Improving Environmental Data and Information Management in NOAA

Prepared for the Data Archiving and Access Requirements Working Group (DAARWG)

Helen Wood, EDMC Chair
Deirdre Jones, EDMC Deputy Chair

December 8, 2010
Introduction and Motivation
Recent High Visibility Drivers
• Deep Water Horizon Incident
• Ocean Policy Task Force Recommendations
• NOAA Climate Service
Role of Environmental Data Management Committee
Summary of FY 2011 Activities
Next Steps
“Environmental information is critical to achieving the objectives of all of NOAA’s goals.

NOAA is, at its foundation, an environmental information generating organization.”

- From the Next Generation Strategic Plan, 2010
Broad Scope for Environmental Data Stewardship
- ~150 Research & Operational Observing Systems
- ~4-5 Petabytes of data/year (~15 Pb total)

Data Management Challenges are Changing
- No longer just about data volume
- Data discovery and integration
- Data stewardship and information
Deep Water Horizon Lessons Learned

Need for:
• Clear and consistent data management policy
• Better data documentation (metadata)
• An overarching response plan

Immediate post-incident
Data management challenge
• Quickly identify applicable / available data

Pre-incident (routine) data management challenges
• Volume and diversity of data
• Collected for one-use; potential for multi-use
• Infrastructure needed to support multiple access mechanisms

Post-incident data management challenges
• Releasability of data (organizational vs. technical)
• Coordination of collection and distribution
• Comparison and utilization of modeling outputs and observations

Prepared by Environmental Data Management Committee
**ACTIONS: Data and Tools Theme Team**

1. Develop and implement a comprehensive strategy for identifying and integrating relevant NOAA data sets to support regional Coastal and Marine Spatial planning.
2. Contribute to the design and creation of the interagency NIMS.
3. Work with NOAA and partners to assess, enhance and provide access to robust and flexible analytical tools to evaluate alternative ocean use scenarios.

**ACTIONS:**

1. Develop source to repository data management system.
2. Adopt and implement common standards for data transport, metadata, data models, and data discovery.
NCS Core Capability 1: Observing Systems, Data Stewardship, and Monitoring

- NCS users will obtain *easy* and *timely access* to the nation’s trusted data and information about the current state of the climate system in context with the past.
Principles for Effective Environmental Data Management

1. Data should be archived and accessible
2. Adequate resources for end-to-end management
3. Management activities should involve users
4. Interagency and international partnerships
5. Metadata are essential
6. Expert stewards required for management
7. Process to decide what data to archive
8. Archive must support discovery, access, and integration
9. Effective management requires a formal, ongoing planning process

National Research Council Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA, 2007
“This [effective utilization of NOAA’s environmental information] will only be possible if operational and scientific users and environmental information providers work collaboratively.” — From the NGSP, 2010

- EDMC kicked off in January 2010
- Good cross-NOAA participation
- Working collaboratively with other NOSC and CIO Council Committees
- Focus on harmonizing, leveraging, and integrating effective data management efforts across NOAA
- Advised by NOAA SAB DAARWG
Revision of NOAA-wide Policy (NAO) 212-15

Environmental data will be visible, accessible and independently understandable to users, except where limited by law, regulation, policy or by security requirements.

• Maintains NOAA’s policy of “full and open access” to environmental data

• Provides mechanism for EDMC to develop procedural directives for more detailed guidance

• Presents an end-to-end lifecycle framework for the management of environmental data

Prepared by Environmental Data Management Committee
As specified in revised NAO 212-15

- Determining what environmental data are required to be preserved for the long term and how preservation will be accomplished
- Developing and maintaining metadata throughout the environmental data lifecycle that comply with standards
- Obtaining user requirements and feedback
- Developing and following data management plans that are coordinated with the appropriate NOAA archive for all observing and data management systems
- Conducting scientific data stewardship to address data content, access, and user understanding
- Providing for delivery to the archive and secure storage
- Providing for data access and dissemination
- Enabling integration and/or interoperability with other information and products
### FY 2011 EDMC Operating Plan Activities

**End-to-End Environmental Data Management Functions**

<table>
<thead>
<tr>
<th>Planning</th>
<th>Observation Acquisition and Transmission</th>
<th>Scientific Data Management</th>
<th>Archive and Access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maintain/Monitor</td>
<td>Collect/Rescue</td>
<td>Calibrate/Validate</td>
</tr>
</tbody>
</table>

**Contingency Planning**

- **Develop Data Management Plan Guidance**
- **Improve Data Documentation**
- **Develop Archive Architecture ConOps**
- **Enhance Data Access Activities**
- **Develop Data Sharing Policy for NOAA Grants**

Prepared by Environmental Data Management Committee
### FY 2011 EDMC Activities

- **= work in progress**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Long term preservation (what &amp; how)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop and maintain metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developing and following data management plans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scientific data stewardship</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery to the archive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data access</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration and/or interoperability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Prepared by Environmental Data Management Committee
External Engagement

GEO Group on Earth Observations

Open Geospatial Consortium

Federation of Earth Science Information Partners

Making Data Matter

IOC

NSTC

WMO

Prepared by Environmental Data Management Committee
Next Steps

- Roll out new policies and procedures
- Entrain NOAA programs
- Assess progress
- Address gaps

Prepared by Environmental Data Management Committee
Improving Environmental Data and Information Management in NOAA

Prepared for NOAA Leadership by the Environmental Data Management Committee

BACKUP SLIDES
Purpose

The Environmental Data Management Committee (EDMC) coordinates the development of NOAA’s environmental data management strategy, and policy, and provides guidance to promote consistent implementation across NOAA, on behalf of the NOSC and CIO Council. Environmental data management is an end-to-end process that includes acquisition, quality control, validation, reprocessing, storage, retrieval, dissemination, and long-term preservation activities. The goal of the EDMC is to enable NOAA to maximize the value of its environmental data assets through sound and coordinated data management practices.
<table>
<thead>
<tr>
<th>Member Name</th>
<th>Role</th>
<th>Representing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helen Wood</td>
<td>Chair</td>
<td>NOAA</td>
</tr>
<tr>
<td>Deirdre Jones</td>
<td>Deputy Chair</td>
<td>NOAA</td>
</tr>
<tr>
<td>David Layton</td>
<td>Member</td>
<td>NOAA (Enterprise Architect)</td>
</tr>
<tr>
<td>Ken McDonald</td>
<td>Member</td>
<td>NOAA (Data Architect)</td>
</tr>
<tr>
<td>Scott Hausman</td>
<td>Member</td>
<td>NCS</td>
</tr>
<tr>
<td>Selina Nauman</td>
<td>Member</td>
<td>NESDIS</td>
</tr>
<tr>
<td>Harry Tabak</td>
<td>Member</td>
<td>NWS</td>
</tr>
<tr>
<td>Tony LaVoi</td>
<td>Member</td>
<td>NOS</td>
</tr>
<tr>
<td>Chris Miller</td>
<td>Member</td>
<td>OAR</td>
</tr>
<tr>
<td>Jim Sargent</td>
<td>Member</td>
<td>NMFS</td>
</tr>
<tr>
<td>Cecile Daniels</td>
<td>Member</td>
<td>OMAO</td>
</tr>
<tr>
<td>Shanna Pitter</td>
<td>Member</td>
<td>PPI</td>
</tr>
</tbody>
</table>

Prepared by Environmental Data Management Committee
The primary focus in FY 2011 will be to establish detailed guidance on utilizing metadata standards across NOAA.

NOAA will be encouraging the use and convergence on International Standard Organization (ISO) Metadata Standard.

Detailed guidance to be built on NMFS Data Documentation Procedural Directive which is nearing completion.

NOAA’s Rolling Deck to Repository (R2R) is an example of end-to-end data documentation.
Selected FY 2011 EDMC Operating Plan Activities
Develop Archive Architecture ConOps

Open Archival Information System Reference Model (OAIS-RM) Functions

• NOAA is following the International Standard for information preservation (OAIS-RM)
  • This standard identifies the functions required to provide long-term preservation

• NOAA has multiple organization and systems fulfilling these responsibilities

• The Concept of Operations will establish the enterprise process and procedures for determining who does what
Selected FY 2011 EDMC Operating Plan Activities

Enhance Data Access Activities

Client Tools / Applications

NOAA Portals / Websites

GEOSS Portal

Standard Access Services

NOAA’s Environmental Data & Information Infrastructure

NOAA Metadata Records

Prepared by Environmental Data Management Committee
Listed FY 2011 EDMC Operating Plan Activities

Develop Data Sharing Policy for NOAA Grants

Dissemination and Sharing of Research Results

NSF Data Sharing Policy

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grants are expected to encourage and facilitate such sharing. See Proposal & Administration Guide (NSF Chapter II.C.4.A) for full policy implementation.

NSF Data Management Plan Requirements

Beginning January 16, 2011, proposed awards to new projects must include a supplementary document of no more than two pages titled "Data Management Plan." This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Proposal & Administration Guide (NSF Chapter II.C.4.A) for full policy implementation.

Requirements by Directorate, Office, Division, Program, or other NSF Unit

Links to data management requirements and plans relevant to specific Directorates, Offices, Divisions, Programs, or other NSF units are provided below. If guidance specific to the program is not provided, then the requirements established in Proposal & Administration Guide (NSF Chapter II.C.4.A) shall apply.

Please note that specific program information provides guidance on preparation of data management plans, such guidance must be followed.

- Engineering Directorate (ENG)
  - Directorate-wide Guidance
- Geosciences Directorate (GEO)
  - Division of Earth Sciences
  - Geosciences Data Sharing Program
- Social, Behavioral, and Economic Sciences Directorate (SBSE)
  - Data Sharing for the Division of Social and Economic Sciences (DEE)

Data Management & Sharing Frequently Asked Questions (FAQs)

Email Print Share