National Environmental Satellite, Data, and Information Service

US NODC Archival Management Practices and the OAIS Reference Model

Donald W. Collins

National Oceanographic Data Center

SSMC3 4TH Floor, 1315 East West Hwy, Silver Spring MD 20910

Donald.Collins@noaa.gov
www.nodc.noaa.gov/Archive/Search

NASA/IEEE MSST 2004
12th NASA Goddard/21st IEEE Conference on
Mass Storage Systems & Technologies
The Inn and Conference Center
University of Maryland University College
Adelphi MD USA
April 13-16, 2004

NODC AMS and the OAIS RM

- **What is the Open Archival Information System Reference Model?**
- **What is the NODC Archive Management System?**
- **How are they related?**
- **What is the benefit to the oceanographic community?**

Open Archival Information System (OAIS) Reference Model

- **Authored initially by the Consultative Committee for Space Data Systems**
- **ISO 14721:2002**
- **Common language to describe digital archives components, functions, and processes**
- **Main components of an OAIS**
  - Participants
  - Information packages
  - Process definitions
  - Function definitions

OAIS RM Participants

- **Producers - collects or creates original data and documentation**
- **Consumers/Designated Community - data users**
- **OAIS/Archive - system to manage data for the Long Term**
- **Management - manages and sets policy for the OAIS/Archive**

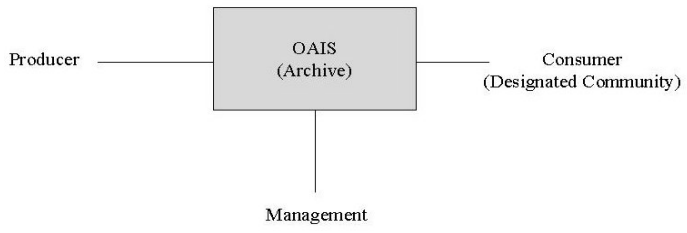


Figure 1. OAIS environment model. (After CCSDS, 2002, p. 2-2)

OAIS RM Components

- **Data and Representation Information**
- **Information Object**
  - Submission Information Package (SIP)
  - Archival IP (AIP)
  - Dissemination IP (DIP)
- **An IP contains**
  - Content Information
  - Packaging Information
  - Description Information
  - Preservation Description Information

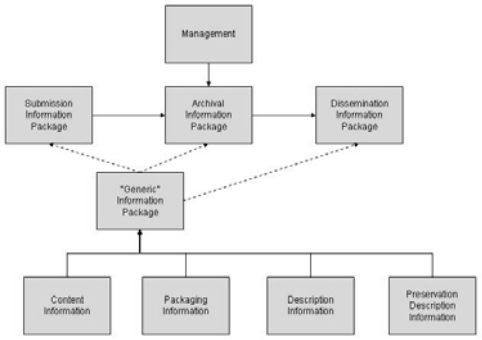
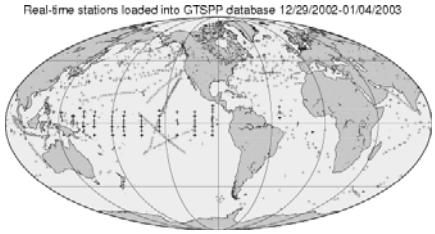


Figure 2. Schematic diagram of information Package objects and the high-level relationships between SIP, AIP, DIP, and Management (after *Sawyer*, 2002).

OAIS RM Processes and Functions

- **Data Submission processes**
  - Data Submission session(s)
  - Data Submission Agreement - terms of data transfer and responsibilities of Producer and OAIS
- **Ingest processes**
  - Procedures for bringing data into the OAIS
  - May require modification of data format(s)
- **Archival processes**
  - Media Migration
  - AIP Management
- **Data Dissemination processes**
  - Package Description
  - Access Aids, Finding Aids, Ordering Aids
  - Data Dissemination Session(s)

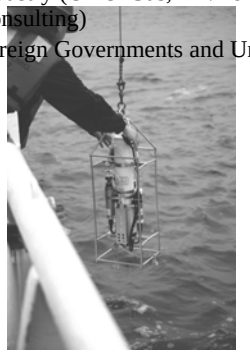
National Oceanographic Data Center



- **NODC: Largest collection of *in situ* oceanographic observations**
  - >20,419 AIPs of *in situ* data
    - Physical, biological, chemical and geological data
  - Satellite observations and analyses
  - Model output
- **Management - Director, Deputy Director, 5 Divisions**
  - Coastal Ocean Laboratory/Database Management Division
  - Information Systems Management Division
  - NOAA Central Library
  - Ocean Climate Laboratory
  - National Coastal Data Development Center

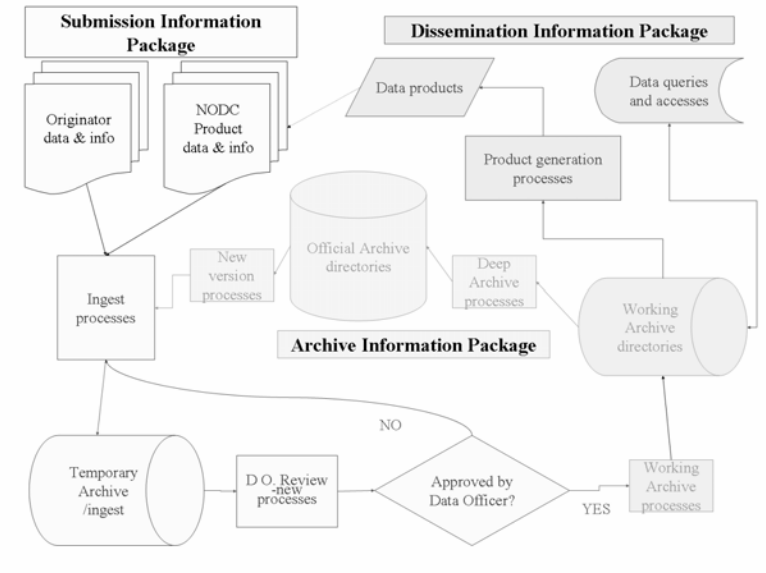
NODC Participants

- **Producers**
  - International and Intergovernmental Organizations
  - Universities and Private Research Institutions
  - Federal, State and Local Government Organizations
  - Industry (Oil & Gas, Environmental Consulting)
  - Foreign Governments and Universities



- **Designated Community / Consumers**
  - Federal, State and Local Government Organizations
  - Universities and Private Research Institutions
  - International and Intergovernmental Organizations
  - Industry (Oil & Gas, Environmental Consulting, Insurance)
  - K-12 Educators
  - General Public

AMS Data Flow and Processes

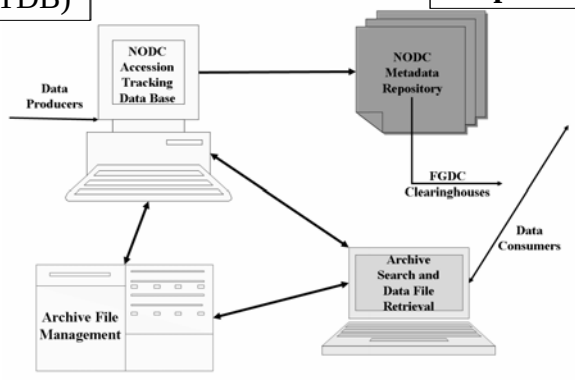


NASA/IEEE MSST 2004

NOCD Archive Management System Components

Accession Tracking Data Base (ATDB)

NODC Metadata Repository (NMR)



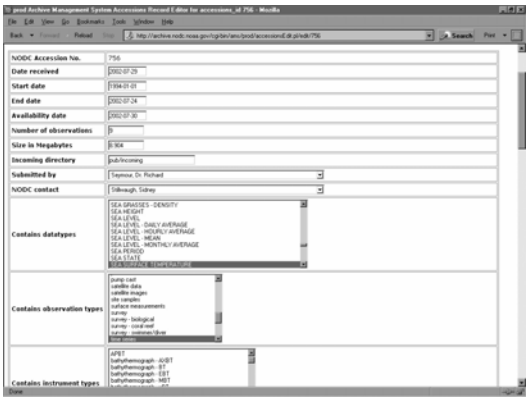
Archive File Management

Ocean Archive System (OAS)

NASA/IEEE MSST 2004

Accession Tracking Data Base

- **Generates unique Accession Number [1]**
- **Captures basic metadata and internal management information**
- **Exports FGDC XML metadata**
- **Uses controlled vocabularies**
- **Oversight/management of each AIP**



Archive File Management System

- **Generates uniform directory tree structure for each AIP**
- **Creates MD5 message digests for file validation [1]**
- **Performs virus checks**
- **Provides for automated backups**

#######/ - NODC Accession Number

#######.01-version.md5 - checksum digest for all files in this version

01-version/ - version directory (may be multiple versions)

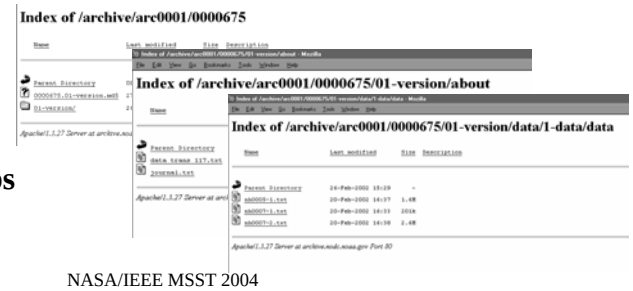
NODC-Readme.txt - file to explain this directory structure

about/ - directory containing NODC-created information about this accession

data/ - directory containing all translations of the data from the Producer

0-data/ - directory containing exact copy of Submission Information Package

1-data/ - directory containing translations of data in this SIP



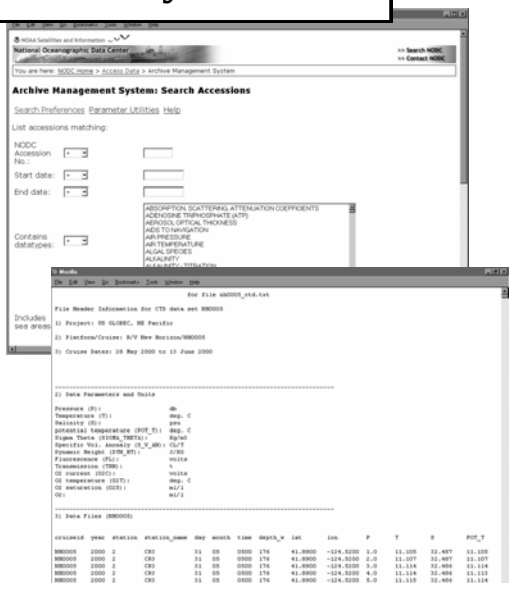
NODC Metadata Repository

- **Stores and manages FGDC CSDGM-compliant metadata for each AIP [1]**
- **Allows for import and export via XML of metadata**
- **Search and discovery via FGDC CSDGM metadata (not yet available)**

NASA/IEEE MSST 2004

Ocean Archive System

- **Consumer-driven search, discovery and retrieval via browser tools [1]**
- **Download DIP directly as a copy of the AIP**
- **Provides message digests to ensure download validity**
- **Data can be discovered, but not directly obtained, while in ingest process**



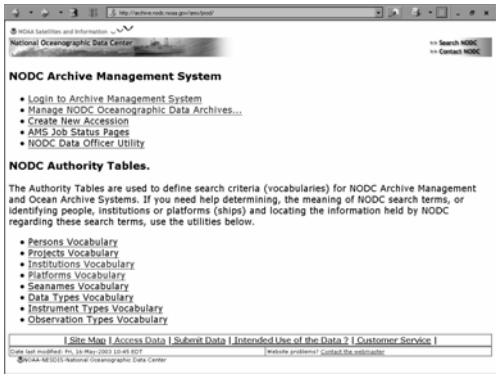
NASA/IEEE MSST 2004

What's Missing?

- **Data Submission Agreement**
  - Provides details of data to be submitted and the responsibilities of the NODC and the Producer regarding the SIP
  - Development of a standardized Data Submission Agreement is underway
- **Preservation Description Information**
  - Goes beyond metadata useful for discovery or re-use
  - AMS provides improved Content Information management
  - Difficult to obtain Detailed Preservation Information for historic data holdings

AMS Benefits the Ocean Community

- **Long Term data management, archival maintenance and access at no direct cost to Producers**
- **Fulfills contractual obligations for federally funded research and NOAA Administrative Orders [1]**
- **Low cost access to global data from a reliable source**
- **Improved data management capabilities within NODC**



NOAA National Data Centers are utilizing the OAIS RM in planning for the future

Acknowledgements

Additional Resources:

[1] Collins, Rutz, Ogata, Mitchell, Shirley, and Thailambal, 2004, Introducing the US NODC Archive Management System. Poster presented at the AGU Ocean Sciences 2004 Conference, Portland OR, January 2004.

Animated GIF image of GTSPP stations courtesy of Steven B. Rutz.

Thank you for support and assistance with this paper:

NODC Management: Kurt Schnebele, Parmesh Dwivedi, Janice Beattie, Wayne Wilmot, and Lee Dantzler

Colleagues: Steve Rutz, Mai Edwards, and Ted Habermann

University of Maryland College of Information Studies: Lauren Brown and the students in the "Seminar in Archives, Records, and Information Management" Spring 2003

Donna Collins and the boys