

ONLINE DATA PORTALS: ORGANIZING OCEAN DATA FOR THE SCIENTIFIC COMMUNITY *

Monica Holland

NASA's Global Change Master Directory,
Greenbelt, MD USA


ABSTRACT

NASA's Global Change Master Directory (GCMD), an online Earth science directory (<http://gcmd.nasa.gov>), directs users to over 3,000 ocean data set descriptions and over 10,634 Earth science descriptions. The search for oceanographic data (as well as Earth science data) is a simple task when using the GCMD. In response to requests from representatives from partner organizations who expressed an interest in the directory software, the GCMD now offers focused views of the directory through portals. The portals have made it easier for partner organizations to maintain and document their data holdings in one place, without duplicating the effort to create another online directory.

1.0 SEARCHING FOR DATA

Finding data related to ocean science online can be an overwhelming task while using internet search engines. Although search engines adequately provide links to ocean related web sites, finding appropriate links to data sets can be laborious. Conducting a search for ocean data using NASA's Global Change Master Directory (GCMD), an online Earth science directory (<http://gcmd.nasa.gov>), explicitly provides data set sources, direct access links to data sets, and additional Earth science resources.

The GCMD's "search engine" directs users to over 3,000 ocean data set descriptions and over 10,634 Earth science descriptions. The search for ocean data (as well as other Earth science data) is a simple task when using the GCMD. Searching from the GCMD home page allows users to select from controlled keywords or search using free-text. The data descriptions within the GCMD comprise a collection of metadata fields, which provide specific information about the data. Metadata, defined as data about data, distinguishes how, when, and by whom a particular set of data was collected, and how the data are formatted (Webopedia, 1997). The metadata standard used to create the records within the directory is based on the Directory Interchange Format (DIF) (Olsen, 2002).

The controlled keyword search is based on a hierarchical set of Earth science keywords. Each metadata record, has one or more keywords to ensure that the user can find a data set of interest. The controlled keyword search includes a topical list of Earth Science categories (*Agriculture, Atmosphere, Biosphere, Human Dimensions, Hydrosphere, Land Surface, Oceans, Paleoclimate, Radiance/Imagery, Snow or Ice, Solid Earth, and Sun-Earth Interactions*). Choosing the *Oceans* topic keyword from the homepage narrows the search to ocean terms that are arranged in a hierarchy (http://gcmd.gsfc.nasa.gov/Data/portals/gcmd/param_search/OCEANS.html). The number of records, as well as a definition, are displayed for each of the terms. An icon , provides a link to the definition for the ocean term.

* Presented at the Seventh International Conference on Remote Sensing for Marine and Coastal Environments, Miami, Florida, 20-22 May 2002.

A complete list of GCMD controlled keywords can be reviewed online (<http://gcmd.gsfc.nasa.gov/Resources/valids/index.html>). Other controlled keyword topics including locations, instruments, platforms, or projects can also be searched within the GCMD.

The free-text search provides an option for a user to search the GCMD using any term(s). A query from the free-text search simply requires typing a descriptive word or phrase within the free-text box on the home page. An enhanced free-text search allows Boolean, fielded, geospatial or temporal searches.

2.0 DATA DESCRIPTIONS

The data set description presents pertinent information to help the user determine if the data may be relevant for their research interest. A complete list of the types of descriptors (*DIF Fields*) that may be used within a data set description are documented in the DIF Writer's Guide Version 8 (<http://gcmd.nasa.gov/User/difguide/difman.html>). Two display formats, the *Brief Record* and the *Full Record* are available to the user. Within the *Brief Record*, the interface display includes only the *Summary*, the *Data Center*, the *Data Set Citation*, the *Geographic Coverage* and the *Related URL* sections. The *Full Record* display includes additional detailed descriptors, which may clarify or expand upon the information about the data set.

Through the controlled keyword search, resulting DIF records can be navigated using a tabbed layout. The tabbed layout provides the option to display the *Full Record*, specific segments of the DIF (i.e. *Distribution*, *Attributes*, *Coverage*, and *Personnel*), or an option to update the record using an online DIF modification tool (see Figure 1).



Figure 1. DIF Tab Layout

The user can expand their search within the record using hyperlinked fields that are associated with other data set descriptions within the GCMD database. For example, parameters are hyperlinked within the DIF to include a link to all of the data sets that have the same parameter (Figure 2).



Figure 2. Example of Parameter Hierarchy DIF Display

2.1 ACCESS TO DATA

Accessing data is possible through the Related_URL field within the record. Through this field, the user can directly link to the data. Providing direct access to the data source enables the user to locate additional data set related resources outside of the GCMD website (Figure 3). The Related_URL field provides the URL, a description about the URL, and a URL_Content_Type field where appropriate. Another way the GCMD provides access to data is through the *Data Set Citation* field. The *Data Set Citation* for the data set properly credits the data set producer.

Related URL
Content Type: ASSOCIATED DATA SET(S) URL: http://daac.gsfc.nasa.gov/data/inter_disc/hydrology/sea_ice/ Description: Access to the SMMR and SSM/I sea ice concentration data

Figure 3. Example of Related_URL Field

3.0 PORTALS : COMMUNITY BUILDING

Customized subset views of the GCMD (“Portals”) have made it easier for partner organizations to maintain and document their data sets in one place, without duplicating the effort to create another online directory. The GCMD has recognized the importance of customization for participating organizations and is generating subset views of the directory through a simple “niche portal”. The term “niche portal”(or vertical portal; vortal) describes a website that caters to a given demographic (Kleinschmidt et al, 2001). “Niche portals” have increased in popularity and have been identified as “natural community building machines” (Ozersky, 2000). The integration of customized portal searches with the ability to organize data make the portal a powerful tool (Ozersky, 2000). GCMD offers customized subset views for the following partners (<http://gcmd.gsfc.nasa.gov/Data/portals/index.html>):

- Antarctic Master Directory (AMD), Joint Committee on Antarctic Data Management (JCADM),
- The Committee on Earth Observation Satellites (CEOS),
- Climate Variability and Predictability (CLIVAR),
- The Distributed Oceanographic Data System (DODS),
- The Earth Science Information Partners (ESIP),
- The Global Ocean Ecosystems Dynamics program (GLOBEC),
- The Global Observation of Forest Cover (GOFC),
- The Global Observing System Information Center (GOSIC) (including G3OS; the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS), and the Global Terrestrial Observing System (GTOS)),
- The Rosenstiel School of Marine and Atmospheric Science (RSMAS),
- The World Data Centers (WDCs).

Two of the largest ocean-related data providers are GLOBEC and DODS. The GLOBEC and DODS portals are used as a metadata directory, ensuring long-term stewardship for related data set descriptions (GLOBEC, 1999).

3.1 GLOBEC

The common interest in global change data management has attributed to the international collaborative effort between GCMD and GLOBEC. The aim of GLOBEC is to advance understanding of the structure and functioning of the global ocean ecosystem, its major subsystems, and its response to physical forcing so that a capability can be developed to forecast the responses of the marine ecosystem to global change (GLOBEC, 1999). GLOBEC functions as one of the nine core projects of the International Geosphere-Biosphere Program (IGBP), an interdisciplinary scientific activity established and sponsored by the International Council for Science (ICSU).

GLOBEC data managers decided to use the GCMD as a metadata inventory as a secure way of preserving a record of the results and achievements of the GLOBEC program (GLOBEC 1999). GLOBEC 's portal is used within the international GLOBEC website (<http://www.globec.org>) as a key resource in disseminating information to their scientific community (GLOBEC 2002). The GLOBEC data policy includes the adoption of the DIF metadata format as the recommended format for all data set descriptions. Currently, over 110 metadata records have been contributed to the portal by using an online tool (DIFBuilder). DIFBuilder prompts the user to enter in required fields to create an online data set description. The DIFBuilder tool is available on the GCMD website (<http://gcmd.gsfc.nasa.gov/cgi-bin/difbuilder/difbuilder>).

The success of the portal can be attributed to the high quality of metadata, and its open accessibility to the GLOBEC community. The portal has been popular within the international community (see Figure 4). A detailed analysis of the portal usage can be reviewed in Table 1. The GLOBEC portal can be accessed online here: <http://gcmd.gsfc.nasa.gov/Data/portals/globec/>

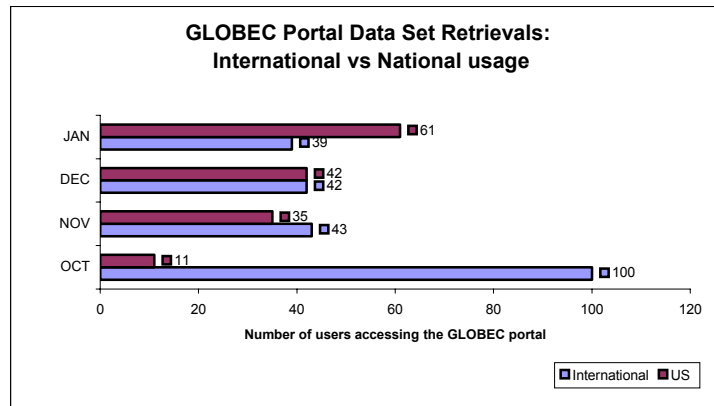


Figure 4. GLOBEC Portal Data Set Retrievals: International versus National usage

3.2 DODS

DODS is a software framework that simplifies all aspects of scientific data networking, allowing simple access to remote data. Local data can be made accessible to remote locations regardless of local storage format by using DODS servers (DODS, 2002). Since 1994, GCMD and DODS have met frequently to plan a strategy for linking the user to a data server. GCMD has taken an active role in ensuring that metadata records provide additional information about the data served by DODS. DODS has incorporated the use of the GCMD as their data locator service within their website <http://www.unidata.ucar.edu/packages/dods/home/data.shtml> (Figure 5).

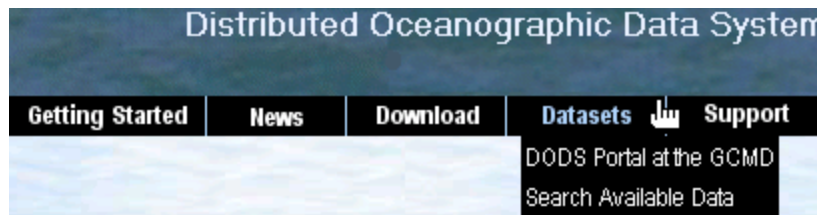


Figure 5. DODS website link to the DODS Portal as the data locator

Currently there are more than 195 data sets served by DODS. All data set descriptions are created to reflect consistency between the number of data sets presented on the DODS website and the DODS portal. The DODS portal uses a similar method to identify the URLs within the data sets (Figure 6) to ensure that the user will be able to locate data set content using either website. Currently statistics show increasing usage of the portal (Figure 7). Additional statistics provide a detailed analysis of the portal usage (Table 1). The DODS portal can be accessed online: <http://gcmd.gsfc.nasa.gov/Data/portals/dods/>

A):

[Atlas of Surface Marine Data 1994 on CD-ROM from NOAA/NODC](#) - [URL](#) | [LAS](#)

(B):

▶ [NOAA - Pacific Marine Environmental Laboratory \(PMEL\)](#)
♦ [Atlas of Surface Marine Data 1994](#) [gcmd](#) | [html](#) | [info](#) | [site-info](#) | [dir](#) | [live-access](#)

Figure 6. A comparison of the DODS Portal search results (A) and the DODS Website Data set Menu page (B)

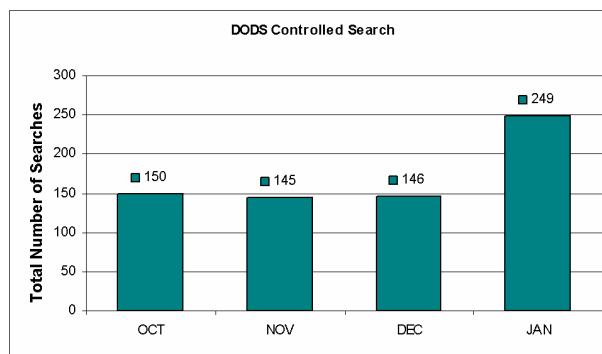


Figure 7. DODS Portal: Controlled Keyword Statistics Oct. 2001-Jan. 2002

3.3 OCEAN PORTAL USAGE

Data set Search Retrieval Statistics per month	OCT. '01	NOV. '01	DEC. '01	JAN. '02	Average
GLOBEC Portal					
Users viewing Data Set Titles	361	628	255	397	410
Data set descriptions retrieved	172	147	89	71	120
International users	100	43	42	39	56
United States users	11	35	42	61	37
DODS Portal					
Users viewing Data Set Titles	450	921	323	603	574
Data set descriptions retrieved	11	53	49	69	46
International users	0	7	1	18	7
United States users	38	76	75	85	69
Controlled Keyword Search Statistics per month					
GLOBEC Portal					
Total hits to GLOBEC Parameter search	159	161	177	241	185
Total International users	27	20	22	95	41
Total National users	105	78	101	21	76
DODS Portal					
Total hits to DODS Parameter Search	150	145	146	249	173
Total International users	1	8	1	14	6
Total National users	125	85	107	134	113
All controlled keyword topic search statistics per month					
GLOBEC Portal	275	292	269	390	307
DODS Portal	233	254	232	375	274
Total number of unique users for all 10 GCMD Portals Nov 2001 through Jan. 2002		5578	5420	6413	5804

Table 1. Users accessing DODS and GLOBEC Portal

4.0 OUTREACH

Several other online resources are available to participating organizations. Resources include 1) listings of websites related to Earth Science for each discipline (<http://gcmd.gsfc.nasa.gov/Resources/pointers/pointwais.html>), 2) a group of discipline-specific Earth Science FAQ from the scientific community of users (<http://gcmd.gsfc.nasa.gov/Resources/FAQs/faqpage.html>), and 3) a conference calendar, which provides an interactive calendar to review the Earth Science conferences worldwide (<http://gcmd.nasa.gov/Resources/calendar/conf/calendar.html>). The combination of these Earth Science resources with the discovery of data sets makes GCMD a valuable online resource.

5.0 CONCLUSION

The GCMD continues to assist the scientific community in the discovery of and linkage to Earth science data, as well as to provide data holders a means to advertise their data to the Earth science community (Olsen, 2002). The ocean science portals are effectively serving the GLOBEC and DODS community by increasing the visibility of their data holdings. The GCMD staff is working with other ocean related organizations including: Rosenstiel School of Marine and Atmospheric Science (RSMAS), Global Observing System Information Center (GOSIC), and Antarctic Master Directory (AMD), Joint Committee on Antarctic Data Management (JCADM). The customized subset views of their datasets descriptions are also available (<http://gcmd.gsfc.nasa.gov/Data/portals/index.html>).

5.0 REFERENCES

- Distributed Oceanographic Data System (DODS). 2002. "What is DODS",
Internet resource: <http://www.unidata.ucar.edu/packages/dods/home/faq/whatIsDods.shtml>
- GLOBEC, "GLOBEC Implementation Plan", *IGBP Report 47/GLOBEC Report 1*, 1999
- GLOBEC. 2002. International GLOBEC website. Internet resource: <http://www.globec.org>
- C. Kleinschmidt and A. Goodman, 2001. " Frequently Asked Questions About Portals".
Internet resource: Traffick.com,
- L. Olsen. 2002. "What is a DIF?" Internet resource: <http://gcmd.gsfc.nasa.gov/User/difguide/difman.html>
- J. Ozersky. 2000. "From Portals to Vortals". Internet resource: business2.com
- Webopedia. 1997. "Metadata". Internet resource: <http://webopedia.com/TERM/m/metadata.html>

ACKNOWLEDGEMENT

I wish to acknowledge the editorial assistance of NASA's GCMD Project Manager, Lola M. Olsen, and SSAI's GCMD Task Manager, Gene Major on this paper. I also wish to thank Hester Willson (International GLOBEC Data Manager) and the DODS team for their partnership and collaboration with the GCMD. Work was performed under NASA contract #NAS5-00220.