

## 8. SERVER CONFIGURATION AND TECHNICAL ISSUES

The FGDC Metadata Clearinghouse is a decentralized system of Internet servers you can use to search for available geospatial data. The PICES Metadata Federation (Fig. 8.1) has successfully connected servers in Japan, Republic of Korea, Russia, and the U.S.

An overview of the steps required to prepare metadata and make it available on an Internet server is summarized in Figure 8.2. In this section we will go through each step in detail.

### 8.1. Hardware recommendations

A typical computer (*i.e.*, 2.3 GHz Intel CPU, 512 MB RAM and 120GB HD) is all that is required to act as a server node. In the U.S. the cost is approximately \$1,200.

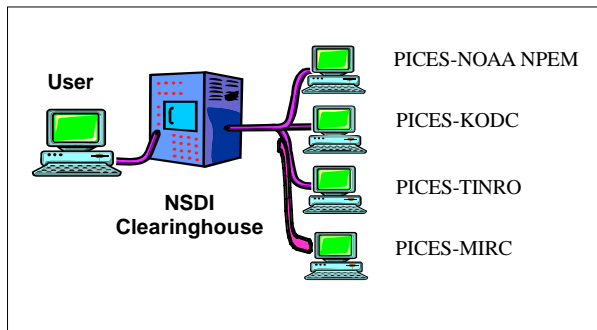


Fig. 8.1. Schematic of the PICES Metadata Federation.

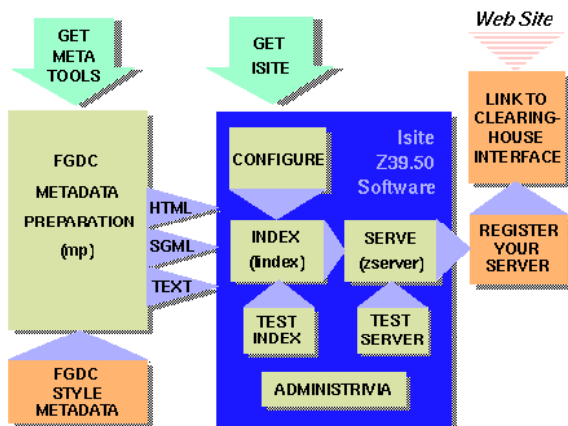


Fig. 8.2 Steps required to prepare metadata and make it available on an Internet server.

### 8.2. Operating system – Red Hat Linux Fedora

We recommend using the freely available Linux operating system by Red Hat. The Fedora Project (<http://fedoraproject.org/>) is a Red Hat sponsored and community supported open-source project. The goal of the Fedora Project is to work with the Linux community to build a complete, general-purpose operating system exclusively from free software.

Red Hat Fedora can be downloaded from <http://fedoraproject.org/wiki/Distribution/Download> or from the Red Hat Fedora Linux Operating System website found at <http://fedora.redhat.com/>.

### 8.3. Web services – Apache HTTP Server version 2.2.3

Web services such as HTTP (Hypertext Transfer Protocol), *etc.* are provided by Apache HTTP Server. Apache HTTP Server version 2.2.3 (sometimes called Apache httpd) is a project of the Apache Software foundation aimed at creating a robust, commercial-grade, feature-filled, and freely-available source code implementation of an HTTP (Web) server. For software, installation instructions and documentation, browse <http://httpd.apache.org/download.cgi>.

### 8.4. Create and validate FGDC-compliant Metadata in XML format

Your clearinghouse server (node) will need a copy of every metadata record in XML format stored in a predetermined subdirectory on the server.

You will need to decide if you want to store your metadata in a relational database or in discrete XML metadata documents. Deciding between holding your metadata in a database or producing discrete metadata documents for each data set is somewhat dependent on the variety and volume of your data sets, as well as how often they (and the metadata) are updated. If you store your metadata in a relational database, then you will need a script of some kind to extract each metadata record from

the database and prepare it as a FGDC record in XML format.

Several tools are available to create and validate FGDC metadata. Free products include:

- Tk Metadata Editor (tkme)  
<http://geology.usgs.gov/tools/metadata/tools/doc/tkme.html>
- Xt Metadata Editor (xtme)  
<http://geology.usgs.gov/tools/metadata/tools/doc/xtme.html>
- MetaScribe  
<http://www.csc.noaa.gov/metadata/metascibe/>
- NOAA's Metadata Enterprise Resource Management Aid (MERMAid)  
<http://www.ncddc.noaa.gov/Metadata/tools>
- Metavist 2005  
<http://doc.nprb.org/web/metadata/workshop/MetaVist/?C=S;O=A>.

Some capable commercial products (not free) are:

- ArcCatalog  
<http://www.esri.com/software/arcgis/about/desktop.html>
- GeoMedia Catalog  
<http://www.intergraph.com/geomediacatalog/default.asp>
- SMMS (Spatial Metadata Management System)  
<http://marinemetadata.org/tools/refs/SMMS/>.

We prefer Metavist as it is easy to use and produces compliant XML metadata records. For documentation, see the website [http://ncrs.fs.fed.us/pubs/gtr/gtr\\_nc255.pdf#search=%22Metavist%22](http://ncrs.fs.fed.us/pubs/gtr/gtr_nc255.pdf#search=%22Metavist%22).

Metadata can be checked with any of the metadata validation utilities listed below.

- Chew 'n Spit (CNS)  
<http://marinemetadata.org/tools/refs/CNS/>
- Metadata Parser (MP)  
<http://geology.usgs.gov/tools/metadata/tools/doc/mp.html>

- Enumerated Domain Helper  
<http://geology.usgs.gov/tools/metadata/tools/doc/ctc/edom.shtml>
- MP Batch Processor  
<http://support.intergraph.com/Geospatial/Downloads/Tools.asp?ID=48&SORT=Title>
- Metadata Validation Service  
<http://geo-nsdi.er.usgs.gov/validate.php>.

MP and MP Batch Processor do a good job for individual records or a library of records.

## 8.5. Obtaining, installing and configuring Isite for UNIX/Linux

### 8.5.1. Download the appropriate Isite software

Download the appropriate Isite software by going to <http://clearinghouse4.fgdc.gov/ftp/>. Figure. 8.3 shows an image of the Isite web site listing versions of Isite for different operating systems.

For this example, we will be using the Linux 2.6.8

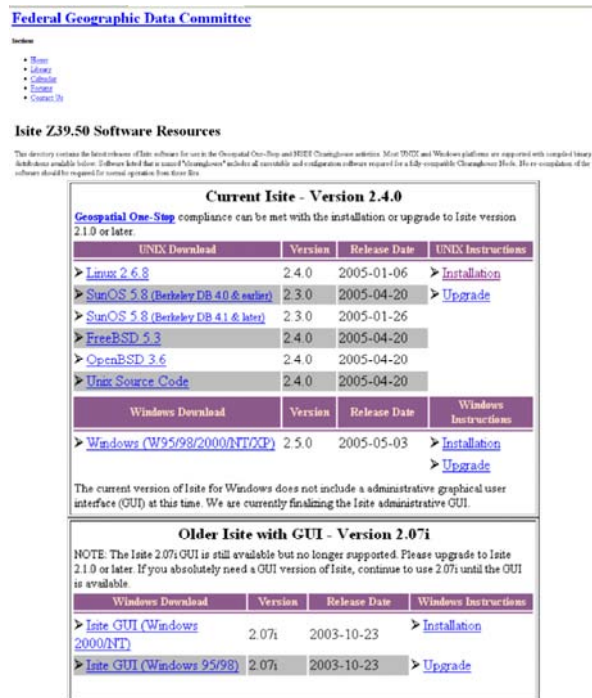


Fig. 8.3 Image of the Isite web site showing versions of Isite for different operating systems.

version of Isite. We will also be using GNU tar (available from the server at <http://www.gnu.org/software/tar/tar.html>) to extract it.

**Conventions:** Commands should be entered all on one line unless specified otherwise. We will also assume that your database name will be in **bold**, and the string you are searching for will be in **bold italics**. The \$ at the beginning of each command is the Linux shell prompt.

### 8.5.2. *Untar the file*

For this example, we will be using the /opt directory to untar Isite using the command

```
$ tar xzvf clearinghouse2-  
v2.3_Linux_2.6.8-24.10-i386-  
2005010602-static.tar.gz
```

This should will generate an Isite2 directory and output something like the following:

```
/opt/Isite2/bin/GetScript  
/opt/Isite2/bin/Key1  
/opt/Isite2/bin/SearchScript  
/opt/Isite2/bin/Iindex2  
/opt/Isite2/bin/Isearch2  
/opt/Isite2/bin/Iutil2  
/opt/Isite2/bin/Iget2  
...
```

Isite2 contains two directories of documentation and software

The /opt/Isite2/bin subdirectory contains the following executables:

- Iindex2 – indexes metadata entries
- Isearch2 – provides local search test
- zclient – provides command-line query of remote or local Z39.50 service
- Izclient – interactive Z39.50 client using ANSI terminal node
- zserver – the Z39.50 server process
- zping – checks the availability of a given Z39.50 service (local or remote)

In the /opt/Isite2/bin subdirectory, there are two initialization files, the Search Application Program Interface initialization file (sapi.ini) and the zserver process initialization file (zserver.ini). The sapi.ini file contains specific configuration information for your indexes. The zserver.ini file

specifies the port number to be used to connect to the Internet, and the DBList parameter lists the names of the databases described in further detail in the sapi.ini file.

Images of the server directory structure and a listing of files contained in each subdirectory are shown in Appendix 15.5.

### 8.5.3. *Name your database*

Pick a name for your database, preferably one without spaces in it. It should be a unique name (not, for example, FGDC) that is indicative of your data content. It *does* matter how you capitalize this name, so be consistent.

Examples: **arizona**, **FrameworkData**, **ABCCounty**. For this example, we will use our DB name of **NPEM**.

### 8.5.4. *The zserver.ini file*

Open and edit the file zserver.ini. Confirm that the following parameters are set as they are below:

```
ServerType=STANDALONE  
Port=6668  
SAPI=sapi.ini  
AccessLog=zserver_access.log  
DBList=NPEM
```

Note that the port number is the communication link with the Internet. If you have a firewall installed then this port must be explicitly open to the Internet in order for the Clearinghouse node to communicate with your Isite server.

You will need to create a location to store your database. For this example we created the directory /opt/Isite2/index.

A listing of the zserver.ini file used for the NPEM node can be found in Appendix 15.6.

### 8.5.5. *The sapi.ini file*

Edit the sapi.ini file to look like the following structure:

```
[DEFAULT]  
DBList=NPEM
```

```
[NPEM]
type=ISEARCH
location=/opt/Isite2/index
FieldMaps=/opt/Isite2/bin/bib1_fgdc.m
ap,\
  /opt/Isite2/bin/gils_fgdc.map,\
  /opt/Isite2/bin/geo_fgdc.map
```

In this example, the database is called “**NPEM**”, and this must match the `DBList` parameter in `zserver.ini`. The database is stored in the `/opt/Isite2/index` directory. The type is `ISEARCH`.

Verify that Fieldmaps are in the right location. The pathnames to Fieldmaps *must* be set to the location where you just installed Isite on your system. Be sure that you have no spaces in your Fieldmaps declaration line. Verify that these files are in the right locations by typing the following commands:

```
$ file /opt/Isite2/index
$ file /opt/Isite2/bin/bib1_fgdc.map
$ file /opt/Isite2/bin/gils_fgdc.map
$ file /opt/Isite2/bin/geo_fgdc.map
```

Each command should return either “directory” or “ASCII text”. If it returns “(No such file or directory)”, check your pathnames.

A listing of the `sapi.ini` file used for the NPEM node can be found in Appendix 15.6.

### 8.5.6. Indexing metadata

To index all metadata files, first create the directory `/opt/Isite2/index`. Then go to the `/opt/Isite2/bin` directory and use the `Iindex2` command with the following arguments from the `Isite2/bin` directory. The lines below are all one command:

```
$ /opt/Isite2/bin/Iindex2 -d
/opt/Isite2/index/NPEM -t fgdc -o
fieldtype=/opt/Isite2/bin/fgdc.fields
/opt/Isite2/data/*.xml
```

**Note:** For this example we will assume your metadata is located in `/opt/Isite2/data`. We have saved the indexing commands above in a script file, `DoIndex`, which we placed in the `/opt/Isite2/bin/` subdirectory.

We execute the script from the `/opt/Isite2/bin/directory` with the command

```
$ sh DoIndex
```

The database **NPEM** will be written to the `/opt/Isite2/index` directory, using FGDC metadata format, using FGDC search fields, and using as input all files in the `/opt/Isite2/data` directory that end in `*.xml`. You will need to adapt these pathnames to match your system.

You can find advanced indexing options by browsing:

<http://clearinghouse4.fgdc.gov/ftp/advancedindexing.html>.

If all is successful, you should receive a message similar to the following:

```
Index v2.2
Building document list ...
Creating new cache
[/opt/Isite2/index/NPEM.db]
Building database
/opt/Isite2/index/NPEM:
Parsing files ...
  Parsing
/opt/Isite2/data/npem993.xml
key=276712911580
```

(This line will repeat for every XML file that is indexed.)

```
Indexing 21624 words ...
Writing compressed blocks to sequence
table 1
.....
Wrote 36 blocks of compressed
strings.
Done with sequence table 1
Dumping sparse table 1...
Done with sparse table 1
Database files saved to disk.
```

It is recommended that you save this command line as a script that can be run over and over again and edited if changes are warranted.

### 8.5.7. Start the Clearinghouse Server process

Start your Z39.50 Clearinghouse Server process by changing directory to `Isite2/bin` and invoking the `zserver` command. You may wish to start the process in the background with the Linux command:

```
$ ./zserver &
```

You could also consider including it in startup scripts typically stored in the `/etc/rc.d` directory so it will get restarted after the system reboots. Otherwise, the `zserver` process will have to be manually started after every reboot.

### 8.5.8. Testing the `zserver` installation

Once you have the `zserver` running (on port 6668 and with a database named “**NPEM**”), you can test it by confirming that it is running by executing the following command:

```
$ ./zping localhost 6668
```

You should get the response:

```
Z39.50 server at localhost on port
6668 is alive.
```

The following title search for the string “*data*” (or some other word you know will be in title records within your database) should return some hits:

```
$ ./zclient localhost 6668 "NPEM"
"data[1,4]"
```

Confirm that a full text search for the string *data* (or some other word you know will be in your database) returns some hits:

```
$ ./zclient localhost 6668 "NPEM"
"data [1,1016]"
```

Finally a global spatial search should return every record in the database:

```
$ ./zclient localhost 6668 "NPEM" "90
-180 -90 180[1,2060]"
```

## 8.6. Registering with NSDI Clearinghouse

Registering your node is a five-step process. Screen-shots of each step are given below.

For on-line documentation, see:

- Guide to Clearinghouse Node Creation [http://www.fgdc.gov/dataandservices/clearinghouse\\_qanda](http://www.fgdc.gov/dataandservices/clearinghouse_qanda),
- Install and configure Isite <http://clearinghouse4.fgdc.gov/ftp/unixinstall1.html>.

To begin the registration process, go to <http://registry.gsdi.org/>, and click on the “Add Your Node” icon (see Fig. 8.4).

For the first step, supply either a valid DNS name or an IP address (Fig. 8.5). Click on the Next-> Perform Server Check button.

Next, a test is performed to validate your server name and compatibility. Enter the communication port number for your Z39.50 server (Fig. 8.6). Click on the Next-> Perform Z39.50 Check button.



**Fig. 8.4** Entry point for registering your node with the Clearinghouse.

[Add Your Node](#)
[Browse Registry](#)
[Browse Diagnostics](#)
[Modify Your Node](#)
[International Server Status](#)
[Domestic Server Status](#)
[Canadian Server Status](#)
[Home](#)

# THE CLEARINGHOUSE REGISTRY

Federal Geographic Data Committee (FGDC)

[Add Your Node Tutorial](#)

## Step 1 of 5: Add Your Node to the Clearinghouse (Server Check)

*This form is used to collect nominations of Z39.50 Version 2 or 3 servers which support the GEO profile. Entries made via this form are made accessible to the Clearinghouse Search Forms in HTML and Java.*

Enter the Host Name of your Z39.50 server:  
 (Example: registry.gsdi.org) Host Name:

If you do not have a Host Name, select this option and enter your Machines IP Address in the Host Name field.

No Host Name:

**Fig. 8.5** Step 1: Enter the name or IP address of the Z39.50 server.

[Add Your Node](#)
[Browse Registry](#)
[Browse Diagnostics](#)
[Modify Your Node](#)
[International Server Status](#)
[Domestic Server Status](#)
[Canadian Server Status](#)
[Home](#)

# THE CLEARINGHOUSE REGISTRY

Federal Geographic Data Committee (FGDC)

[Add Your Node Tutorial](#)

## Step 2 of 5: Add Your Node to the Clearinghouse (Server Check)

*This form is used to collect nominations of Z39.50 Version 2 or 3 servers which support the GEO profile. Entries made via this form are made accessible to the Clearinghouse Search Forms in HTML and Java.*

**Status:** Your server at [codline.afsc.noaa.gov](http://codline.afsc.noaa.gov) has been located and tested via the internet.  
 Please continue the clearinghouse registration by providing the following information.

Enter the Port Number of your z39.50 server  
 (example: 210)

Port Number:

**Fig. 8.6** Step 2: Submit server port number.

[Add Your Node](#)
[Browse Registry](#)
[Browse Diagnostics](#)
[Modify Your Node](#)
[International Server Status](#)
[Domestic Server Status](#)
[Canadian Server Status](#)
[Home](#)

# THE CLEARINGHOUSE REGISTRY

Federal Geographic Data Committee (FGDC)

[Add Your Node Tutorial](#)

## Step 3 of 5: Add Your Node to the Clearinghouse (Search Test)

*This form is used to collect nominations of Z39.50 Version 2 or 3 servers which support the GEO profile. Entries made via this form are made accessible to the Clearinghouse Search Forms in HTML and Java.*

**Status:** Your server at [codline.afsc.noaa.gov](http://codline.afsc.noaa.gov) and Port 6668 has been successfully tested. Please continue the clearinghouse registration by providing the following information.

Enter the Database Name of the z39.50 server  
 Database Name:

Enter a valid word contained within the title of one of the metadata records  
 Title Search:

Enter a valid word contained within the full text of one of the metadata records  
 Full Text Search:

Perform Spatial Search Test:

Un-check this box ONLY if your collection has ZERO records with bounding coordinates in them. Unchecking this box will skip a spatial test on this collection.

**Fig. 8.7** Step 3: Test search function.

Now add the name of your database and two search words, as shown in Fig. 8.7. Click on the Next->Perform Z39.50 Search Tests button.

Inspect the status report (Fig. 8.8) of the search test made using your database name and your two search phrases or words. Click on Next-> Continue Registration.

For the final step, supply information that describes your server (Figs. 8.9–8.13). Please fill out all fields. Add “PICES” to the beginning of the title of the server. In this way all PICES servers will be seen at the same time in the alphabetized list of active servers connected to the Clearinghouse.

To continue, click on the Submit Node Entry to Registry button and you will be asked to enter a password for future modifications or updates.

That completes the registration process. Optionally, you can register your node with Geodata.gov by clicking on the Modify Your Node icon at the top of the screen, entering your node name, contact name and password (Fig. 8.13) and clicking on the Register on Geodata.gov button.

Status of nodes can be determined by browsing <http://registry.fgdc.gov/serverstatus/> (Fig. 8.14).

The Clearinghouse Registry  
Federal Geographic Data Committee (FGDC)

[Add Your Node](#) [Browse Registry](#) [Browse Diagnostics](#) [Modify Your Node](#) [International Server Status](#) [Domestic Server Status](#) [Canadian Server Status](#) [Home](#)

[Add Your Node Tutorial](#)

### Step 4 of 5: Add Your Node to the Clearinghouse (Test Results)

*This form is used to collect nominations of Z39.50 Version 2 or 3 servers which support the GEO profile. Entries made via this form are made accessible to the Clearinghouse Search Forms in HTML and Java.*

**Status:** Your z39.50 server at [codline.afsc.noaa.gov](http://codline.afsc.noaa.gov) on Port [6668](#) with the [NPEM](#) has successfully passed the required clearinghouse search tests with the following results:

- ◆ **fish** Title Search: [74](#) Metadata Record(s) Found.
- ◆ **data** Full Text Search: [2775](#) Metadata Record(s) Found.
- ◆ Geospatial Coordinate Search: [2775](#) Metadata Record(s) Found.
- ◆ Implementation Name: [CNIDR zserver](#)
- ◆ Implementation Version: [2.2.4-SUSE92-Linux, Release 2005042001](#)
- ◆ GEO Value: [GEO22](#)

Please continue the clearinghouse registration by continuing to the next page.

[Next-> Continue Registration](#)

**Fig. 8.8** Step 4: Status report.

[Add Your Node](#)
[Browse Registry](#)
[Browse Diagnostics](#)
[Modify Your Node](#)
[International Server Status](#)
[Domestic Server Status](#)
[Canadian Server Status](#)
[Home](#)

# THE CLEARINGHOUSE REGISTRY

Federal Geographic Data Committee (FGDC)

## Step 5 of 5: Add Your Node to the Clearinghouse (Registration)

*This form is used to collect the remaining required information for Clearinghouse registration. Entries made via this system are made accessible to the Clearinghouse Search Forms in HTML and Java.*

### Server Description

- ◆ Long Title of Server: (to be used on select list)
- ◆ Short Title of server:  (to be displayed on selection map)
- ◆ Abstract: (Describe the collection criteria, data coverage, themes, and formats of information served by this 239.50 service. This field may be used by clients to evaluate the types of information and coverage to perform a selective query or to evaluate query results.)

**Fig. 8.9** Step 5a: Supply title of server (long and short versions) and provide an abstract that describes the service.

◆ Cost information: (Enter cost basis description)

◆ Server Categories:

Agriculture and Farming:	<input type="checkbox"/>	Atmospheric and Climatic Data:	<input checked="" type="checkbox"/>
Biologic and Ecologic Information:	<input checked="" type="checkbox"/>	Environmental Monitoring and Modeling:	<input checked="" type="checkbox"/>
Business and Economic Information:	<input type="checkbox"/>	Administrative and Political Boundaries:	<input type="checkbox"/>
Earth Surface Characteristics and Land Cover:	<input checked="" type="checkbox"/>	Elevation and Derived Products:	<input type="checkbox"/>
Geologic and Geophysical Information:	<input checked="" type="checkbox"/>	Health and Disease:	<input type="checkbox"/>
Ocean and Estuarine Resources and Characteristics:	<input checked="" type="checkbox"/>	Cultural and Demographic Information:	<input type="checkbox"/>
Images and Photographs:	<input type="checkbox"/>	Fresh Water Resources and Characteristics:	<input type="checkbox"/>
Cadastral and Legal Land Descriptions:	<input type="checkbox"/>	Utility Distribution Networks:	<input type="checkbox"/>
Facilities, Buildings and Structures:	<input type="checkbox"/>	Geodetic Networks and Control Points:	<input type="checkbox"/>
Transportation Networks and Models:	<input type="checkbox"/>	Base Maps, Scanned Maps and Charts:	<input type="checkbox"/>
Tourism and Recreation:	<input type="checkbox"/>		

◆ This Clearinghouse Node contains metadata that provide the following access to data through the Online Linkage field through a WebURL:

Data download via ftp:	<input type="checkbox"/>
Data download via http:	<input checked="" type="checkbox"/>
Data ordering via http form:	<input type="checkbox"/>
Data information via http form:	<input type="checkbox"/>
Data information via offline access:	<input checked="" type="checkbox"/>
OpenGIS Web Mapping Service:	<input type="checkbox"/>
OpenGIS Web Feature Service:	<input type="checkbox"/>

**Fig. 8.10** Step 5b: Provide cost basis, server categories and methods you use to link metadata to data.

## Server Host Information

*This block of information is required to enroll your server in the list of GEO servers. The hostname, port, and database name information is used to automatically generate select lists for the Java and HTML query interfaces.*

- Full Internet Host Name: **codline.afsc.noaa.gov**
- Full Internet IP Number: **161.55.120.200**
- TCP/IP Port Number for Z39.50 Service: **6668**
- Data index name (Z39.50 Database Name): **NPEM**
- Valid word in a database record title: **fish**
- Valid word found in a database record text: **data**
- Perform Spatial Search: **Yes**
- GEO Value: **GEO22**
- Software Implementation: **CNIDR zserver**
- Software Version: **2.2.4-SUSE92-Linux, Release 2005042001**
  
- Platform:
- Full WebURL of related Web site:
- Server latitude in Decimal Degrees (DD.D):
- Server longitude in Decimal Degrees (-DDD.D):
- What Country best describes the significant coverage of your database? Note: If your database has significant coverage for multiple countries you may need to select either a continent or global selection from the pulldown list:
- Data Coverage:
- Does this collection include data that covers the United States in part or full?  Yes
- Collection Scope ([What is this?](#))

## Predominant Geographic Extent of Data Served

*This rectangular footprint defines the general area for which you manage geographic metadata and data. This information may be used in a preemptive query to identify suitable servers from a long list of servers.*

**min longitude**  **max longitude**   
**min latitude**  **Max latitude**

**Fig. 8.11** Step 5c: Provide server host information and geographic coverage summary.

## Server Contact Information

Contact information to permit clients and servers to contact you with service problems or queries.

- Name:
- Organization:
- Address:
- City:
- State or Province:
- Postal Code:
- Country:
- E-mail address:
- User Support Hours of Service:
- Telephone:
- Fax:

**Fig. 8.12** Step 5d: Provide contact information.

# THE CLEARINGHOUSE REGISTRY

Federal Geographic Data Committee (FGDC)

[Modify Your Node Tutorial](#)

## Modify the Registry

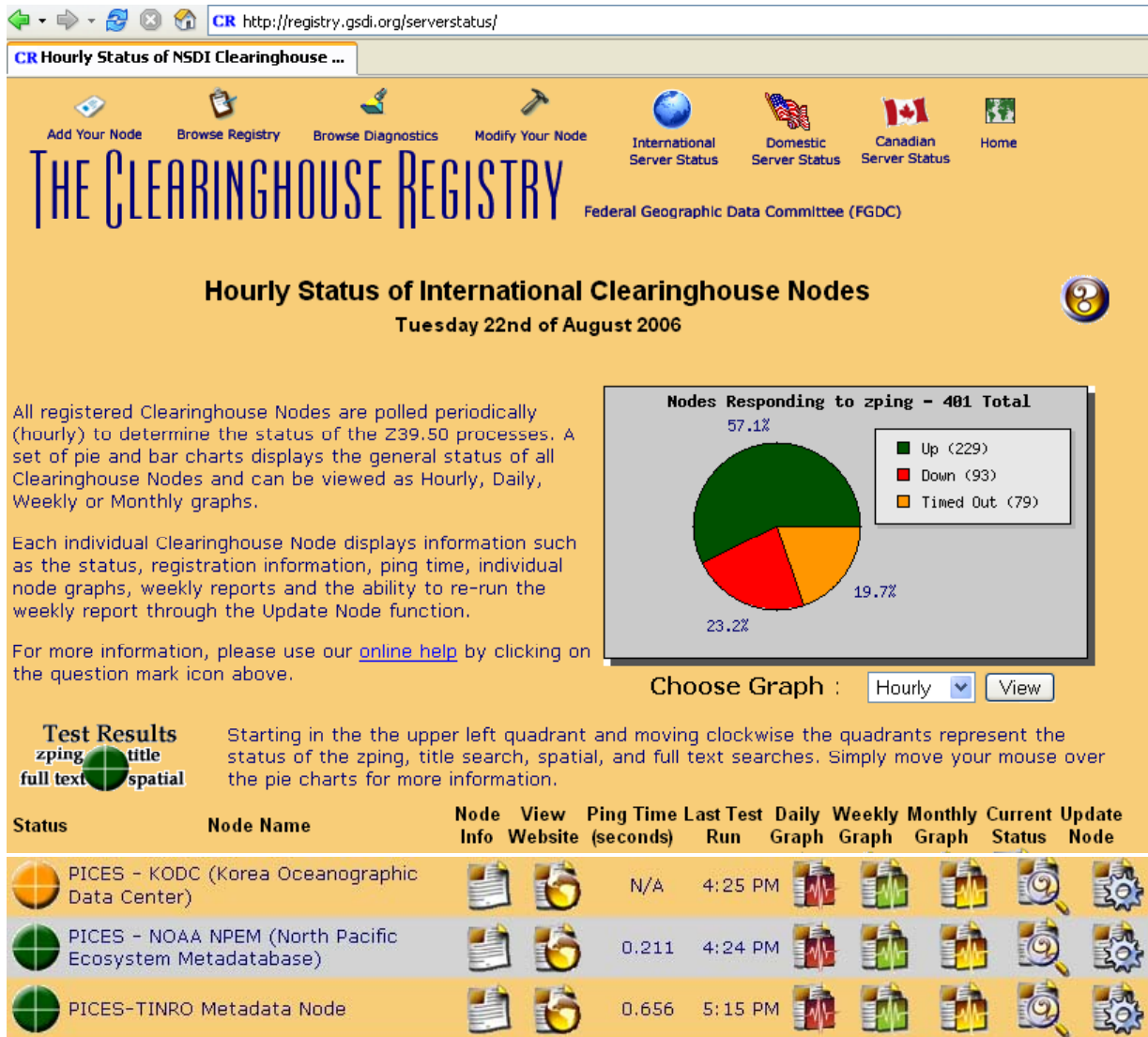
Select the name of the node, contact name and password for the node that you wish to modify. If you provide the correct information, you will be able to modify the node registration information.

Node:

Contact:

Password:

**Fig. 8.13** Step 5e: Optional registration at Geodata.gov.



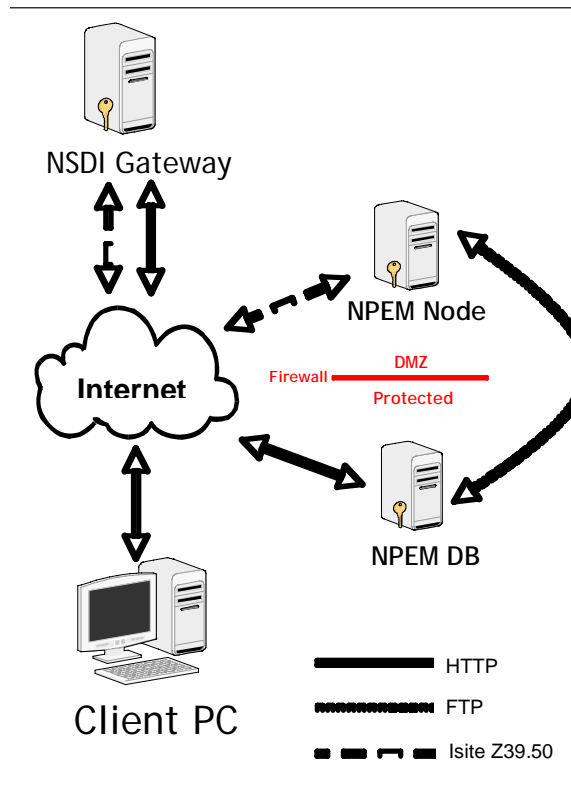
**Fig. 8.14** Hourly status of international clearinghouse nodes (PICES nodes shown).

## 8.7. Security issues

The diagram in Fig. 8.15 shows how we have arranged our computer hardware for the PICES–NPEM node. Recall that the communication port on the NPEM node Isite server has to be open to the Internet. This presents a serious security problem. Our solution was to open the port (6668) on the firewall, but place the NPEM node server outside our firewall but inside the DMZ (Demilitarized zone), still under control of the firewall router.

Our metadata reside in a relational database on a server inside the firewall. Records from the metadata relational database are read, one-at-a-time, and converted into an XML record with a

Java script (listed in Appendix 15.7). At the end of this process, there is an XML file for each record in the metadata base. These XML files are transferred to the NPEM node server via FTP as updates as additions are made. After each update, the XML files are re-indexed on the NPEM node. The NPEM node and the NSDI Clearinghouse Gateway communicate via Z39.50 protocols. Users can access the NPEM metadata base server through our custom search interface <http://www.pmel.noaa.gov/np/mdb/> using HTTP or they can search the same metadata records through the NSDI clearinghouse Gateway along with other shared Gateway holdings via HTTP and Z39.50.



**Fig. 8.15** Arrangement of computer hardware for a secure PICES–NPEM node.