Designed By:
Christina Chiddo
Edited By:
Lola Olsen

CEOS IDN NEWSLETTER

Issue 16
May 2004

INSIDE THIS ISSUE:

Agenda for CEOS Meeting

By Lola Olsen, Task Team Leader, CEOS IDN

Upcoming CEOS Meeting Will Be Held North of the Arctic Circle
See you @ 69.65 N, 18.9667 E

We are looking forward to seeing you in Tromso, Norway for the CEOS Subgroup meetings in May. The IDN Task Team will meet at 3:45 p.m. on Monday, May 10th. Tromso is a delightful spot - although you may have some difficulty sleeping because of the activity that goes on around the clock at this time of year. Here are a few photos (among hundreds) taken by Lee Belbin in June, 2001 at a meeting of the Joint Committee on Antarctic Data Management (JCADM) in Tromso.

Lee Belbin (JCADM) and Lola Olsen (CEOS IDN Task Team Lead) dancing at midnight on top of a mountain in Tromso.

A view from the Norsk Polar Institute in Tromso in June 2001.

The International Directory Network has been very active since the last CEOS meeting. The release of MD9-ISO is scheduled for May 2004. At the upcoming meeting, you'll hear from Viktor Pusztai (UNEP), who has been working to provide database support with PostgreSQL for MD9. You'll also hear about the status of the exchange of metadata between the

Cont. on next page >

< **cont. from previous page**

IDN and the Canadian Centre for Remote Sensing's (CCRS) GeoConnections Discovery Portal – which is due to resume this month. Ms. Andrea Buffam of CCRS continues to serve as the IDN representative with the GCMD's Science User Working Group (UWG). (See next article.)

All numbers are approximate times in minutes.

Review of Minutes from Chiang Mei, September 2003 and Status of American Coordinating Node:
Olsen (15)
- Metrics for content and usage
- Keyword Modification Summary
- Collaborations – (e.g., portals)
- Software Development and the status of Local Database Agents
- Maintenance Reductions
- Science User Working Group

IDN Node Reports:
(Collective Estimate – 16)

The CEOS IDN Interoperability Voting Committee, Status of Proposals, and Related Requirements
Voting Committee
Lorant Czaran (UNEP/GRID)
Status of ISO Modifications
Lorant Czaran (UNEP/GRID) (5)
Evaluation of Spatial and Temporal Resolution Query
Ivan Petiteville (ESA) (6)
Requirements for Model Output
Glenn Rutledge (NOAA/NCDC) (9)
AVISO—Paul Kopp (5)

Review of Evaluation of Geospatial Search
Allan Doyle (International Interfaces) (8)

Joint Committee on Antarctic Data Management (JCADM)
Stein Tronstad (Norwegian Polar Institute) (10)

Use of PostgreSQL in MD9
Viktor Pusztai (UNEP/GRID-Budapest) (8)

Efficient use of XML in Data Exchange
Brian McLeod (CCRS) (8)

Utility of Keywords for CIP
Jolyon Martin (ESA) (8)

Utility of Services in the IDN
Bernhard Buckl (DLR) (6)

"Getting Started with the IDN's docBUILDER," paper by Craig Walters (USGS/EROS Data Center, Land Processes DAAC), presented by
John Faundeen (USGS) (10)

Schedule of Content of New Releases – MD9
Olsen (NASA/GCMD) (6)

GCMD's Science User Working Group Plans Next Session in Greenbelt, MD

By Acting Chairman Martin Ruzek
Universities Space Research Association

Plans are being finalized for the next session of the Global Change Master Directory's Science User Working Group (UWG), which will be held on May 25th and 26th, 2004. Several new members will participate: Dr. Walter R. Hoegy, Assistant Chief for Science and Education, Laboratory for Atmospheres at NASA; Dr. Erick Chiang, National Science Foundation/Office of Polar Programs; Mr. Glenn Rutledge, NOAA/National Climatic Data Center (NCDC); and Dr. Doug Beard, USGS/ Biological Resources Division (BRD). Returning members include Dr. Wendell Brown, University of Massachusetts-Dartmouth; Dr. Hubert Staudigel, Scripps Institution of Oceanography, University of California, San Diego, Institute for Geophysics and Planetary Physics; and Ms. Andrea Buffam from the Canadian Centre for Remote Sensing, representing the CEOS IDN. Dr. Christopher Kerr, NOAA GFDL, will also join the group for this meeting. The GCMD Science User Working Group works under the GCMD UWG Terms of Reference.

See http://globalchange.nasa.gov/Aboutus/UWG/uwg.html

Recommendations from previous UWG meetings have helped to guide the progress of the directory. We anticipate an active and valuable working meeting.

Martin Ruzek, Acting Chairman, and Hubert Staudigel, Scripps Institution of Oceanography, at the last UWG.

Revamping Data Center and Location Keyword Hierarchies

By Gene Major, Heather Weir and Lola Olsen, IDN, American Coordinating Node

Considerable efforts have been placed on organizing the Earth science keywords and Earth science services keywords, but less attention has been focused on the growing number of data and services descriptions that can be searched by other IDN controlled vocabularies - such as data centers, locations, instruments, platforms/sources, projects, chronostratigraphic units, and URL_Content_Types. Searches based on data centers and geographic locations are the focus of this article.

Data Centers:

---------- ---_____

Tom Northcutt, Systems Administrator in the Global Change Data Center

Users currently can search or refine their search for data set information by selecting a "Data Center" keyword. (Searching by "data center" for data services is not yet available). Each Data_Center_Name is a hierarchy consisting of a Category (or bucket), Short_Name and Long_Name (i.e., Category > Short_Name > Long_Name).

A major reorganization to assist users in locating the data center of choice is in progress. The first step in this process was to evaluate and rename the categories (or buckets) in which the data centers are classified. Some categories remained the same, such as "Commercial", while many others were broadened to allow for a more inclusive classification. For example, the "University" category became "Academic", which includes school systems (K-12) as well as Universities (both US and non-US).

The second step was to create a logical naming convention for all data centers. This step has required extensive review of each individual Data_Center_Name and Data_Center_URL in making the appropriate changes. Each Data_Center_Name consists of a Short_Name and a Long_Name. The Short_Name consists of a multi-level hierarchy (usually acronyms by which the data center in known) with the broadest name first, followed by more detailed names.

For example, the Short_Name for the Institute for Space and Astronautical Sciences within the Japan Aerospace Exploration Agency in Japan would be found under the category (bucket), "Government Agencies Non-US" as: JP/JAXA/ISAS.

Here, JP is the country ISO identifier; JAXA is the agency; and ISAS is a department within JAXA.

The Long_Name is also a hierarchy and usually spells out the acronyms of the Short_Name in reverse order - from more detailed to broad. In this example, the logical order of the Long_Name would be:
Institute for Space and Astronautical, Japan Aerospace Exploration Agency, Japan

A third step, to be implemented, will expand the Data Center category (or bucket) to three levels. Currently there are 18 categories, which will be reduced to 8 through the three-level hierarchical category. For example, the category "NOAA" would become "U.S. Federal Agencies > DOC > NOAA". All of the NOAA data centers would then appear under this new hierarchy.

Cont. on next page >

< cont. from previous page

Geographic Locations:

Currently, data sets describing observations within a particular region or country are not consistently indexed with the parent continent. The problem is more acute for ocean observations, where one may find data sets from the North Atlantic Ocean that are not indexed with Atlantic Ocean. Authors have often forgotten to choose the continent or major ocean body keyword from the list in the metadata authoring tool, docBUILDER.

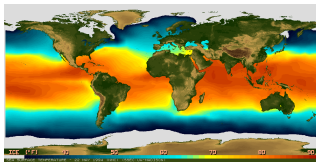
A revised geographic location keyword hierarchy will improve searching for data sets and aid authors in choosing geolocation keywords to describe their data sets. Thus, users will always be able to locate all the data sets available for a selected region.

A 5-level geographic location keyword hierarchy will be defined as:

Category > Division > Subregion1 > Subregion2 > Subregion3
Category is defined as:
- Continent
- Ocean
- Extraterrestrial locations
- Vertical locations

Division if Continent Category is specified:
- Africa
- Antarctica
- Asia
- Australia
- Europe
- North America
- South America

Division if Ocean Category is specified:
- Atlantic Ocean
- Pacific Ocean
- Indian Ocean
- Arctic Ocean
- Southern Ocean



Subregion1, Subregion2, and Subregion3 define specific countries, regions, ocean basins, seas, and islands as necessary.

For example:

Italy would thus be defined as:
Continent > Europe > Southern Europe > Italy

and the North Atlantic Ocean as:
Ocean > Atlantic Ocean > North Atlantic Ocean

This location keyword hierarchy is largely based on the geographic hierarchy used by the USGS Coastal and Marine Geology program, InfoBank Atlas and Gazetteer:

http://walrus.wr.usgs.gov/infobank/programs/html/main.atlas.html

Redesigned Website Released on March 1st

By Rob Gutro, Team Leader, NASA's Earth Science News Team

Published in GEO World, April 2004,
Volume 17 No.4, p12
Internet Resource of the Month

http://globalchange.nasa.gov or http://gcmd.nasa.gov

Using NASA's Global Change Master Directory (GCMD), students, scientists, teachers, reporters and the scientifically curious can locate earth science data easier and faster than before. The redesigned Web site was launched March 1, 2004.

The site, updated daily, now is easier to navigate, with nine tabs running atop the home page: Home, Data Sets, Data Services, Portals, Authoring, What's New, Community, Calendar and Links. GCMD topics, found under the "Data Sets" tab, provide summaries and information such as data over time and location, database creators, and direct links to data and services.

Available topics range from tiny airborne particles (aerosols) to the continental-sized ozone hole to global sea-surface temperatures. General categories include the following: Agriculture, Atmosphere, Biosphere, Climate Indicators, Human Dimensions, Hydrosphere, Land Surface, Oceans, Paleoclimate, Snow and Ice, Solid Earth, Spectral/Engineering and Sun-Earth Interactions.

Users can search more than 15,000 datasets and services as well as link to more than 76,000 resources within the descriptions. Individual dataset descriptions were contributed by more than 1,300 data centers, government agencies, universities, research institutions and private researchers worldwide.

For scientists and others who want to add or modify GCMD data- sets, they can do so under the "Authoring" tab by using the new "docBUILDER" Web-based tools. Under the "Data Services" tab are available services ranging from analysis and visualization tools to education and environmental advisories.

"Perhaps the greatest contribution to the public is the ability to create customized subsets of the directory that can be displayed, in turn, by special-interest groups," said Lola Olsen, directory project manager at NASA's Goddard Space Flight Center, Greenbelt, Md. "These groups save major development and maintenance costs by re-using the directory capabilities."

For example, member countries of the Joint Committee on Antarctic Data Management contribute directory entries using GCMD tools and may, in turn, host individual, customized subsets of the database through "portals" through which they can display their own contributions.

Students and teachers can benefit from a "Learning Center" found under the "Community" tab. In addition, the "Links" tab acts as a Web-based search engine for access to more than 2,500 earth science Web resources.

User Statistics Update for New Website

By Jason Divock, CEOS IDN, American Coordinating Node

The restructuring of the American Coordinating Node's (the GCMD) web site required that the user statistics collection scripts be modified to accommodate the new structure. It is vital to track the changes in usage on web pages that have been relocated and to monitor user response to new pages that have been added. Without these reports, it would be difficult to evaluate the needs of a changing user base.

The current scripts now:

- Track hits to the home page and other web pages on the GCMD web site.
- Monitor popularity of DIFs being searched and retrieved.
- Monitor retrievals of DIFs from outside the GCMD web space.
- Track use of the different portals within the GCMD.
- Determine the domains associated with incoming hits (i.e., .gov, ...edu, .com, .net).
- Monitor the popularity of searches using Isite.

The analysis of usage is vital to the success and usability of the site. For example, "Data Services" and "Portals" are now more prominent on the home page. Comparisons over time will help to determine if these pages are attracting more users and/or if there is information that may need to be more prominently displayed to attract users' attention. Links that are buried too deeply can be identified and moved, and portal usage can be assessed so we can better interact with our partners.

A Note From Jason

As a new member of the IDN Task Team, the analysis of usage was my first task and has provided an opportune way to learn about the structure of the web site and the overall mission of the IDN. The task has helped in understanding other aspects of the project and has allowed me to take an open approach to writing the scripts.

Getting Started with the IDN's docBUILDER

By Craig Walters, Land Processes Distributed Active Archive Center

Craig Walters is a Publication Specialist working for the LP DAAC at the EROS Data Center (EDC) near Sioux Falls, SD. Craig has worked for the LP DAAC over the past three years creating their documentation.

As a technical writer for the LP DAAC (Land Processes Distributed Active Archive Center), I am responsible for creating our data set documentation using the Directory Interchange Format (DIF). I have witnessed firsthand the evolution of that process—from building ODL files on a UNIX-based server, to using the DIFBuilder and now the docBUILDER. The International Directory Network / Global Change Master Directory (IDN / GCMD) continues to evolve in its use of authoring tools. Its latest offering, docBUILDER, represents its marriage of a web-based tool and XML.

cont. on next page >

< **cont. from previous page**

Creating a DIF document with docBUILDER

To launch the IDN's docBUILDER DIF tool:

1. From the Global Change Master Directory's (GCMD) Home page <http://gcmd.nasa.gov/index.html>, click the *Authoring* tab (located in the center of the navigation running across the top of the page). As a result, the *Add Data Set and Services Descriptions to the GCMD* page loads.
2. From the *Add Data Set and Services Descriptions to the GCMD* page, select the *Data Set Description* link (found under the heading, "Add a description to the GCMD"). As a result, the docBUILDER main window loads.

   NOTE: You may choose to create an Earth science "Services" description or provide additional information on relevant projects, instruments, platforms, or data centers.

To begin a new DIF document:

1. On the docBUILDER home page, type an Entry_ID into the document identifier window.

   NOTE: If you need help determining the requirements for your Entry_ID, click the document identifier link.
2. Click the *Begin a new description* button to complete this action. As a result, the document page loads.

Notice that the ENTRY_ID field has a check mark next to it. As you complete each field of your DIF, docBUILDER tracks the document's state of completion. Additionally, if you need to edit a completed field, simply click that field and make the appropriate changes.

> **NOTE**: If at any point in the process of building your DIF you should have questions about the requirements for any particular fields, click the ("i") graphic located at the end of that field's name, and a pop-up window with entry requirements will display.

Using docBUILDER to create your DIF

As you complete the various DIF fields, docBUILDER will update the viewing window (located below the list of DIF fields) to reflect additions to your document. You may only view the document with this window. Additions (or edits) to your DIF document occur through following the process:

The process for populating the fields of any DIF follows the same general procedure:

1. Click the field name. The corresponding field's page loads.
2. Provide all data to complete the page.
3. Click the *Continue* button (at the bottom of the page). You are then returned to your document page showing the completed field denoted by a check mark.

NOTE: Every DIF requires completion of the first five fields for publication:

Entry_ID
Entry_Title
Science_Keywords
Data_Center
Summary

The additional fields are optional but may be crucial to the complete description of your data set. You may complete these fields in a manner similar to the one described here.

Generally, using docBUILDER to create (or edit) a DIF is a straightforward process - as docBUILDER uses an intuitive design. However, a few additional instructions relevant to specific fields follow in the hope of providing a clearer understanding of how the docBUILDER tool functions.

To populate the *Entry_Title* field:

1. Click the *Entry_Title* field. As a result, the *Entry_Title* page loads.
2. Type your Entry Title.
3. When you have completed the *Entry_Title* window, click the *Continue* button to return to your document.

To populate the Science Keywords field:

1. Click the *Science Keywords* field. As a result, the *Science Keywords* page loads.
2. Select the appropriate topic of your keyword phrase. Your browser will update the scrolling display window.
3. Locate your keyword phrase by either 1) scrolling through the category and topic list or 2) typing the keyword's term or variable.

   NOTE: Science Keyword format:

   Category > Topic > Term > Variable > Detailed Variable
4. Click to highlight and select your keyword phrase.
5. Click the *Add Selected Items to the List Below* button. Repeat steps 2 and 3 as needed to add additional keyword phrases, and click the *Continue* button at the bottom of the page when you have completed adding your Science Keywords.

Cont. on next page >

< **cont. from previous page**

NOTE: If you need to clear your Science Keywords, click the *Delete All Instances* button.

6. To complete the addition of Science Keywords, click the *Continue* button at the bottom of the window. You are then returned to your document page showing the completed field: in this case, *Science Keywords*, denoted by a check mark.

To populate the Data Center field:

1. Click the *Data Center* field. As a result, the *Data Center* window loads.
2. Completing the *Data Center Names* section of the *Data Center* window:
    a. In the Data Center Names section of the window, click the alphabet grouping within which your data center falls. Your browser will then display a corresponding list of the appropriate data centers for that portion of the alphabet.
    b. Using the *Search the List* search box, type in the first few characters of your data center. Your browser will respond with the corresponding portion of the data center list.
    c. Select the data center from the list of choices OR scroll through the list to find your data center.
    d. Click the *Set Data Center Name* button and navigate to the *Personnel* section of the window.
3. Completing the *Personnel* section of the *Data Center* window:
    a. In the *Personnel* section, search for your personnel identification information by using the same approach as above. First, click the letter of the alphabet where your last name is located, and then select it from the list before clicking the *Set the Personnel* button.
    NOTE: You may review the personnel information before loading it by clicking the *Show Selected Person's Full Record* button.
    b. After reviewing the information, click the *Set the Personnel* button. Your browser will load this information into the corresponding fields below.
    NOTE: You may edit these fields as necessary.
4. Complete this window by typing in your *Data Center URL* and *Data Set ID*. The Data Set ID may also be the DIF's Entry_ID.
5. When you complete the *Data Center* window, click the *Continue* button to return to your document.

To populate the Summary field:

1. Click the *Summary* field. As a result, the *Summary* page loads.
2. Copy and paste your summary text into the *Summary* window.
    NOTE: You may compose your summary in this window, but you would be better served composing your text with the benefits of a word processor.
3. When you complete the Summary window, click the *Continue* button to return to your document.

Complete other DIF fields pertinent to your needs and requirements before reviewing and submitting your DIF.

Submitting your DIF for sharing

To submit your DIF document to the IDN (GCMD):

1. From the Document Menu pull-down, select *Submit document to the GCMD*.
2. After reviewing the document, type your e-mail address and click the *Mail it!* button. You will receive confirmation via e-mail from the GCMD.

    NOTE: If you wish to send your document to yourself only, select *Email document to yourself* from the pull-down Document Menu.

Local Database Agent

By Chris Gokey

In the recent months, the Local Database Agent has been re-engineered and can now also be deployed in a system that may not necessarily be running MD8/9. This software is used by the Global Change Master Directory to mirror metadata content among participating distributed nodes. It provides a peer-to-peer data sharing network - allowing organizations to have autonomy over their own metadata repository while still being able to be part of a larger community where their metadata can be shared with each other.

The software allows nodes to develop and plug in their own Metadata Repository object (or they can use the provided IDN Repository object) to work with their own database. The LDA will interact with their repository object, retrieve changes from their database, and update their database with any changes propagated from other nodes on the network.

The core component in the system that actually handles the synchronization is called JSync. This component provides a way to synchronize two systems in a manner similar to the Unix rsync utility. It works by generating a "manifest" file on the server. Nodes connect, download the manifest file, and compare their local manifest file against the servers. The result is the differences between the two systems. All changes are captured and automatically loaded into the database.

Filtered subsets of the IDN content can also be supported. Nodes that only request a subset of the content (e.g., only Antarctic data) can synchronize against a specialized repository. Also included in this release is a built-in SchemaTranslator component that allows nodes to map schemas using Document Object Models and XSL files. Therefore, nodes with an FGDC repository can participate in this way.

THE CEOS IDN NEWSLETTER
MAY 2004