Evolving Requirements for Data and Information Management

David McKinnie
Office of Ocean Exploration and Research

Sharon Mesick
National Centers for Environmental Information

Megan Cromwell
Northern Gulf Institute/Mississippi State University
National Centers for Environmental Information
OER Mission and the Data Imperative

• OER’s mission is to explore the deep ocean:
  – map to modern standards
  – conduct baseline characterization
  – make that data and information available *and accessible* to whomever needs it in a form that is *useful* for science, for decisions, for general information about the ocean

• If ocean exploration results are not available, accessible, and reusable, we have not succeeded in our mission
Data Management: Evolution in a changing landscape

- 2000: OE / NOAA Data Center Agreement
- 2001: Data Quality Act
- 2002: OE Data IPT
- Seminal Documents
- 2005: OE / NOAA Data Center Agreement
- 2007: OE / NURP Merger
- 2009: Ocean Exploration Act
- 2013: Open Data Access Policy
- 2015: NCEI: NOAA Data Center Consolidation
- 2015: NOAA Public Access Policy (PARR)
- 2019: Open Government Data Act

Expedition Portfolio by Year:
- Partnerships / NOAA Ship Okeanos Explorer / Competitive Grants
Evolving View of Characterization

- **2008-2015**
  - Sonar (multibeam/mapping operations)
  - Video (ROV operations)
  - Some oceanographic data

- **2015-2018**
  - Water Column?
  - Acoustics?
  - Other measurements?

- **2019**
  - Processes
  - Connectivity
  - Rapid evolution
OER’s conceptual model for an expedition

- We explore a 3-D polygon
- We need to think in terms of:
  - How all data collected by the expedition in that polygon might relate
  - Developing new CONOPS using gliders, other platforms and sensors (Deep Argo? Saildrone? Silixia?)
  - Integrating data collected in that polygon in an extended time domain (e.g. moored arrays, Argo floats).
• NOAA Ship *Okeanos Explorer* retires in 2023-24. There is no dedicated replacement; OER will begin operate from multiple platforms soon

• By 2025, OER must be heavily invested in autonomy

• The new CONOPS that will result mean much more complex and diverse data work flows, more diverse data types, and radically increased data volume
SCIENCE AT SEA

THE PAST

SCIENTIFIC PAPERS

DATA AVAILABILITY: 2+ YEARS
THE PRESENT

REAL-TIME SCIENCE ON SHORE

OPEN ACCESS DATA

DATA AVAILABILITY: 6 WEEKS
THE FUTURE

DATA AVAILABILITY: REAL-TIME

SCIENCE ANYWHERE

TWO-WAY INTERACTION
National Centers for Environmental Information: Making Ocean Exploration Data Available, Accessible, and Useful
Outward-facing Data Management
Community of Practice

LEARN

INNOVATE

DEVELOP

LEVERAGE TECHNOLOGY

COLLABORATE
Learn: Creating and Archiving Born-Digital Video

Federal Agency Digitization Guidelines Initiative
Audio-Visual Working Group

- Library of Congress
- National Archives and Records Administration
- Smithsonian Institution

Fiscal ‘15 Report

- Federal Case Studies (OER one of eight case studies)
- Recommendations
- Resources

Learn: National Science Foundation

2016 Video Workshop: Establishing Community Standards for Underwater Video Acquisition, Tagging, Archiving, Access

- Recording
- File Naming Conventions
- Video Formats
- Compression
- Metadata Priorities
- Timecode embedding
- Audio Channels
- Annotation / Event Logs
- Archive / Open Access

Original (Off-line)
• Program saves original

Full Archive (Near-line)
• Authoritative Standardized version

Proxy (Online)
• Rapid public access
**Innovate: Video Data Management**

1. User discovers video through metadata catalog search
2. Low-Res Video retrieved & previewed from NCEI archive
3. Full-res order placed; ftp link returned to user email
Innovate: Video Data Management

- Seatube/ONC
  - Annotations
  - Environmental variables / ROV sensors
  - Video

- Timecode
  - National Centers for Environmental Information

- Value Added
  - Rich ISO Metadata
    - Available
    - Accessible
    - Useable
Innovate: Video Data Management
Innovate: Sharing Video Management Approaches

• Lessons learned
  - Video guidance shared across NOAA
  - Best practices and templates

• Legacy data integration
  - Career collections and key expeditions available online
  - Recovery from disparate media
  - Data now available in standard formats

• Benefits for Future
  - Legacy data recovery has implications for machine learning
  - Reinforces need for data architecture in advance
Develop: End-to-end Sample Data Management
Develop: End-to-end Sample Data Management

“Museum samples are only as valuable as the data associated with them. The Okeanos collected material has extremely high scientific value because of the finely detailed electronic data associated with it. SODA ensures that all this data remains associated with the proper samples and streamlines importing that information to the Smithsonian Natural History Museum catalog with minimal (if any) errors, where that data is publicly available to anyone interested. As a result, the Okeanos samples remain high quality for scientific studies performed now and well into the future.” – Dr. Abigail Reft, Smithsonian National Museum of Natural History
Develop: End-to-end Sample Data Management

Sample Operations Database Application

GIS SERVICES

Smithsonian National Museum of Natural History

Oregon State University MARINE AND GEOLOGY REPOSITORY
Supporting Earth, Ocean, and Antarctic Sciences

Benthic Deepwater Animal Identification Guide

Sample management
Electronic records
Public access to physical and digital data
Develop: Sample Data Access

OceanExplorer.NOAA.gov
Museum samples are only as valuable as the data associated with them.

The *Okeanos* collected material has extremely high scientific value because of the finely detailed electronic data associated with it.

SODA ensures that all this data remains associated with the proper samples, and streamlines importing that information to the Smithsonian Natural History Museum catalog with minimal (if any) errors, where that data is publicly available to anyone interested.

As a result, the *Okeanos* samples remain high quality for scientific studies performed now and well into the future.

Dr. Abigail Reft,
Smithsonian National Museum of Natural History
Develop: Sample Data Analysis

https://oceanexplorer.noaa.gov/okeanos/animal_guide/animal_guide.html

Benthic Deepwater Animal Identification Guide
<table>
<thead>
<tr>
<th>Region</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-JNA</td>
<td></td>
</tr>
</tbody>
</table>

**Polynoidae**
D2-EX1504-L4-06-01:27:22

<table>
<thead>
<tr>
<th>phylum</th>
<th>class</th>
<th>order</th>
<th>suborder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annelida</td>
<td>Polychaeta</td>
<td>Phyllodocida</td>
<td>Aphroditiformia</td>
</tr>
</tbody>
</table>
OER is a key partner for the DSCRTP. As the major NOAA funder of deep-sea exploration and research, OER’s expertise, research, and information products are central to the DSCRTP mission. As a management-oriented research program, the DSCRTP complements OER’s work and directly links it to resource managers’ needs.

Together, our programs have been able to leverage additional ship time, make connections between exploration and management needs, and achieve more together than either program could alone.

*Thomas Hourigan, Ph.D.*  
*Chief Scientist,*  
*Deep Sea Coral Research and Technology Program*
Leverage: NOAA’s GeoPlatform

https://noaa.maps.arcgis.com/home/index.html
Leverage: NOAA’s GeoPlatform

New geospatial web services

Shaded relief imagery, 3D visualization

Near real-time updates, automatically updated when ship is at sea

Web Services: Okeanos Explorer Bathymetric Grids
Leverage: NOAA’s GeoPlatform
Leverage: NOAA’s GeoPlatform
The Future of Data:
Direct-to-Cloud-to-Users with Telepresence X.0 and Data X.0
Future: Telepresence X.0 and Data X.0

Video, Communications, and Data

BROADBANDS SATELLITE

STREAMING SERVICES
live/real time video, data, communications, two-way

Science Services
Video

Science Party
Public

Commercial Network Provider

CLOUD SERVICES
All processing, QA/QC, product generation in cloud

Data, Daily Products, & Data Visualizations

Research
Deep Archive (Long Term Preservation)

Any Deep Ocean Exploration Platform
Science Benefits: Telepresence X.0 and Data X.0

Any Deep Ocean Exploration Platform

Video, Communications, and Data

BROADBAND SATELLITE

STREAMING SERVICES
live/real time video, data, communications, two-way

Science Services

Science Party

Video

Commercial Network Provider

CLOUD SERVICES
All processing, QA/QC, product generation in cloud

Data, Daily Products, & Data Visualizations

Research

Public

Deep Archive (Long Term Preservation)
Interactive Science Tools from Web Services

Combining output from multiple web services to:
• drive operations;
• verify data quality; and,
• develop intelligent data in near-real time
Science Benefits: the Virtual Whiteboard

Envisioning the **Virtual Whiteboard** for collaborative scientific analysis

- Support annotations and future data accessibility and usability
- Contributes to scientific understanding
Operational Guidance: 3D Chemistry Model
Integrated GIS Analysis Tools
Data Management Evolution

EXPLORE:
2002: Archive
Define and document best data management practices

EXPERIENCE:
2010: Access
Best practices implemented and refined aboard NOAA Ship Okeanos Explorer

UNDERSTAND:
2019: Reuse
Ocean Exploration data readily accessible and usable

EMPOWER USERS:
2024: Service Economy
Telepresence X.0 and Data X.0 will empower users to build their own experience