 5) or click on the button to show more results. If the user uses the Metadata link, he/she will be directed to a page created by ArcCatalog showing geodata information. The Metadata page will have a few standard features. There will be three tabs to choose from, Description, Spatial, and Attributes, each of which will have the same outline of metadata filled out to describe a particular set of data. In the text, headings, which will be colored green, will expand when clicked to show more detailed descriptions, and will contract when clicked a second time. Some metadata pages will have a thumbnail picture to visualize the data.

2 TECHNICAL BASIS: PHYSICAL AND MATHEMATICAL MODEL

The software will not solve any mathematical equations; it will locate and format data files to be used in a relational database. There will be three CNWRA generated parts to this software, the (i) *Harvester* Perl script, (ii) ArcCatalog Olympus button Visual Basic code that links to ArcCatalog, and (iii) HTML/JavaScript based web pages.

Data flow of Olympus (Figure 1) will basically consist of a web page front-end interacting with the web server (Apache), which will interface with the Z39.50 Gateway. Isite and its components zserver and zgate, developed and maintained by CNIDR, will provide access to database systems via the ANSI/NISO Z39.50 search and retrieval protocol. The Z39.50 communications server, Zserver, will be linked with the CNIDR Search Application Programming Interface (API) in order to accomplish this task. Olympus will have four major components, (i) Data Viewer/Metadata Builder, (ii) Data/Metadata Repository, (iii) Relational Database, and (iv) Harvester (Figure 1). The GIS Data/Metadata Repository will be filled with Metadata created by ArcCatalog and located by the Harvester. The Harvester will be composed of the Converter, the File Locator, the Metadata Compiler (mp), and the Database Builder. The Database Builder will be based on Iindex software, included in the Isite package, and the Metadata Compiler will be based on mp, both packages being managed by the internally generated Harvester code. For a brief description of each processor, please see paragraph 3.4.

ArcCatalog provides a metadata editor that will be used by Olympus to document any newly added set(s) of geodata. The Catalog will fill in as much information as it can, using the data set's properties, however the user will need to fill specific fields such as the dataset abstract, purpose, point of contact or calendar date. The complete list of the minimum specific fields required provided in Appendix 1. If the data changes, the Catalog will automatically update the associate record with the new information. Metadata will become an integral part of the Geographic Coverage and will follow the coverage when copied or moved to a new location.

In ArcCatalog there are several ways that the same metadata record can be viewed. This will be done by using different stylesheets from the dropdown list on the Metadata toolbar (Figure 6). However, Olympus will use only the ESRI style sheet for "ingesting" Metadata. The ESRI stylesheet organizes data in three categories of information. The Description tab will include information about the status of the data source, its location, and any enclosed files. The Spatial tab will show the data's extent, as well as detailed feature or raster properties. The Attributes tab will describe each attribute and lists the relationships in which the data source participates. To "ingest" the data the user will click on the ArcCatalog Olympus Export button to save the current XML Metadata file in HTML enriched format.

JavaScript will be used in the web page to communicate with the LiveMap 3.0 Java applet. The LiveMap applet will be used to retrieve geographic coordinates for interacting with the geodata entries in the database. The layout of the interface screen will be based on a template generated by the Federation of Earth Science Information Partners (FESIP, 2001).

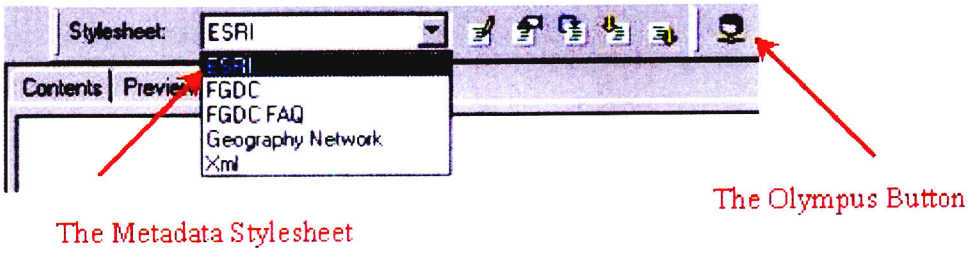

Figure 6. ArcCatalog – The Metadata Stylesheet and Olympus. Conceptual Placement of the Olympus Button on the ArcCatalog Menu.

3 COMPUTATIONAL APPROACH

3.1 Data Flow and User Interface

3.2 Hardware and Software Requirements

- Target platform(s):

  The Olympus database will be developed and run on Silicon Graphics/Irix. The Graphical User Interface will be developed on Windows and will run on any platform with the Microsoft Internet Explorer web browser (IE) application available. Olympus will not work with Netscape.

- Operating System(s):

  The database will operate on Irix, but the web-based interface will be able to run on any operating system.

- Programming language(s):

  Java, JavaScript, HTML, VB and Perl

3.3 Graphics Requirements

The Olympus GUI will be designed for Microsoft Internet Explorer.

3.4 Pre- and Post-Processors

There are no pre or post processors. All of the processing is considered a part of the Olympus system. A brief description of each processor follows.

ArcCatalog, creates metadata and saves it as HTML in Unicode format.

Metadata Compiler (mp), parses formal metadata, checking the syntax against the FGDC Content Standard for Digital Geospatial Metadata and generating output suitable for viewing with a web browser or text editor. It runs on UNIX systems and on PC's running Windows 95, 98, or NT.

lindex, is a component of Isite package which reads FGDC SGML records, created by mp and creates a set of index files. These files are stored in the Isite/bin/db directory where the zserver (another component of Isite package) will use them to answer external queries.

Zserver, another component of the Isite package, is the application which serves the metadata records to the Intranet. Zserver uses the indexes to answer queries. One of the client (i.e., zgate) is used by the Olympus system. Configuring zserver involves installing the package and editing two files, zserver.ini and sapi.ini.

Harvester, will be a CNWRA generated Perl script that will localize newly generated xml files from a predefined directory location, create an sgml version of them, and feed an lindex database. Also, the program will copy and customize html rich versions of the metadata files by

placing them in the same location with the xml files. *Convert*, will convert HTML Unicode to HTML ASCI files so that the Z39.50 server (zgate) can read them.

3.5 Software Validation

Validation of version 1.00 is scheduled for completion on March 25, 2002.

4 REFERENCES

Gateway based on CNIDR Isite Copyright © MCNC/CNIDR and A/WWW Enterprises, 1994-2000

LiveMap 3.0 JAVA applet is provided courtesy of Jonathan Callahan at Pacific Marine Environmental Laboratory: callahan@pmel.noaa.gov

Federal Geographic Data Committee. FGDC-STD-001-1998. Content standard for digital geospatial metadata (revised June 1998). Federal Geographic Data Committee. Washington, D.C.

FESIP, 2001. Federation of Earth Science Information Partners search page, web document, http://mercury.ornl.gov/esip/

Necsoiu, Marius. Capability Development: Olympus CNWRA's Distributed Data and Information Sharing System (White Paper). June 26, 2001.

APPENDIX

REQUIRED FIELDS FOR THE METADATA TO BE USED BY ARCCATALOG FOR THE OLYMPUS SYSTEM

APPENDIX

REQUIRED FIELDS FOR THE METADATA TO BE USED BY ARCCATALOG FOR THE OLYMPUS SYSTEM

Under General **Tab**

Field	**Description of the required information**
Abstract	A brief narrative summary of the data set
Purpose	A summary of the intentions with which the data set was developed
Access Constraints	Restrictions and legal prerequisites for accessing the data set
Use Constraints	Restrictions and legal prerequisites for using the data set after access is granted
Point of Contact	Define the primary contact as person or organization

Under the **Citation** Tab

Field	**Description of the required information**
Originator	The name of an organization or individual that developed the data set
Publication Date	The date when the data set is published or otherwise made available for release

Under the **Time Period** Tab

Field	**Description of the required information**
Currentness reference	The basis on which the time period of content information is determined
Calendar Date	The year (and optionally month, or month and day) for which the data set corresponds to the ground condition

Under the **Status** Tab

Field	**Description of the required information**
Progress	The state of the data set. The options are: Complete, In Work, and Planned

Update Frequency	The frequency with which changes and additions are made to the data set after the initial data set is completed. The options are: Daily, Weekly, Monthly, Quarterly, Annually, Continually, As needed, Irregular and Unknown

Under the **Spatial Domain** Tab

Field	**Description of the required information**
Bounding Coordinates and G-Polygon	Northern/Southern/Eastern/Western-most coordinate of the limit of coverage expressed in latitude

Under the **Keywords** Tab

Field	**Description of the required information**
Keyword	Common-use word or phrase used to describe the subject of the data set
Thesaurus	Reference to a formally registered thesaurus or a similar authoritative source of theme keywords
Place	Dataset Location

Software Validation Test Plan

SOFTWARE VALIDATION TEST PLAN
Olympus Data and Information Sharing System™, Version 1.00 (Olympus DISS™)

Prepared by

D. Marius Necsoiu
Brandi L. Winfrey

Center for Nuclear Waste Regulatory Analyses
San Antonio, Texas

October 2003

Approved by:

H. Lawrence McKague, Element Manager
Geology and Geophysics

11/7/03
Date

Table of Contents

1 SCOPE OF VALIDATION

Olympus was designed to systematically organize and search metadata files by geographic location, data type, or time period. It is an intranet web-based geographical Data and Information Sharing System (DISS) designed to search and retrieve geographic data at its source [Software Requirements Document (SRD) Feb 2002]. Software validation of Olympus Data and Information Sharing System™ (Olympus DISS™) should test the following capabilities commensurate with their use in regulatory review.

a) The system shall have a user-friendly interface.

b) The system shall allow ingestion of different data formats (i.e., raster, vector shapefiles/coverages, and tabular data) based on the Federal Geographic Data Committee (FGDC) standard.

c) The system shall have the capability to access data over the intranet, and if necessary, internet.

d) The system shall be able to handle large numbers of files.

e) The system shall respond accurately to queries in a reasonable amount of time.

2 REFERENCES

- Software Requirements Description – Olympus Data and Information Sharing System, Version 1.00, February 2002.
- The following is a list of registered trademarks used in this document:

ESRI® and ArcINFO®, are registered trademarks of ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE, INC.

ArcCatalog™ is a trademark of ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE, INC.

ENVI® is a registered trademark of Better Solutions Consulting LLC.

ERDAS® and ERDAS IMAGINE® are registered trademarks of Erdas, Inc.

MrSID® is a registered trademark of LizardTech, Inc.

Netscape Navigator® is a registered trademark of Netscape Communications Corporation.

Microsoft Internet Explorer® is a registered trademark of Microsoft Corporation

Apache™ is a trademark of The Apache Software Foundation.

iSite Development has registered for federal trademark protection.

3 ENVIRONMENT

3.1 Software

The following software must be installed on the computer(s) used to perform testing activities. Software listed below describes the environment Olympus DISS™ will be tested on as a minimum. Older versions may work but will not be tested. Newer versions may or may not work.

SERVER

Silicon Graphics/Irix version 6.5.14
Apache HTTP Server version 1.3.22
Java version 1.2.2
Perl: perl5.00503-n32 (4.36 or newer)
ArcGIS version 8.1
Metadata Compiler (mp) version 2.7.6
Isite version 1.04
Harvester version 1.0
LiveMap 3.0 Java Applet

CLIENT

Microsoft Internet Explorer version 5.50.4134.0600

3.2 Hardware

There must be a server machine with the Olympus DISS software installed (refer to Section 3.1) and a client with a web browser (see Section 3.1) for remote searches. The server and client may reside on the same computer or on separate, networked computers. For clients on separate computers, there must be a network or internet connection.

4 PREREQUISITES

The prerequisites necessary to perform testing activities are that the Apache HTTP Server be running on the Olympus DISS server.

5 ASSUMPTIONS AND CONSTRAINTS

- Netscape Navigator will NOT work with Olympus DISS.
- The term "reasonable," when referred to performance (i.e., time), is relative to the method with which the database is queried. For example, it will take a longer amount of time to query when using a Virtual Private Network (VPN) connection than it will take to query from the same machine that the database is located on. Also, for the purpose of testing, a "reasonable" amount of time will be defined as a period of time that affords satisfaction to the user and is not extreme or excessive.

- Assume that testers have a general knowledge of spatial data.

6 TEST CASES

6.1 10,000 Data Sets – Capability to Handle Large Numbers of Files

6.1.1 Objective

Test system performance of the Olympus DISS client and server with large quantities of data files.

Requirements:

- System will respond accurately to queries in a reasonable amount of time.
- The system will be able to handle large numbers of files.

6.1.2 Test Input

Any search from the Olympus DISS Graphical User Interface on any client, that will query the database can be used as input for this test. For this test, the database will contain a minimum of 10,000 entries.

6.1.3 Test Procedure

TEST A: Verify that the server can handle a large quantity of data sets

1. The maintainer of the database system will insert a minimum of 10,000 data records into the database.
2. Does ingesting this quantity of data take an excessive amount of time? It should only take 1.3–2 hours.
3. Does the ingestion of this many records cause the system to halt? Or fail?

TEST B: Verify that the client can successfully query the server when it contains a large quantity of data sets.

1. Open Microsoft Internet Explorer on any client machine on the network
2. Go to the Olympus DISS Graphical User Interface main screen at "http://io/"
3. Select online or offline data
4. Select any method for querying the database
5. Click on the Search button
6. Wait for the results to return

7. Is the result accurate?
8. Was the result returned in a reasonable amount of time?

6.1.4 Expected Test Results

TEST A: PASS/FAIL The test will be successful if the system ingests all of the data records with minimal impact on system load and without halting or failing.

TEST B: PASS/FAIL The test will be successful if the correct results for the query are returned.

6.2 Access System Over An Intranet

6.2.1 Objective

Test system performance of the Olympus DISS client and server over a local network (*intranet*) and over a virtual private network (VPN) (A VPN requires *internet* access.)

Requirement:

- The system will have the capability to access data over an intranet, and if necessary, the internet.

6.2.2 Test Input

Any search from the Olympus DISS Graphical User Interface on a client machine that will query the database can be used as input for this test.

6.2.3 Test Procedure

1. Log on to any client computer that is not also the server, within the intranet
2. Open Microsoft Internet Explorer
3. Go to the Olympus DISS Graphical User Interface main screen via its URL "http://io/"
4. Select online or offline data
5. Select any method for querying the database
6. Click on the Search button
7. Wait for the results to return
8. Are the results accurate? Do the returned results match the parameters queried?
9. Repeat the above steps with a Virtual Private Network (VPN)

6.2.4 Expected Test Results

PASS/FAIL. The test will be successful if the client machine is able to remotely query the server successfully over a network or internet connection when it is on a machine separate from the server.

6.3 Allow Ingestion of Different Data Formats

6.3.1 Objective

Verify that the Olympus DISS™ server machine can ingest dissimilar data file formats.

Requirement:

- The system will allow accurate ingestion of different data formats (i.e., raster, vector shapefiles and coverages, and tabular data) based on the FGDC standard.

6.3.2 Test Input

File formats: ERDAS IMAGINE, ERDAS 7.5 LAN, MrSID, TIFF, JPEG, ENVI BIL, ESRI GRID, ESRI Shapefile, ESRI ArcINFO Coverage

6.3.3 Test Procedure

1. The database maintainer creates metadata files with ArcCatalog using each of the aforementioned file formats.
2. The database maintainer runs the Harvester perl script to ingest the new .xml metadata files into the database.
3. Query the database through the Olympus DISS Graphical User Interface using search terms appropriate for searching the newly ingested data.
4. Review returned results to see if the new metadata files have, in fact, been ingested.

6.3.4 Expected Test Results

PASS/FAIL. The test will be successful if each of the above mentioned file formats has been, in a repeatable manner, ingested into the database on the server.

6.4 Functional User Interface

6.4.1 Objective

The web-based graphical user interface, consisting of a login page, search page, result page, and metadata pages, will be easy to navigate, functional, and appropriate.

Requirement:

- The system will have a user-friendly interface.

6.4.2 Test Input

There is no test input.

6.4.3 Test Procedure

1. Have an individual with no prior experience using Olympus DISS, but having general knowledge of spatial data, navigate through the Graphical User Interface to find the results they desire.
2. Survey response to the experience. Was it useful, easy, helpful? Did it enhance the quality of data search? Was it functional? Did all of the links work? Was data available? Were help topics appropriate?
3. Repeat steps one and two with another individual.

6.4.4 Expected Test Results

PASS/FAIL. The test will be successful if, based on email and/or in-person responses to the beta tests, individuals are reasonably satisfied that Olympus DISS is easy to navigate, simple to understand, and overall enhances their ability to find data.