

A review of DAARWG issues

Agenda topics at the previous eight DAARWG meetings show a pattern of concerns with major NOAA data issues.

A review of past activities may help guide future DAARWG work.

Presented by
Ferris Webster, chair, DAARWG

27 June 2012

DAAR 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.

1. December 2006

NOAA presented the following issues for consideration:

- **GEO-IDE** (An architecture for integrated observing, data-processing, & information-management systems as being addressed by NOAA's Global Earth Observation Integrated Data Environment)
- **Size** (Managing exponentially growing data volumes)
- **Metadata** (Appropriately describing data to ensure long-term utility)
- **Integration** (Providing data in standard formats and protocols to enable integration)
- **Access** (Clear and easy discovery of and access to data and products)
- **Usability** (Assisting data use)
- **At-risk data sets** (Collecting data at risk to extend the environmental record)

DAARWG identified 3 issues:

Comprehensive Large-Array Stewardship System (CLASS)

“The WG sensed that the model for CLASS was not clearly defined. Shouldn’t the aim be clarified before a lot of work is done in designing and building a system? Part of the clarification is the connection of CLASS to NOAA’s Mission Objectives. “Don’t build the house, and then hire the architect”.

Is CLASS for all data in NOAA? The system architecture to handle large arrays of satellite data may need to be significantly different from that needed to handle fisheries biological samples.”

Global Earth Observation Integrated Data Environment (GEO-IDE)

What are links between GEO-IDE and other programs in NOAA, other agencies, and internationally?

What resources are available to support GEO-IDE development, and are these sufficient for reasonable progress?

Integration

Coordinate NOAA’s multiple data system to provide access with uniform procedures have having standard formats and protocols

2. May 2007

- Archive policy
 - multiple versions
 - non-NOAA data
- CLASS
- GEO-IDE
- Coping with data volume

3. January 2008

- National Research Council report on environmental data management at NOAA
- CLASS
- GEO-IDE
- NPOESS
- Archiving policy workshop

4. June 2008

- What-to-archive procedure (NAO 212-15)
- NOAA active archive (GEO-IDE)

5. February 2009

- Data-retention policy
- NOAA centers of data
- Information heterogeneity

6. December 2010

- Data-sharing policy for NOAA grants
- NMFS Data Management Planning
- Archive Architecture Concept of Operations
- GEO-IDE
- Integrated Ocean Observing System (IOOS)
- Unified Access Framework (UAF)
- Data Documentation
- CLASS
- Climate Service

7. June 2011

- External dataset policy
 - NMFS
 - NWS
 - IOOS
 - National Climate Assessment

8. November 2011

- GEO-IDE, UAF
- Active Archive
- Data documentation
- National Earth Observation Strategy
- Data Management Framework
- Data management planning
- Data Sharing Policy for NOAA Grants
- Data Management Vision

Overarching issues

a. Structural issues

- Coping with data inhomogeneity
- Data integration and consolidation
- Legacy systems, centers of data
- Data for research vs. operations
- Allocation of resources

Overarching issues

b. Technical issues

- Metadata/documentation
- What to archive
- External (non-NOAA data)
- Data-sharing policy
- Data discovery and access

Conclusions

- Structural/policy issues are inherent and difficult to resolve. Most have been on the DAARWG agenda since the first meeting, and are likely to remain as issues of concern. Ultimate solutions will be difficult to obtain. DAARWG can usually only recommend incremental improvements.
- Technical issues can usually be dealt with. The expertise of DAARWG members may help NOAA staff in finding appropriate solutions.