NASA'S GLOBAL CHANGE MASTER DIRECTORY: FOSTERING COLLABORATIONS FOR EARTH SCIENCE INFORMATION AND DATA RETRIEVAL

Gene R. Major
NASA Global Change Master Directory
Raytheon ITSS
4500 Forbes Blvd. Suite 300
Lanham, MD 20706
major@gcmd.nasa.gov

ABSTRACT: NASA's Global Change Master Directory (GCMD), a freely available online information system (http://gcmd.nasa.gov), provides access to a wide variety of global change and Earth science data and information. The GCMD offers search and retrieval of Earth science information from a variety of user interfaces including simple free-text searching, controlled keyword lists, and Java-based query tools. One of the successes of the GCMD has been to form collaborative partnerships with Federal and international agencies and organizations. Through these partnerships, individual agencies and organizations can meet their own metadata needs, yet ensure that data set holdings are known in the context of a multidisciplinary Earth science information catalog.

1.0 Global Climate Change Information

The discipline of climate change science involves one of the more hotly debated scientific issues today. The debate over whether or not environmental change is human-induced has been ongoing among scientists, politicians and policymakers, industrialists and economists. In 1995, a report by a panel of over 200 international scientists added to the debate by declaring that the "balance of evidence suggests a discernable human influence on global climate" (IPCC 1995).

Many U.S. Federal agencies as well as international organizations are undertaking observations and research on the question of global change in order to provide data and information to make informed policy decisions. The cornerstone of informed decision making is the availability and access to global environmental change data and information. The National Research Council (1998) states that "[u]ltimately, the [U.S. Global Change Research Program] USGCRP is about information...the subject of the Program's research demands that information flow effectively to the public at large as well as to researchers."

2.0 Global Change Master Directory

NASA has developed the Global Change Master Directory (GCMD) to provide access to Earth science and global climate change data worldwide. The GCMD, available on the World Wide Web (http://gcmd.nasa.gov), is an online information system that assists the research community and others in locating global change and Earth science data held at Federal agencies, universities, research laboratories, international organizations, and commercial enterprises worldwide.

Much like a bibliographic catalog, the GCMD provides citation-like descriptive information about Earth science data sets and guides the user to the location of these data sets, whether in national data centers, individual investigators' collections, project data repositories, or data journals. The GCMD database contains over 8000 descriptions of data sets covering climate change, the biosphere, hydrosphere and oceans, geology, geography, and human dimensions of global change. Data sets that are being collected from national and international programs are described along with characteristics and attributes of the data and information on how to obtain the data. New data descriptions are added daily and over 500 new data set descriptions are added annually. Since the GCMD is accessed through the Web, hyperlinks are often provided to link the user directly to the data.

Data set descriptions are entered in the GCMD in the Directory Interchange Format (DIF). The DIF is a metadata standard that consists of a specific set of information fields to assist in normalizing the search for data set information. Fields such as data set title, summary (or abstract), and data center is available along with 30 other DIF fields. The GCMD philosophy for using the Directory Interchange Format is that the predetermined set of fields is the critical set needed for a user to determine if the dataset information returned from a database query is that that defines viable alternative data sets for the user's needs. Online metadata authoring tools have been developed to assist authors in developing metadata.

Several search and retrieval interfaces for accessing the GCMD database are offered including searching from a set of controlled science keywords or using a free-text (which uses an implementation of the Z39.50 search and retrieval protocol). The GCMD also offers two Java-based interfaces for more advanced searching. Data set information can be retrieved and displayed in a number of different formats. Translation from DIF to several metadata formats permits greater flexibility in handling the needs of diverse communities that require varying levels of metadata detail.

3.0 Partnerships and Collaborations

Peter F. Drucker, a leading authority on management and leadership, said that organizations would survive only through alliances and flexible spheres of influence. He was speaking of business organizations, but the words can equally apply to information organizations such as the GCMD. The content and quality of the records within the

GCMD database could not be attained without the cooperation of many different organizations. All organizations have limited resources, budgets, and time. By combining resources and expertise, users can gain wider access to Earth science data.

3.1 An International Directory

Since 1987 the GCMD has been a partner in the Committee on Earth Observation Satellites (CEOS) International Directory Network (IDN) (http://gcmd.nasa.gov/ceosidn/idn_home.html). The GCMD is the North American Coordinating Node of the IDN and is primarily responsible for coordinating software installations and database content among all 15 IDN nodes worldwide. Most of the nodes serve directory-level Earth science content in the DIF metadata format, which are then passed to the GCMD as a centralized repository. The content is then replicated at other IDN nodes worldwide so scientists in those regions can have access to information in the directory. The result of this collaboration has been a dynamic and growing network of online Earth science information systems with recent additions from the United Nations Environment Programme (UNEP) and the Joint Committee on Antarctic Data Management (JCADM).

3.2 A Directory for Federal agencies

As part of NASA's commitment to the U.S. Global Research Program (USGCRP), the GCMD is a contributing partner to the Federal Global Change Data and Information System (GCDIS) (http://www.gcdis.usgcrp.gov/). Information and resources concerning global change science are shared among U.S. Federal agency participants with the GCMD providing dataset descriptions from nearly every Federal agency. The GCMD has over 1000 descriptions from NASA's Earth Observing System (EOS) Program (http://eospso.gsfc.nasa.gov/).

3.3 A Biological Partnership

In 1996, The Department of Interior's U.S. Geological Survey (USGS) Biological Resources Division (BRD) and NASA's GCMD entered a formal partnership to add information on existing biological and ecological databases to the GCMD and to the emerging National Biological Information Infrastructure (NBII) (http://www.nbii.gov/). Through this partnership, the two agencies are sharing resources to add dataset information to the GCMD and NBII databases and are cooperating in developing new techniques to assist users of both data systems to find existing sources of data and information applicable to their needs. In working together, the agencies are reducing overlapping costs and "by sharing expertise and resources, the scientific community and the public will have more efficient access to a broader range of information about existing sources of biological data" (Solomon and Gaines 1998)

Searching the metadata of the NBII Clearinghouse and those of the GCMD, scientists can locate and obtain ecological and biological data sets of interest. The collaboration

between BRD and GCMD takes advantage of the strengths of each of the organizations. The BRD has 17 science and technology centers with valuable data sets. By combining BRD data sets descriptions with the broader coverage within the GCMD database, scientists can have access to many more multidisciplinary data sets. Since the project started, approximately 400 USGS/BRD data set descriptions have been added to the GCMD.

3.4 Resource Discovery for Agriculture

In late 1998, the NASA GCMD and the U.S. Department of Agriculture (USDA) entered a partnership to create an Agricultural Online Data Directory (ARGOS) to support the USDA's Global Change Research program (http://agros.usda.gov). The initial focus of the collaboration is to identify, collect, describe, and provide access to USDA research data conducted on carbon sequestration and emissions in agricultural systems. GCMD staff has developed an online tool to access USDA-specific data set information. A comprehensive controlled vocabulary of agriculture-related keywords was developed to aid in searching and retrieving agriculture data set information from the GCMD and to aid in indexing agriculture data sets. Over 500 USDA data set descriptions have been added to the GCMD.

4.0 Other International Partnerships

In addition to the CEOS IDN partnership, the GCMD has been engaged in a number of other collaborations with international global change science projects. The Joint Committee on Antarctic Data Management (JCADM) (http://www.jcadm.scar.org/) has adopted the DIF as a metadata standard for cataloging an extensive collection of scientific research data sets from international researchers in the Antarctic. The JCADM, with the assistance of GCMD staff, is building an Antarctic Master Directory (AMD) which would contain a subset of the data set descriptions cataloged in the GCMD. The National Science Foundation (NSF) Office of Polar Programs (OPP) has recently stipulated that all NSF-funded Antarctic projects must develop metadata descriptions in the Directory Interchange Format for entry in the AMD/GCMD catalog (NSF, 1998).

Other international collaborations have begun with the International Geosphere-Biosphere Program (IGBP) to provide access to metadata from all of IGBP's core programs. Specifically, GCMD staff has already developed metadata authoring tools and format translation software for describing data held by the Joint Global Ocean Flux Study (JGOFS) program, a core project of the IGBP. A recent recommendation by another IGBP core project, PAst Global ChangES (PAGES), is to provide paleoenvironmental data descriptions in the Directory Interchange Format for inclusion in the GCMD (Anderson and Webb 1999).

5.0 Information Systems Partnerships

Not only has GCMD staff entered into partnerships to develop database content (the catalog records), there have also been efforts to form partnerships to explore state-of-the-art information systems development for searching and retrieving information from the database. For example, the GCMD staff worked closely with the University of Maryland's Human-Computer Interaction Laboratory (HCIL) to develop a Java-based dynamic query search and refinement system (Greene, et al 1999).

Recently, the GCMD entered into a collaboration with Purdue University and the San Diego Supercomputing Center (SDSC) to develop an object-oriented architecture for distributed systems that could have widespread applications within the International Directory Network (IDN)

6.0 Conclusions

Forming alliances and partnerships are vital to the mission of the GCMD and essential for access to Earth science data catalog records and Earth science data. By forming partnerships and collaborations, collaborators take on an ownership for data set catalog records that accurately describe their data holdings. Global climate change science is intensely multidisciplinary, bringing in all disciplines of Earth science from solid earth geophysics to ecosystem biology, from oceanographic research to solar-terrestrial impacts on climate. Cataloging Earth science data set records is an arduous task and, without partnerships, the availability of accurate, descriptive information would be limited. In the future, more collaborative partnerships are anticipated as data management requirements call for cataloged metadata records that will allow the community to search and discover Earth science data.

7.0 References

Anderson, D.M. and R.S. Webb, eds. 1999. *PAGES: The data guide. Results from the second workshop on global paleoenvironmental data*, February 9-12, 1998, Boulder, CO.

Greene, S., Tanin, E., Plaisant, C., Shneiderman, B., Olsen, L., Major, G., Johns, S. 1999. The end of zero-hit queries: Query previews for NASA's Global Change Master Directory, *International Journal on Digital Libraries* Vol. 2 No.2+3, pp.79-90

IPCC. 1995. *Climate Change 1995 - The science of climate change: Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, J.T. Houghton, L.G. Meira Filho, B.A. Challander, N. Harris. A. Kattenberg and K. Maskell (eds.), Cambridge University Press, Cambridge, UK

National Research Council (NRC). 1998. *Global environmental change: research pathways for the next decade*, National Research Council Committee on Global Change Research, Board on Sustainable Development Policy, Policy Division, National Academy Press, Washington, D.C.

National Science Foundation. 1998. *Guidelines and award conditions for scientific data.* Office of Polar Programs, National Science Foundation, Arlington, VA. (Online at: http://www.nsf.gov/cgi-bin/getpub?opp9)

Solomon, C. and J. Gaines. 1998. *The collaboration of the U.S. Geological Survey Biological Resources Division and the NASA Global Change Master Directory in writing and serving metadata*, presented at the Ecological Society of America – 1998, Baltimore, MD, August 2-6, 1998. Symposium Abstracts, p. 218.

Acknowledgement

The author gratefully acknowledges the NASA Global Change Master Directory and the NASA GCMD Project Manager, Lola Olsen, Code 902, NASA/GSFC, Greenbelt, MD 20770, olsen@gcmd.nasa.gov