

RESEARCH METADATA ON THE WEB: SELECTED GEOSPATIAL DATA AND METADATA DIRECTORIES

By Stephanie Haas
Marston Science Library
University of Florida
Gainesville, FL 32611-7011

ABSTRACT: This paper will present an overview of the types of research metadata that are available over the Web. It will include a discussion of the electronic tools which are being used to create metadata online and the directories which are being created at the national, federal, and state levels.

The Value of Metadata

Metadata allow researchers to identify potentially useful data sets collected by others. Gary Waggoner, Biological Research Division, U.S. Geological Survey, states "Metadata refers to data that are used to describe a database (e.g., describing the extent of the data, coverage, scale, what methods were used to collect the data, by whom and when the data was collected, etc.) With valid and complete metadata, someone can learn enough about a database (without communicating with the "owner" of the data) to determine if the data would be of use or interest to them."

Metadata's Multiple Meanings

There are several standards for metadata. The situation is aptly described in the introduction to NOAA's Environmental Services Data Directory: "Environmental data is often stored in vastly different formats. Over the years, there have been many competing standards. These include: the Federal Geographic Data Committee (FGDC) Metadata standard, the paragraph formats used in the NEDRES databases, the Government Information Locator System (GILS), and the Directory Interchange Format (DIF). Each format has different field names and a different look and feel.

These competing standards generally make database searches across multiple databases in multiple formats impossible. However, all of the formats contain certain common fields, including geographic location, title, summary, and data center." ("How Can Users Search the NOAA Environmental Services Data Directory?" NOAA Environmental Data Directory Overview. URL: <http://www.esdim.noaa.gov/NOAA-Catalog/Overview.html> Updated on December 5, 1996)

GEOSPATIAL METADATA

Geospatial metadata refers to geographically-referenced data sets that are used in geographic information systems (GIS). In the U.S., the impetus for establishing standards

and directories for this type of metadata came from the federal government. Many agencies, e.g., U.S. Geological Survey, Environmental Protection Agency, U.S. Department of Agriculture, share a common base of spatial data needs upon which their unique projects are overlaid. The federal government mandated the development of metadata standards and directories to reduce the costly duplication of geospatial data collection. The Federal Geographic Data Committee (FGDC) was established through the Office of Management and Budget (OMB) Circular A-16 "...to promote the coordinated development, use, sharing, and dissemination of surveying, mapping, and related spatial data..." Major responsibilities of FGDC were developing a national database of spatial data, developing and implementing data standards, promoting technology developments, and promoting interagency cooperation in spatial data activities. The FGDC was also charged with promoting cooperation among federal, state, and local agencies and the private sector in spatial data collection, production and sharing.

Products resulting from the FGDC include:

Federal Geographic Data Committee Homepage

<http://www.fgdc.gov/index.html>

Content Standard for Digital Geospatial Metadata

<http://www.fgdc.gov/Metadata/ContStan.html>

Geospatial Data Clearinghouse

<http://www.fgdc.gov/Clearinghouse/Clearinghouse.html>

The Geospatial Data Clearinghouse is a collection of over 50 spatial data servers, primarily in North America, that have digital geographic data primarily for use in Geographic Information Systems (GIS) image processing systems, or other modeling software. These data collections can be searched through a single interface based on their descriptions or "metadata."

With mixed results, the efforts of the Federal Geographic Data Committee have been adopted and/or modified by federal, state, and international organizations.

SELECTED U.S. FEDERAL GEOSPATIAL SITES

Environmental Protection Agency sites:

EPA Scientific Metadata Standards Project

<http://www.lbl.gov/~olken/epa.html>

EPA Spatial Data Library System (ESDLS)

http://www.epa.gov/enviro/html/esdls/esdls_over.html

Data sets are contained in ESDLS for county, state, and national levels. These data are available at a scale of 1:100,000 (county), 1:250,000 (state), and 1:2,000,000 (state and national). The geographic extent for ESDLS data covers the conterminous United States, Alaska, Hawaii, Puerto Rico, and the Virgin Islands.

EPA Databases and Software

<http://www.epa.gov/epahome/Data.html>

Overview of the facets of EPA's data sites, metadata tools, etc.

EPA's EMAP site (Environmental Monitoring and Assessment Program)

<http://www.epa.gov/emap/>

This site contains data, metadata, and information about the program.

EPA Geospatial Data Clearinghouse

<http://nsdi.epa.gov/nsdi/>

The metadata in this clearinghouse provide the spatial data used by Maps On Demand.

GILS (Government Information Locator System)

http://www.access.gpo.gov/su_docs/gils/gils.html

The Government Information Locator Service (GILS) is a networked-based approach to assist users in locating government information resources. The U.S. Federal implementation of GILS began in December 1994. Records are searchable through the GPO Access site listed above.

National Environmental Data Index

<http://www.nedi.gov/>

NEDI currently provides Full-text Search of the environmental information compiled by several agencies of the federal government including Agriculture; Commerce; NOAA; National Institute of Standards and Technology; Census; Patent and Trademark Office; Defense; Master Environmental Library: Air Force Combat Climatology Center (AFCCC), Air Force Global Weather Center (AFGWC), Center for Air/Sea Technology, Stennis Space Center (CAST, SSC), Coastal and Hydraulics Laboratory (CHL), National Geophysical Data Center (NGDC), Naval Research Laboratory (NRL), Simulator Data Base Facility (SDBF), National Imagery and Mapping Agency (NIMA); Defense Environmental Network and Information Exchange (DENIX); Energy; Interior; USGS including the Biological Resources Division; Health and Human Services; EPA; NASA, National Science Foundation; and Library of Congress (text of bills).

National Aeronautic and Space Administration (NOAA) sites:

Global Change Master Directory

<http://gcmd.gsfc.nasa.gov/>

NASA's Global Change Master Directory (GCMD) is a comprehensive directory of descriptions of data sets of relevance to global change research. The GCMD database includes descriptions of data sets covering climate change, the biosphere, hydrosphere & oceans, geology, geography, and human dimensions of global change.

Earth Observing System Data and Information System

http://spsosun.gsfc.nasa.gov/New_EOSDIS.html

Earth Observing System (EOS) Data and Information System (EOSDIS) is a comprehensive data and information system developed by NASA under the Earth Science Enterprise (ESE) Program.

NOAA Central Library Historical Data Sets

<http://www.esdim.noaa.gov>

NOAA Coastal Services Center

<http://www.csc.noaa.gov/>

NOAA Environmental Services Data Directory

<http://www.esdim.noaa.gov/NOAA-Catalog/index.html>

From this directory, you can choose to search any or all of the data sets from the National Data Centers, other NOAA sites, and the two non-NOAA sites listed below:

NOAA National Climatic Data Center
NOAA National Geophysical Data Center
NOAA National Oceanographic Data Center
NOAA National Snow and Ice Data Center

NOAA National Marine Fisheries Service
Historical Data Sets from Central NOAA Libraries
Historical DATA Sets from NCDC Libraries
All other NOAA Sites

NOAA US-Japan GOIN (Global Observing and Information Network) Node
National Environmental Data Referral Service

U.S. Geological Survey sites:

USGS Geospatial Data Clearinghouse

<http://water.usgs.gov/nsdi/>

U.S. Geological Survey Geoscience data

<http://geo-nsdi.er.usgs.gov/>

U.S. Geological Survey Mapping & Remotely Sensed Data

<http://mapping.usgs.gov/nsdi/>

Includes links to the National Atlas of the United States (<http://www-atlas.usgs.gov/atlasvue.html>)

U.S. Geological Survey Water Resources Spatial Data

<http://water.usgs.gov/public/GIS/>

U.S. Geological Survey EROS Data Center

<http://edcwww.cr.usgs.gov/dsprod/prod.html>

Earth Resources Observation Systems (EROS) Data Center, or the EDC houses millions of aerial photographs of the United States, as well as images from several series of satellites covering the entire Earth. EDC also is host to the sales data base for digital products and maps of the USGS's National Mapping Division, of which it is part. Specific products can be searched through the Global Land Information System (GLIS) (see "Specific USGS projects" below)

The **Biological Research Division**, U.S. Geological Survey created its own metadata standards because taxonomic data, collection data, and other biological information that were not implicitly geospatial were not adequately addressed by the FGDC. Metadata tools created as part of the National Biological Information Infrastructure include:

NBII Metamaker Version 2.1 template for entering metadata

(http://www.emtc.nbs.gov/http_data/emtc_spatial/applications/nbiimker.html),

NBII Metadata Clearinghouse

(http://www.emtc.usgs.gov/http_data/meta_isite/nbiigateway.html),

Integrated Taxonomic System (IT IS) for biological nomenclature

([http://www.it.is.usda.gov/it is/](http://www.it.is.usda.gov/it%20is/))

Taxonomic Resources and Expertise Directory

(<http://www.nbio.gov/tred/>)

Specific USGS projects:

Global Land Information System (GLIS)

<http://edcwww.cr.usgs.gov/glis/glis.html>

The Global Land Information System (GLIS) is an interactive computer system developed by the U.S. Geological Survey (USGS) for scientists seeking sources of information about the Earth's land surfaces. Metadata descriptions are included for digital cartographic data, digital climate data, digital geologic data, digital hydrologic data, digital landuse/landcover data, digital satellite and aerial data, digital soil data, photographs, and printed maps.

Arctic Environmental Data Directory

<http://www-ak.wr.usgs.gov/aedd/aedd.html>

The U.S. Geological Survey maintains the Arctic Environmental Data Directory (AEDD) in Anchorage, Alaska, on behalf of the member agencies of the Interagency Arctic Research Policy Committee (IARPC). The IARPC includes the National Science Foundation, the Departments of Commerce, Defense, Energy, Health and Human Services, Interior, State and Transportation, the Environmental Protection Agency, the National Aeronautics and Space Administration, the Smithsonian Institution, the Office of Management and Budget, and the Office of Science and Technology Policy. It contains descriptions of data on global change studies, environmental interactions, earth sciences, social sciences, and policy and management. Contact information is provided in each entry to obtain the data.

INTERNATIONAL METADATA SITES

African Data Dissemination Service

<http://edcintl.cr.usgs.gov/adds/adds.html>

Supported by US AID, this site allows you to download and view data for the African continent as a whole, regions of Africa, and individual countries in Africa. Data are predominantly geospatial, but agriculturally related data sets for crop production, rainfall, etc. are also available.

Antarctic Master Directory (AMD)

<http://scar.org/subsidiary/jcadm/amd/>

The Global Change Master Directory (GCMD) group is working closely with the AMD. This site will be a directory for all Antarctic data sets and the GCMD will also receive copies of their descriptions in the Directory Interchange Format (DIF).

Biodiversity Conservation Information System (BCIS) Metadatabase

<http://www.biodiversity.org/metadatabase.html>

In September 1997 the World Conservation Monitoring Centre (WCMC), based in Cambridge, UK, was commissioned by the BCIS consortium to develop a metadatabase for the consortium members. The BCIS is linked to the home page of the Clearing-House Mechanism (CHM) of the Convention on Biological Diversity (CBD) <http://www.biodiv.org/chm.html>, which currently contains full text articles and discussions, related to key thematic areas of interest to the CBD. A searchable URL Database on Scientific and Technical Cooperation in Biodiversity is available at <http://www.biodiv.org/chm.html>

Canadian Geospatial Data Infrastructure

<http://cgdi.gc.ca/>

The Canadian Geospatial Data Infrastructure (CGDI) is an initiative by the Inter-Agency Committee on Geomatics (IACG) and the Canadian Council on Geomatics (CCOG) to assemble the many governmental and commercial interests related to the production, application and dissemination of geospatial information. Related to this site are the National Atlas of Canada (<http://www-nais.ccm.emr.ca/english/home-english.html>) which provides base and thematic maps online, and an interactive map maker (<http://www-nais.CCRS.NRCAN.GC.CA/schoolnet/issuemap/Home.html>).

ERIN (Australian Environmental Resources Information Network)

<http://www.erin.gov.au/database/db.html>

Maintained by the Department of the Environment, this site provides public access to Data Directories, Environmental Regions, Biological Data, Map Facilities, and Textual / Bibliographic Databases.

The European Catalogue of Data Sources (CDS)

(<http://www.mu.niedersachsen.de/system/cds/>) was established in order to provide a locator system on environmental information in Europe and to support the operation of the European Environment Agency - Environmental Information and Observation Network (EEA-EIONET). It supplies information on who has what information in Europe, in which form and where the data exist as well as how to get access to them. CDS provides meta-information to the users of environmental information and data.

Meta-information is organized in two categories: addresses and data sources. Addresses are hierarchically structured into organizations and persons. Data sources are assigned to six classes. Index terms are taken from the General European Multilingual Environmental Thesaurus (GEMET) (<http://www.eea.dk/Locate/GEMET/default.htm>)

Consortium for International Earth Science Information Network (CIESIN)

<http://www.gateway.ciesin.org/gils/>

“The CIESIN GILS Access System is a prototype catalog service providing access to records that describe key information resources and data access systems associated with the Consortium for International Earth Science Information Network (CIESIN) and CIESIN’s Information Cooperative partner organizations. Metadata records are structured according to GILS (Application Profile Version 2.0) content standard and are searched and retrieved by one or more user-selected metadata attributes. Metadata records are retrieved and displayed using familiar GILS attribute names.”

Socioeconomic Data and Applications Center (SEDAC)

(<http://www.gateway.ciesin.org/cgi-bin/zgate>) is a node of the CIESIN Information Network. Data available through SEDAC is focused in the following categories: population dynamics; social and political structures and institutions; human and environmental health; economic activity; human attitudes, preferences, and behavior; land use; agriculture; and industry.

CEOS Information Locator Service (CILS), Committee on Earth Observation Satellites (CEOS)

<http://cils.dlr.de/what.pl/EOEData>

CEOS, the Committee on Earth Observation Satellites, was created in 1984 upon the initiative of the Economic Summit of Industrialized Nations. In 1995 the need for an electronic information service for remote sensing earth observation data was recognized. DARA (the German space agency) was asked to initiate a demonstration pilot of the CEOS Information Locator System (CILS) which was to be closely related to already existing CEOS information systems and to major European developments in information sharing. Special consideration was to be given to the needs and capabilities of developing countries.

Environment Data Directory (EDD), Australia

<http://www.environment.gov.au/net/edd.html>

“Australia’s Environment Data Directory stores information about data within the Commonwealth Department of the Environment, Sport and Territories. It includes a controlled facility for on-line populating of metadata records. It also supports spatial and temporal searching using a metadata guideline called ANZLIC (Australia, New Zealand Land Information Council). ANZLIC is quite similar to the metadata standard of the U.S. Federal Geographic Data Committee (FGDC), which is fully aligned with GILS.”

GDDD - Geographical Data Description Directory, MEGRIN

<http://www.megrin.org/gddd/gddd.html>

MEGRIN is an organization representing and owned by 19 National Mapping Agencies (NMAs) which aims to meet the needs of the pan-European digital map data market. It is an initiative of CERCO, the forum for heads of European NMAs. MEGRIN’s GDDD has datasets from more than 36 National Mapping Agencies (NMAs).

Global Environmental Information Locator Service (GELOS)

<http://ceo.gelos.org/>

GELOS is part of the G7 Environment and Natural Resource Management project (G7-ENRM). Its main objective is to create a global virtual distributed library of ENRM data and resources. While data is to be included, as of 31 July 1998, the resource types indexed included: company profile, document, database system, on-line service, and project/program. The European Catalogue of Data Sources (CDS) is part of the GELOS project. A complete description of CDS is given above.

Global Information Locator Service

<http://www.gils.net/>

The Global Information Locator System was modeled on Government Information Locator Service. Information providers can describe anything with a Global Information Locator Service record--books, people, meetings, artifacts, rocks, etc. The referenced resource is often not available on networks, nor even electronic. For information that is online, the record can include "hyperlinks" for network access to the resource described or related resources. There is no central interface for this project, as it is a machine-level search interface. Of interest are the implementers of GILS found at <http://www.gils.net/implement.html> including the Nordic Web Index http://nwi.bok.hi.is/index_e.html

Green Pages-Environmental Data Directory, Australia

http://www.environment.gov.au/edd/owa/edd_search2.category_list

This Web page allows you to search by categories, e.g., Boundaries, Climate and Weather, Ecology, Fauna, Flora, Forests, Geosciences, Human Environment, etc. The top page <http://www.environment.gov.au/database/edd/> allows basic or advanced searching.

Instituto Nacional de Estadística Geografía e Informática (INEGI), Mexico

<http://ags.inegi.gob.mx/>

Mexico's National Statistics, Geography and Informatics Institute (INEGI) is in charge of generating, incorporating, processing and publishing information on the country's physical environment, geographical features and natural resources, and on its population and productive activities. This site contains statistical data sets and basic thematic maps.

JAPAN GIS/MAPPING SCIENCES RESOURCE GUIDE

<http://www.cast.uark.edu/jpgis/>

The Guide is designed to serve as an introduction to the world of Japanese GIS, remote sensing, geospatial data products, maps, activities and information sources. It serves primarily to help English-speaking geographic information systems and remote sensing

specialists, engineers, geographers, cartographers, geologists, production managers, computer and data processing professionals, and marketing and sales executives.

International Arctic Environmental Data Directory (ADD), Arendal, Norway

http://www.grida.no/prog/polar/add/add_new.htm

ADD is a membership organization where countries with major Arctic data holdings are represented. Members also include international circum-Arctic organizations with data directories. Member countries include:

Canada, Denmark, Finland, Germany, Greenland, Iceland, Norway, Russia, Sweden, UK, and USA This directory covers information sources and some data, e.g., NERC Arctic Environmental Metadata Centre compiled by the Scott Polar Research Institute.

Southern African Metadata

<http://www.gims.com/metadata/>

A consortium of companies and the government is responsible for this site. At present, the geospatial metadata is that of the CSIR Division of Water, Environment and Forestry Technology (Environmentek). This Division is the largest source of multi-disciplinary competencies in South Africa in the field of natural resource and environmental management. Positioned at the forefront of environmental policy development, integrated environmental management (IEM) and natural resource management, it provides innovative solutions for environmentally sustainable development to industry, government and local and rural communities through contract research, consultation and technology transfer. The Division has a spatial technology group of about 25 people covering the following areas of expertise: Geographical information system applications, Remote sensing applications, 3D visualization and virtual reality technologies, and Decision support systems.

United Nations Environmental Programme (UNEP)

Global Resources Information Database

<http://grid2.cr.usgs.gov/>

United Nations Environmental Programme (UNEP) Global Resources Information Database (GRID) Meta-Data Directory.

<http://www.grid.unep.ch/mdd/home.htm>

This directory primarily houses dataset descriptions from developing countries. They also use the Directory Interchange Format (DIF) "standard". The Directory itself is downloadable for PCs (GCMD helped write the software). The intent was that many nations do not have Internet access, so a PC-version of the directory was made for those folks to search metadata and to write descriptions of their metadata.

The Caribbean Environment Programme

<http://www.cep.unep.org/>

This site has a searchable directory of metadata and an interactive GIS system.

Commercial Sites:

Data Hound

<http://nt1.esri.com/scripts/production/esri/marketing/datahound/main.cfm>

An ESRI service that catalogs and searches Web sites that offer freely downloadable data compatible with ESRI software.

Use of Metadata Directories

The creation of metadata records is labor intensive. While the use of online directories can be measured, the actual value to researchers is hard to assess.

Many Web sites have automatic counters that indicate the total number of times the site has been accessed. Other site administrators have tried to present a more detailed analysis of usage. In the present study, statistics of the Global Change Master Directory and the NBII Metadata Clearinghouse were reviewed.

The statistics discussed below often offer a very distorted view of actual use. In discussions with Gene Major, task leader for the Global Change Master Directory, it was pointed out that when a page is accessed, all associated images, e.g., jpegs, gifs on that page also receive a "request," or hit. Therefore, the raw figures presented in the statistics may be greatly inflated. There seems to be a great deal of discussion in the geospatial metadata world concerning users vs. uses, a debate that has been raging for years in the library/information world.

The Global Change Master Directory statistics

<http://gcmd.nasa.gov/announcements/wwwstats.html>

include: Monthly report, Weekly report, Daily summary, Hourly summary, Domain report, Directory tree report plus a summary report (Figure 1) of usage.

Figure 1: Summary report of statistics for date of record specified

Server: <http://gcmd.gsfc.nasa.gov/> (NCSA Common)

Local date: Wed Aug 26 11:18:45 AM EDT 1998

Covers: 08/01/98 to 08/26/98 (26 days).

All dates are in local time.

Requests last 7 days: 56742

New unique hosts last 7 days: 3089

Total unique hosts: 11332

Number of HTML requests: 26731

Number of script requests: 62701

Number of non-HTML requests: 85888

Number of malformed requests (all dates): 15847

Total number of all requests/errors: 191167

Average requests/hour: 312.7, requests/day: 7505.8

Running time: 1 minute, 38 seconds.

For the 26 day period of August 1-August 26, 1998, there were a total of 191,167 requests, or an average of 7353 requests/day.

The domain statistics also offer some interesting insights. Only those domains with 1,000 accesses or more are shown in Figure 2. A complete listing is available at the Web site.

Figure 2. Global Change Master Directory HTTP Server Domain Statistics

Covers: 08/01/98 to 08/26/98 (26 days). All dates are in local time.
1 level, sorted by number of requests, 93 unique domains.

reqs : # uniq : Last Access (M/D/Y) : Domain

52479 : 522 : 08/26/98 : US Government (.gov)
26868 : 2628 : 08/26/98 : US Commercial (.com)
22375 : 2307 : 08/26/98 : (numerical domains)
19457 : 2333 : 08/26/98 : Network (.net)
13188 : 876 : 08/26/98 : US Educational (.edu)
10803 : 175 : 08/26/98 : United Kingdom (.uk)
3911 : 260 : 08/26/98 : Australia (.au)
3783 : 180 : 08/26/98 : Japan (.jp)
2606 : 274 : 08/26/98 : Germany (.de)
1965 : 17 : 08/26/98 : Hungary (.hu)
1594 : 185 : 08/26/98 : Canada (.ca)
1340 : 101 : 08/26/98 : United States (.us)
1328 : 86 : 08/26/98 : Netherlands (.nl)
1112 : 100 : 08/26/98 : Non-Profit (.org)

From the domain statistics, one can determine that the heaviest users are U.S. governmental, U.S. commercial, undifferentiated numeric domains, networks, U.S. educational. Computers from more than 80 countries accessed this site.

The Server Tree Report (Figure 3) gives the number times a particular directory/file was accessed.

**Figure 3. Global Change Master Directory HTTP Server Tree Report
(Subset of Total Report)**

Covers: 08/01/98 to 08/26/98 (26 days). All dates are in local time.

of Requests : Last Access (M/D/Y) : Dir/File

2018 :	08/26/98 :	FAQS [Frequently Asked Questions]
30 :	08/26/98 :	about.html
52 :	08/25/98 :	climdata.html
268 :	08/26/98 :	co2rise.jpg
24 :	08/25/98 :	elnino.html
20 :	08/24/98 :	entp2090.gif
170 :	08/26/98 :	faqpage.html
471 :	08/26/98 :	gcmd_icon.gif
230 :	08/26/98 :	globwarm.html

While a thorough analysis of the Server Tree Report is not the intent of this paper, it is interesting to note the types of information, which can be gleaned.

As an example, this Directory offers four types of searching interfaces: Guided Interface, Free-Text Interface, and two experimental interfaces: Science Keyword Interface, and Query Preview Interface. By associating appropriate URL file extensions with the tree report, one can compare interface use.

The top page, which displays all four interfaces, is at
<http://gcmd.nasa.gov/search/interfaces.html>

This top page was accessed 1124; 08/26/98 was the last access date.

To some extent individual interface use can also be studied:

of Requests : Last Access (M/D/Y) : Dir/File

629 :	08/26/98 :	mainquery.html is the Guided Interface
2374 :	08/26/98 :	zgate includes but is not limited to the Free-Text Interface. I have been unable to distinguish the free-text interface portion.
3647 :	08/26/98 :	param_search is the experimental Science Keyword Interface

Beneath the param_search entry are the statistics on science keyword searches:

20 : 08/25/98 : OCEANS.html
37 : 08/26/98 : OCEANS_BATHYMETRY.html
40 : 08/26/98 : OCEANS_COASTAL_PROCESSES.html
34 : 08/26/98 : OCEANS_MARINE_GEOPHYSICS.html
33 : 08/26/98 : OCEANS_MARINE_SEDIMENTS.html
33 : 08/26/98 : OCEANS_OCEAN_ACOUSTICS.html
37 : 08/26/98 : OCEANS_OCEAN_CHEMISTRY.html
37 : 08/26/98 : OCEANS_OCEAN_CIRCULATION.html
33 : 08/26/98 : OCEANS_OCEAN_PRESSURE.html
37 : 08/26/98 : OCEANS_OCEAN_TEMPERATURE.html
33 : 08/26/98 : OCEANS_OCEAN_WATER_BUDGET.html
35 : 08/26/98 : OCEANS_OCEAN_WAVES.html
35 : 08/26/98 : OCEANS_OCEAN_WINDS.html
35 : 08/26/98 : OCEANS_SALINITY_DENSITY.html
35 : 08/26/98 : OCEANS_SEA_ICE.html
34 : 08/26/98 : OCEANS_TIDES.html
1 : 08/07/98 : OCEANS_VAL.html

187 : 08/26/98 : ~hcil/ index.html is the experimental Query Preview Interface

It is not possible to examine the queries used in free text searching from the data on the Web; however, Gene Major, GCMD, was able to supply a portion of the July 1998 access log for the system. Words searched included: water, temperature, global, rainfall, ozone, SST, sea, daily, daily and weather, space, dry, carbon, avhrr, Langley, global and ocean, snow, satellite, etc.

Of importance to GCMD is the growth of the records in the directory. From an initial catalog of 2000 records, or DIF's, in August 1993, as of June 1998, there were more than 6000 records.

The NBII Metadata Clearinghouse Gateway has just recently begun to keep data on site use. Monthly Summary Statistic Files for the NBII Metadata Clearinghouse Database can be found at http://www.emtc.usgs.gov/http_data/meta_isite/isite_stats_summary/

At the present time, data are available for the previous month and consists of domain use and unique search terms used.

Figure 4 Sample of the domain listings from NBII Clearinghouse Database

Summary Statistics For the NBII Metadata Clearinghouse Database

Date:08011998

Number Of Searches By Host For Previous Month [July 1998]

Format: Number Of Searches - By Host

```
1 128.163.81.192
1 129.71.57.63
1 131.187.131.190
1 132.210.173.7
22 140.90.236.88
2 141.117.7.179
1 fenwood-ip.fenwood.co.uk
1 host131-084.ghpl.lib.oh.us
1 infon282.jet.es
1 jc-39.connect.more.net
```

Using GCMD domain formats for analysis gives the following view of usage:

NBII Metadata Clearinghouse Directory use in July 1998 by domain

# reqs	Domain
621	U.S. Government (.gov)
115	Numeric domains
10	Network (.net)
9	U.S. Commercial (.com)
4	U.S. Military (.mil)
4	Canada (.ca)
3	U.S. Educational (.edu)
3	Spain (.es)
2	Italy (.it)
2	Australia (.au)
1	Finland (.fi)
1	Russian Federation (.ru)
1	United Kingdom (.uk)
1	United States (.us)

Also available from this site is an analysis of the unique search terms used. A sample of the terms is given in Figure 5.

Figure 5. Unique Search Terms Used in NBII Metadata Clearinghouse, July 1998

Format: Number Of Times Used - Unique Search Term
1 "disease"
1 ""catch and release""
1 ""Air quality""
1 ""Climate""
1 ""Climatic zones""
1 ""Environmental policy""
1 ""Global warming""
1 ""Heating Degree Days""
1 ""Marine pollution""
1 ""Meteorology""
1 ""Public health""
2 "49 48 -122 -123"
1 "49.38 45.56 -96.18 -104.43"
1 "49.39 45.15 -116.53 -125.15"
3 "49.61 43.75 -103.44 -116.66"
2 "49.76 43.11 -89.11 -97.62"
1 "Americanism: menace to the World"
2 "ArcView"
1 "Architectual Acoustics Industrial Noise Control"
1 "Atlantic Forest"
1 "Tae kwon-do green belt to red belt"
1 "Divine Comedy"

The range of unique search terms offers some interesting views of the use of metadata directories. First, many of these directories are Z39.50 compatible which means searches started in remote sites could possibly find their way into the NBII Metadata Clearinghouse. Second, it is equally likely that users of the Web don't really understand the contents of metadata directories. Both situations would lead to futile searches and retrievals.

Conclusions

The evolution of metadata for data sets and their directories offers fertile ground for researchers. Never before have major global initiatives existed to identify data sets; however, as with any neoinformation format, a plethora of problems exist. These include multiple definitions of metadata, researcher hesitancy to complete metadata, competing metadata formats, a profusion of metadata directories with overlapping coverage, and a lack of understanding for both producers and users of how this metainformation will be used.

Nonetheless, it is the author's opinion that the potential rewards offered by metadata for identifying and integrating data sets into research endeavors will inextricably change the way scientific research is conducted. The results are that each study will more fully integrate preceding data leading to a more precise understanding of the world.

References

- Major, Gene R., Hughes STX Corporation, Global Change Master Directory [personal communication, August 27, 1998]
- Michener, William K., Brunt, James W., Helly, John J., Kirchner, Thomas B., and Stafford, Susan G. 1997. "Nongeospatial Metadata for the Ecological Sciences." *Ecological Applications* 7(1):330-342
- Vogel, R.L. & Northcutt, R.T. 1997. "Combining Controlled-Vocabulary and Free-Text Search Options in Order to Enhance the Retrieval of Metadata for Access to Environmental Data." Presented at the Fourth International Conference on Remote Sensing for Marine and Coastal Environments, Orlando, Florida, 17-19 March 1997.
- Vogel, R.L. 1998. "Why Scientists Have Not Been Writing Metadata." *EOS: Transactions, American geophysical Union*. 79(31):373, 380.