

# Data Archiving and Access Requirements Working Group (DAARWG)

---

Report to the  
Science Advisory Board  
National Oceanic and Atmospheric Administration

Presented by  
Ferris Webster, chair, DAARWG

9 March, 2011

# DAARWG update

---

- Met in December 2010 after nearly two years without meeting.
- Had an opportunity to evaluate data-management changes over the period.
- Found improvements in a number of areas.
- Noted improved collaboration between NOAA elements.
  - evolution away from a stovepipe-dominated data-management environment

# NOAA has...

---

- ...made progress towards defining an architecture to provide integrated data services.
- ...included the generation of environmental information in its mission.
- ...learned from the *Deepwater Horizon* oil spill
  - efficient data-management will better prepare it for the next natural disaster.
  - there is need to integrate data from a host of specialized research areas and projects.

# Access technology is changing

- NOAA and its clients expect environmental information to be available with a simple point-and-click.
- Technology to provide such access and to integrate 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.

# Signs of progress

---

- DAARWG noted progress in a number of programs.
  - Excellent bottom-up “volunteer” efforts drive many of these.
  - Assuring the importance of data systems in NOAA priorities still needs improvement.
- These programs have attributes that DAARWG would like to see in other NOAA data management elements.

# Fisheries data management

---

- Progress of the Fisheries Enterprise Data Management (EDM) program impressed us.
- NMFS deals with a variety data types
  - biological, physical, chemical, economic, ...
  - many are difficult to quantify.
- Fisheries Information Advisory Committee uses a new model of collaboration
  - establishes data-management policies, procedures, architecture, and tools
  - fosters documentation, discovery, access, and understanding of Fisheries information.

# Fisheries directories

---

- We were pleased to see development of data directories in the NMFS data system.
- This activity could be enhanced by collaborating with other directories both within and beyond NOAA.
  - we encourage NMFS to connect their efforts to existing community catalogs such as NASA's Global Change Master Directory.
  - make descriptions of Fisheries data sets not only available on the web, but but also readily ingestible into external archives.

# A recommendation

---

- We believe that the Fisheries EDM should extend beyond solely meeting needs internal to NOAA Fisheries.
- We make 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.



# GEO-IDE\*

---

- This program to develop a system of systems is making good progress.
  - It is developing a framework to provide 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 approach uses integration across domains and across functions by data type as an organizing principle.

\*Global Earth Observation Integrated Data Environment

# Unified Access Framework (UAF)

- Individuals from many parts of NOAA have created a bottom-up project: UAF. They...
  - ...adopt shared approaches to achieve data interoperability within NOAA as well as with other agencies and international data groups.
  - ...use evolving national and international interoperability standards.
  - ...focus not only on data discovery, but also data access, use, and understanding.
- We are impressed with their progress
  - hope this effort will be continued and established as a best practice within NOAA.

# DAARWG Context diagram

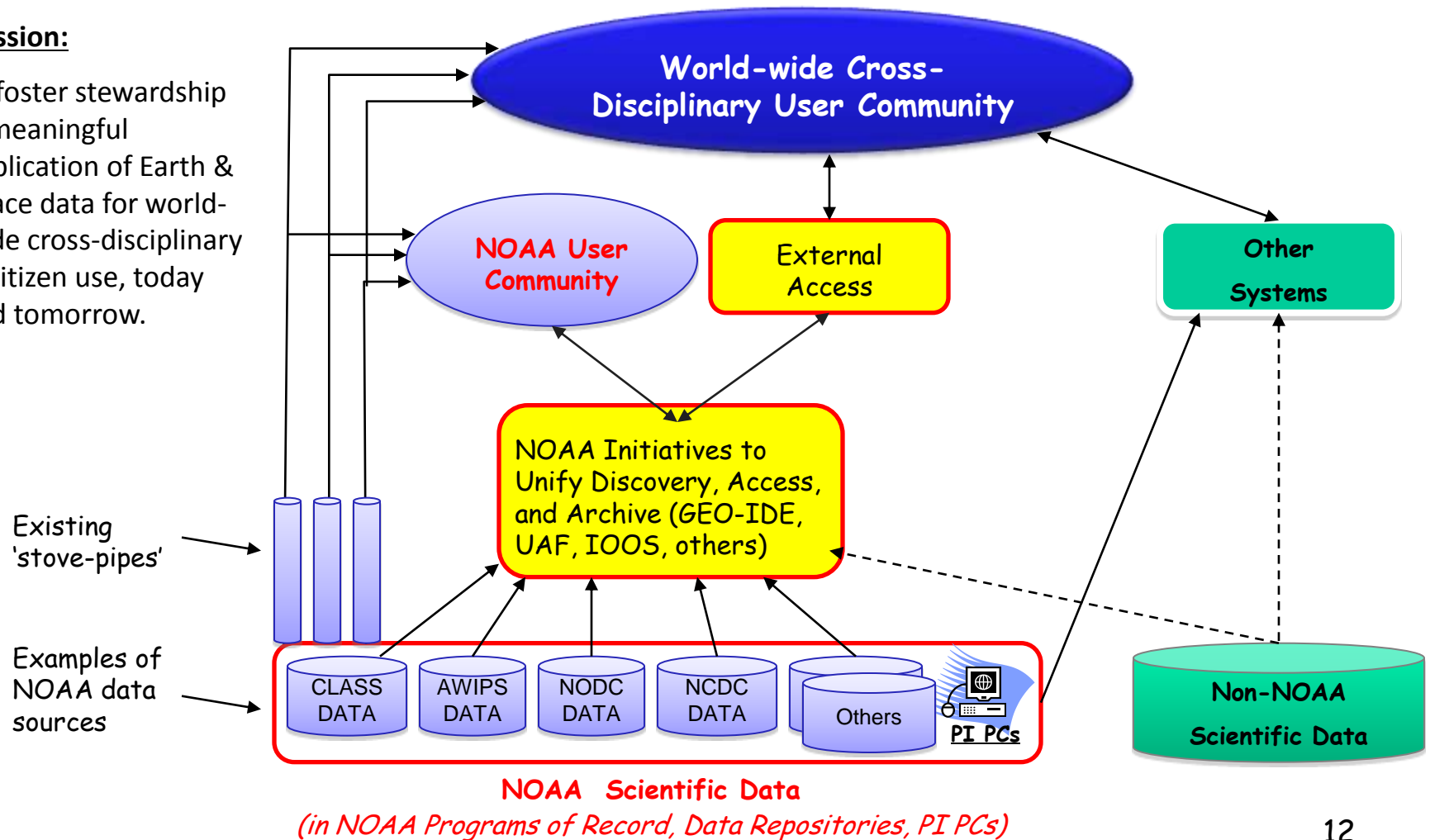
---

- The figure in the next chart is an illustration 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

## Mission:

To foster stewardship & meaningful publication of Earth & Space data for world-wide cross-disciplinary & citizen use, today and tomorrow.



# Data documentation and metadata

---

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

# Data documentation and metadata

---

- We are pleased to see a shift in emphasis from metadata intended for data discovery & long-term archival to metadata that serves the complete range of data-user needs.
- We applaud the increased importance attached to metadata intended to facilitate the *use* of data.
  - This goes to the heart of the needs of today's scientific community.

# DAARWG appointments

---

Three appointments are proposed: two replace members who have rotated off. Fautin is a renewal. Their terms are to run until 31 December, 2013. They are:



- **Daphne Fautin**, University of Kansas
  - Biodiversity, invertebrate biology, marine organisms



- **Chris Lenhardt**, Oak Ridge National Laboratory
  - Earth and environmental sciences, ecosystem science, environmental data science & systems



- **Beth Plale**, Indiana University
  - data-driven computing, long-term preservation of digital data, collaboration with earth and atmospheric sciences.

# Daphne Fautin, U of Kansas

---

- Professor, Invertebrate Zoology
- Will be at NSF for the next year or two
  - An opportunity to improve NOAA/NSF collaboration
- Ph.D., University of California, Berkeley
- Research interests:
  - Sea anemones and organisms that live with sea anemones, including fishes, crustaceans, and algae.
  - Biodiversity
  - Habitat specificity.
- This will be her second term as a DAARWG member.



# Chris Lenhardt, Oak Ridge N.L.

---

- Informatics Scientist. Environmental Data Science and Systems, Oak Ridge National Laboratory
- MA, Political Science, University of Michigan
- Manager, NASA DAAC for Biogeochemical Data at ORNL
- Research interests:
  - Environmental data science & systems as applied to climate science, climate impacts and adaptation, human dimensions of global change, Earth and environmental sciences, ecosystem science, renewable energy systems
  - Science 2.0, Web 3.0
  - Ethical, legal, and social issues associated with emerging science and technology

# Beth Plale, Indiana University

---

- Director, Center for Data and Search Informatics
- Director, Data to Insight Center of Pervasive Technologies Institute
- Ph.D., State University of New York (SUNY)
- Research interests:
  - data management
  - data-driven computing
  - long-term preservation of digital data.
  - Problems involving interdisciplinary collaboration with earth and atmospheric sciences.

# Existing DAARWG members

---



- **Roberta Balstad**, 2012, Columbia University,
  - Societal impacts of climate change



- **John Boreman**, 2013, North Carolina State Univ.
  - Population dynamics and environmental impact assessments of coastal & estuarine fishes.



- **Peter Cornillon**, 2012, University of Rhode Island
  - Satellite oceanography, distributed data systems



- **Sara Graves**, 2012, U. of Alabama in Huntsville
  - Large-scale distributed information systems

# Existing DAARWG members

---



- **Ernest Hildner**, 2012, Boulder, Colorado
  - Space environment research



- **Phil Jones**, 2011, University of East Anglia, UK
  - Climate change, paleoclimatology



- **Krish Narasimham**, 2013, Lockheed Martin Corporation
  - Enterprise IT, data management, communication networks.



- **Roger Wakimoto**, 2011, NCAR
  - Atmospheric sciences; severe weather

# Existing DAARWG members

---



- **Ferris Webster**, 2012, U of Delaware
  - Physical oceanography, ocean data systems



- **Bruce A. Wielicki**, 2011, NASA Langley Research Center
  - Cloud layer properties



- **Dean Williams**, 2011, Lawrence Livermore National Laboratory
  - Climate modeling. IGBP model archive



- **Thomas Zacharia**, Oak Ridge National Laboratory
  - (SAB Liaison)

# DAARWG terms

Name (current term end date)	2011	2012	2013
Roberta Balstad		12/2012	
John Boreman			12/2013
Peter Cornillon		12/2012	
Sara Graves		12/2012	
Ernest Hildner		12/2012	
Phil Jones	12/2011		
Krish Narasimham			12/2013
Roger Wakimoto	12/2011		
Ferris Webster		12/2012	
Bruce Wielicki	12/2011		
Dean Williams	12/2011		
Daphne Fautin (proposed)			12/2013
Chris Lenhardt (proposed)			12/2013
Beth Plale (Proposed)			12/2013

# WG Terms of Reference

---

- Provide scientific advice and broad direction to NOAA regarding the wide range of data, information, and products that NOAA should archive and how NOAA can best provide access to this information.
- The Data Archiving and Access Requirements (DAAR) Working Group will evaluate data archiving and access requirements from all of NOAA's observing systems and computational models, as well as non-NOAA information.