

**Report to the NOAA Science Advisory Board**

from the

Data Access and Archiving Requirements Working Group

March 2011

The SAB's Data Access and Archiving Requirements Working Group (DAARWG) met in December 2010 after nearly two years without a meeting, and four years after its first meeting. DAARWG members thus had the opportunity to evaluate changes in NOAA data-management activities over the period. We were pleased to find improvements in a number of areas. We noted evolution from a stovepipe-dominated data-management environment to one with improved collaboration between NOAA elements.

NOAA is making progress towards defining an architecture and approach, along with policies and procedures to provide integrated data services. NOAA has explicitly included the generation of environmental information in its mission and vision. This by its nature requires integration of data from a host of specialized research areas and projects. Stimulated by the *Deepwater Horizon* oil spill, NOAA shows awareness that efficient data-management procedures will better prepare it for the next natural disaster.

The latest and emerging generation of researchers and data information engineers is a product of the Internet. Though many in NOAA management were not trained in this way, they use the web and expect products and information to be available in that form. The organization, its people, and their clients expect NOAA's environmental information to be available with simple point-and-click access. The technology available for providing such access and for integrating science data to provide environmental information has matured rapidly. However, there is still a lot to do. NOAA should not lose focus before the task is complete.

We note progress in a number of programs, several of which are driven by bottom-up "volunteer" efforts. In fact, DAARWG members wonder if NOAA achieves many of these advances despite its organization rather than because of it. We believe that assuring the importance of data systems in NOAA priorities still needs improvement.

**Fisheries Enterprise Data Management**

DAARWG members were impressed by the progress of the NOAA Fisheries Enterprise Data Management (EDM) program. This is a bottom-up activity driven by needs identified by internal users.

The National Marine Fisheries Service must deal with a vast variety of biological, physical, and chemical, and economic data types, many of which are difficult to quantify. The Fisheries Information Advisory Committee is using a new model of collaboration to establish data-management policies, procedures, architecture, and tools, and to foster documentation, discovery, access, and understanding of the array of information resources within NOAA Fisheries.

The DAARWG was pleased to see the development of directory-level information in the NMFS data system. We believe that this effort could be further enhanced by proactively collaborating with other directory efforts both within and beyond NOAA. Specifically, we encourage NMFS to connect their directory development to existing community catalogs such as NASA's Global Change Master Directory to make sure that not only are the descriptions of Fisheries Service data sets available on the web, but that they are readily ingestible into external archives.

In sum, DAARWG believes that the Fisheries EDM should extend beyond solely meeting the data management requirements internal to NOAA Fisheries. Thus, DAARWG makes the following recommendation:

**DAARWG recommends that approaches used in the Fisheries Enterprise Data Management system should be developed with a broader base of use in mind, to become fully inter-operative with other data management programs both within and outside of NOAA.**

### **Global Earth Observation Integrated Data Environment (GEO-IDE)**

This program to develop a system of systems is making good progress. It is working to develop a framework that provides integration of many quasi-independent systems. If successful, it will break down stovepipes between NOAA elements that have been barriers to needed collaborative work. The organizing principle uses integration across domains [atmosphere, ocean, climate, ecosystem], across functions [discovery, access, interoperability, preservation], and by structural data type [grids, profiles, trajectories, points].

### **Unified Access Framework (UAF)**

The UAF is using and engaging evolving national and international interoperability standards. Their approach allows interoperability of data within NOAA as well as outside NOAA by other agencies and international data groups. They focus not only on data discovery, but also data access, use, and understanding. DAARWG members have been impressed with the progress made during two years and hope this effort will be continued and established as a best practice within NOAA.

## Data documentation and metadata

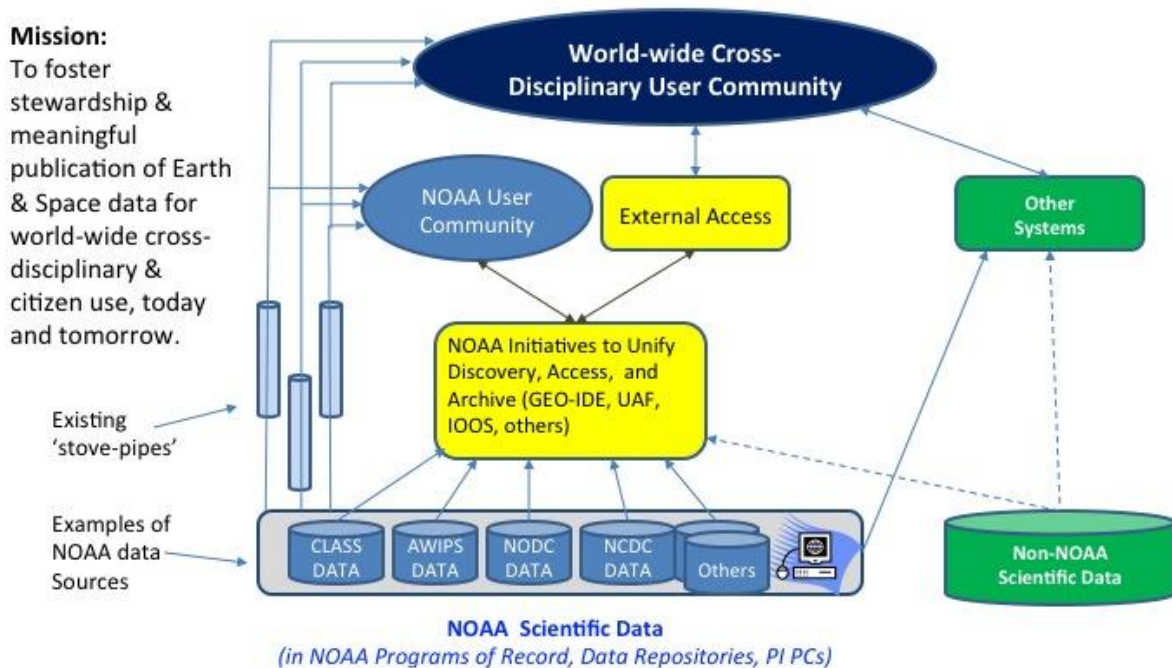
The term “metadata” means different things to different people. The term can address a broad range of functions including data discovery, data use, information required to understand the data, and their pedigree in the distant future. The emphasis that NOAA gives to the function of their metadata shows how NOAA views the importance of different user communities.

DAARWG is pleased to see a shift in emphasis from metadata that is intended for data discovery and long-term archival to metadata that serves the complete range of data user needs. Specifically, we applaud the increased importance attached to metadata intended to facilitate the *use* of data, which goes to the heart of the needs of today’s scientific community.

## DAARWG Context Diagram

- The diagram is illustrative of the context in which NOAA’s status and plans for unified discovery, access, and archiving of all NOAA scientific data, can be assessed.

## DAARWG Context Diagram



### ***Approval of new DAARWG members***

We propose and seek approval from the SAB for three DAARWG appointments. Each of the nominees brings strengths that enhance the working group's abilities to deal with the diversity of data and information activities in NOAA. They will replace members who have rotated off. (The DAARWG renews about one third of its appointment each year.) For each of these three, we propose a term to expire on 31 December 2013.

#### **Daphne Fautin**

Daphne Fautin, Ph.D., is Professor, Invertebrate Zoology, at the University of Kansas. She received a Ph.D. from the University of California, Berkeley.

Dr. Fautin has already served one term with the DAARWG and this proposal is appoint her for a second term.

Dr. Fautin has accepted a short-term position at the National Science Foundation and will be there for the next one or two years. This appointment should be an opportunity to improve NOAA/NSF collaboration since both agencies are active in promoting the development of new programs to improve the scientific use of data and information. With her expertise in marine life and biodiversity, she will enhance the DAARWG's ability to deal with issues in the National Marine Fisheries Service.

Her research interests include

- Studies on sea anemones and organisms that live with sea anemones, including fishes, crustaceans, and algae.
- Biodiversity
- Habitat specificity

Dr. Fautin serves on the International Steering Committee of the Ocean Biogeographic Information System, the US National Committee for the International Union of Biological Sciences, and chairs the US National Committee for the Census of Marine Life.

#### **Chris Lenhardt**

Chris Lenhardt, M.A., is Informatics Scientist, Environmental Data Science and Systems, Oak Ridge National Laboratory, Oak Ridge, Tennessee. He received an MA in Political Science from the University of Michigan.

He is manager of the Distributed Active Archive Center (DAAC) for Biogeochemical Data at ORNL. The DAAC is a NASA-funded data center and is part of NASA's Earth Observing System. He previously had been Deputy Manager of the Socioeconomic Data and

Applications Center (SEDAC), at the Center for International Earth Science Information Network (CIESIN) as well as Associate Director of CIESIN, where he managed metadata, archiving, and data management activities.

At ORNL Chris supervises activities in a number of research areas:

- Earth and aquatic sciences
- Ecosystem science
- Environmental data science & systems
- Renewable energy Systems
- Human health risk & environmental analysis
- Energy-water-ecosystem engineering

### **Beth Plale**

Beth Plale, PhD, is Associate Professor of Computer Science at Indiana University, Bloomington, Indiana. She also holds the position of Director, Center for Data and Search Informatics, and Director, Data to Insight Center of Pervasive Technologies Institute. She received a Ph.D. from the State University of New York (SUNY).

Professor Plale is an experimental computer scientist whose research interests are in data management, data-driven computing, and the long-term preservation of digital data. She has a strong research interest in metadata and provenance of digital scientific data particularly for purposes of long term preservation and focuses on "the first mile" where collection is automatic and close to the generation source.

Plale is deeply engaged in environmental and atmospheric science research and has substantive experience in developing stable and useable scientific cyberinfrastructure. Her current research areas are data provenance and metadata, digital preservation of complex scientific data objects, workflow systems support for real time observational data, and complex events processing. She is an ACM Senior Member and IEEE Member.