



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

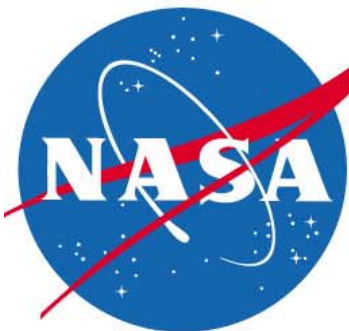
NASA/IEEE MSST 2004

**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

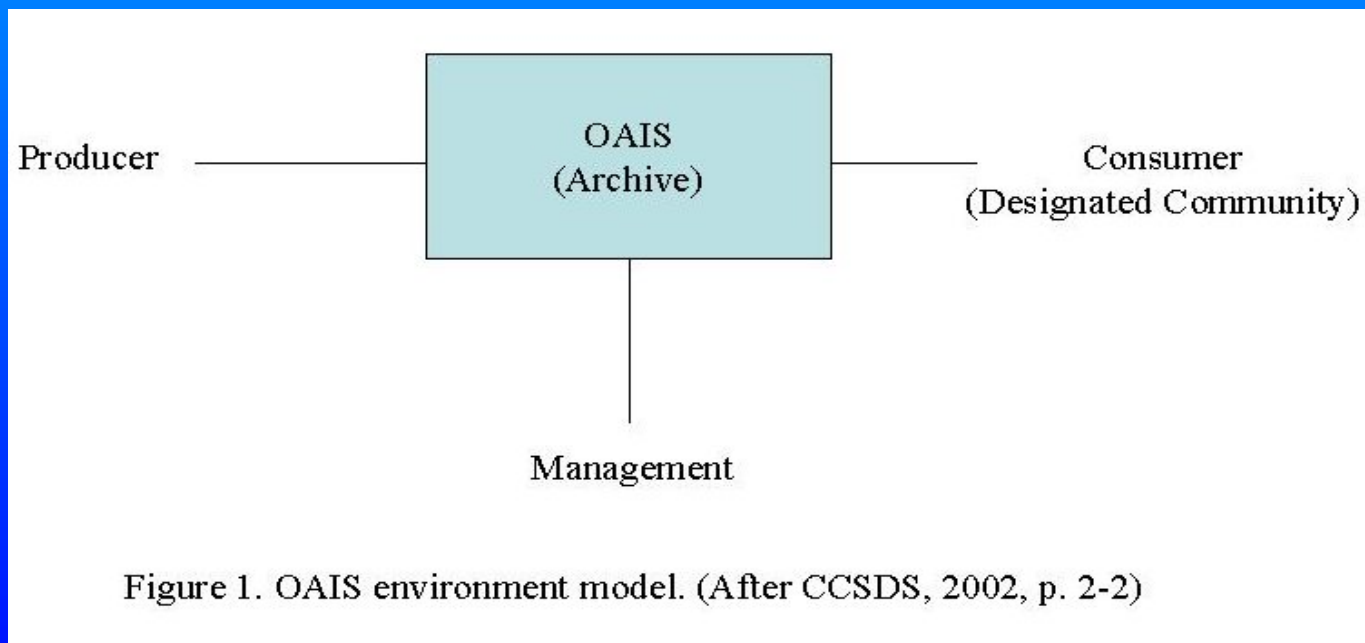


Figure 1. OAIS environment model. (After CCSDS, 2002, p. 2-2)



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

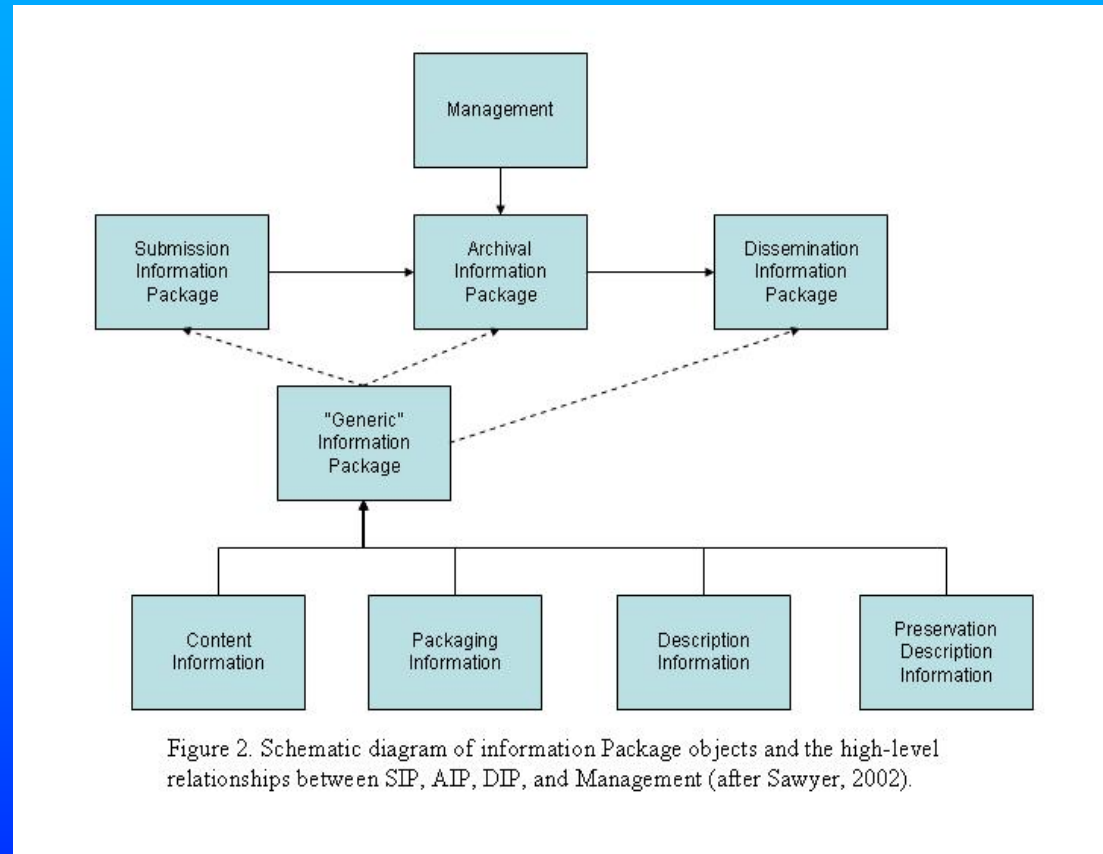


Figure 2. Schematic diagram of information Package objects and the high-level relationships between SIP, AIP, DIP, and Management (after Sawyer, 2002).

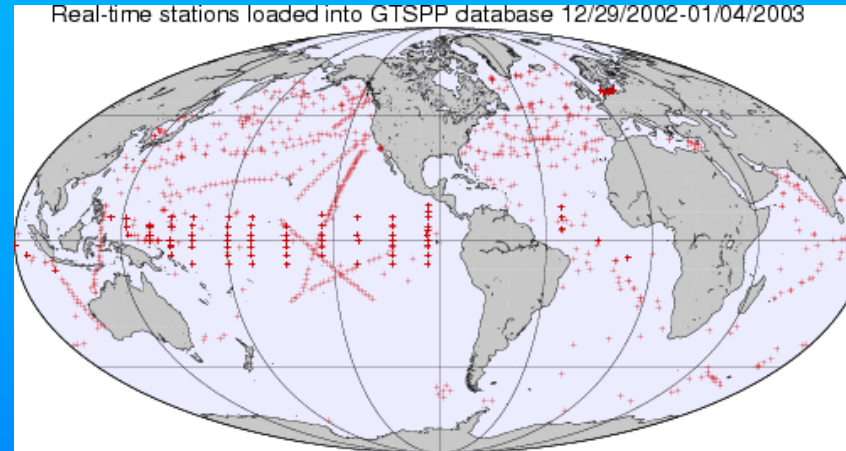


OAIS RM Processes and Functions

- **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



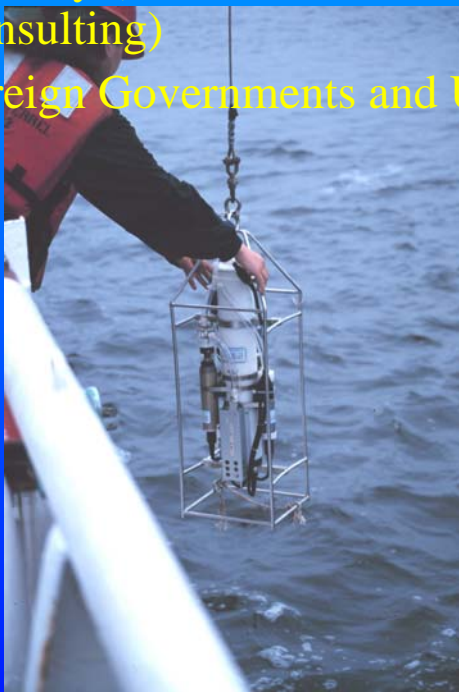
- **NODC: Largest collection of *in situ* oceanographic observations**
 - >20,419 AIPs of *in situ* data
 - Physical, biological, chemical and geological data
 - Satellite observations and analyses
 - Model output
- **Management - Director, Deputy Director, 5 Divisions**
 - Coastal Ocean Laboratory/Database 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

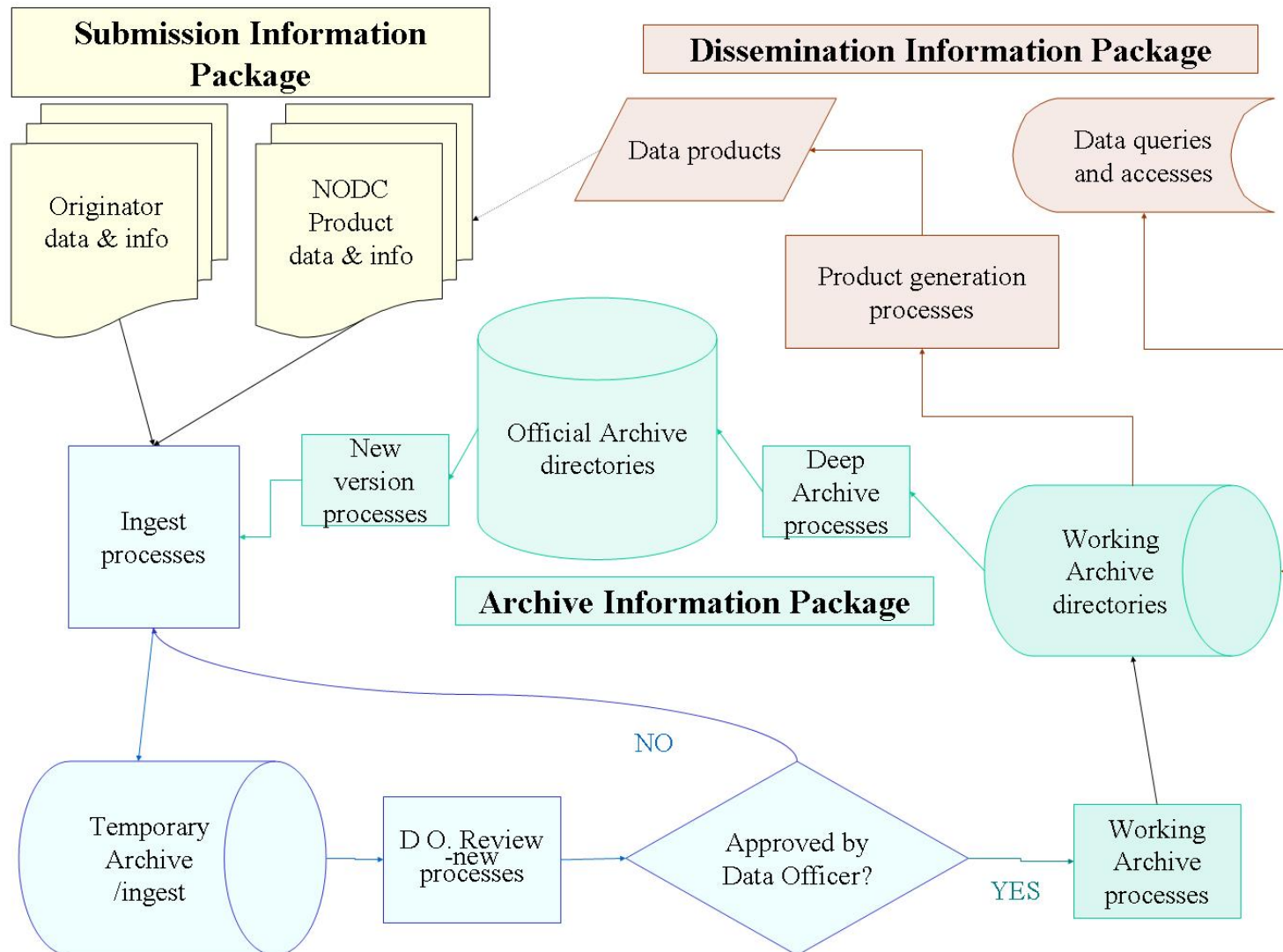


- **Designated Community / Consumers**

- 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

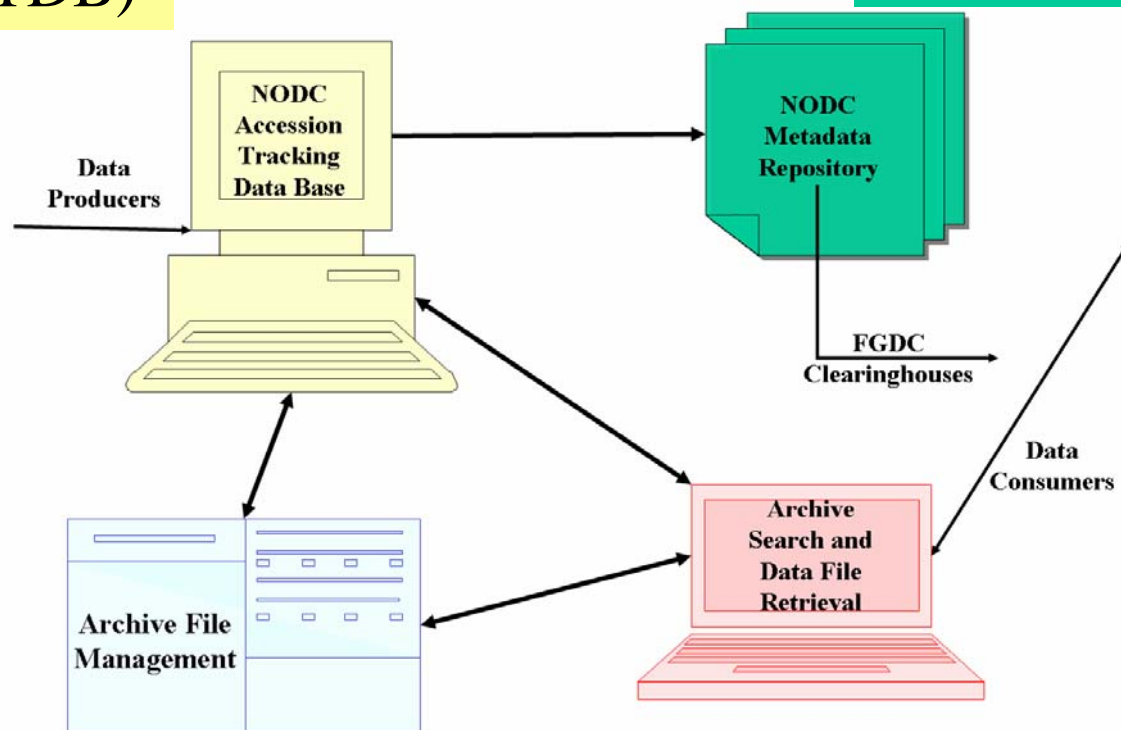




NODC Archive Management System Components

Accession Tracking Data Base (ATDB)

NODC Metadata Repository (NMR)



Archive File Management

Ocean Archive System (OAS)



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

The screenshot shows a web browser window titled "prod Archive Management System Accessions Record Editor for accessions_id 756 - Mozilla". The address bar shows the URL "http://archive.nodc.noaa.gov/cgi-bin/ams/prod/accessionsEdit.pl/edit/756". The main content area is a form with the following fields:

NODC Accession No.	756
Date received	2002-07-29
Start date	1994-01-01
End date	2002-07-24
Availability date	2002-07-30
Number of observations	9
Size in Megabytes	8.904
Incoming directory	pub/incoming
Submitted by	Seymour, Dr. Richard
NODC contact	Stillwaugh, Sidney
Contains datatypes	SEA GRASSES - DENSITY SEA HEIGHT SEA LEVEL SEA LEVEL - DAILY AVERAGE SEA LEVEL - HOURLY AVERAGE SEA LEVEL - MEAN SEA LEVEL - MONTHLY AVERAGE SEA PERIOD SEA STATE SEA SURFACE TEMPERATURE
Contains observation types	pump cast satellite data satellite images site samples surface measurements survey survey - biological survey - coral reef survey - swimmer/diver time series
Contains instrument types	AFBT bathythermograph - AXBT bathythermograph - BT bathythermograph - EBT bathythermograph - MBT



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

Index of /archive/arc0001/0000675

Name	Last modified	Size	Description
------	---------------	------	-------------

Parent Directory 05

0000675.01-version.md5 27

01-version/ 26

Apache/1.3.27 Server at archive.noaa.gov

Index of /archive/arc0001/0000675/01-version/about - Mozilla

File Edit View Go Bookmarks Tools Window Help

Index of /archive/arc0001/0000675/01-version/about

Name	Last modified	Size	Description
------	---------------	------	-------------

Parent Directory

data trans 117.txt

journal.txt

Apache/1.3.27 Server at archive.noaa.gov

Index of /archive/arc0001/0000675/01-version/data/1-data/data - Mozilla

File Edit View Go Bookmarks Tools Window Help

Index of /archive/arc0001/0000675/01-version/data/1-data/data

Name	Last modified	Size	Description
------	---------------	------	-------------

Parent Directory

nh0005-1.txt

nh0007-1.txt

nh0007-2.txt

Apache/1.3.27 Server at archive.nodc.noaa.gov Port 80



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)

Identification_Information:
Citation:
Originator: OREGON STATE UNIVERSITY
Publication_Date: 20020228
Title: DEPTH - OBSERVATION and Other Data from NEW HORIZON and Other Platforms from 20000531 to 20000812 (Accession No. - 0000675)
Geospatial_Data_Presentation_Form: TABULAR DATA
Description:
Abstract: NOT AVAILABLE
Purpose: BASIC RESEARCH
Supplemental_Information: Note: Metadata for this accession were extracted from a legacy database maintained by the U.S. National Oceanographic Data Center (NODC). The design of the database did not exactly reflect the FGDC Content Standard for Digital Geospatial Metadata (CSDGM). Principal Investigator (PI) and organization contact information accurately represents all available information from the legacy database at the time that this description was created. However, proper attribution of a PI to a specific institution or the role (submitting or collecting) taken by an institution may not be correct due to inexact mapping between fields in the legacy database and the CSDGM. Due to this uncertainty, the contact information was initially recorded in the Supplemental Information element of the CSDGM description. To develop more accurate metadata, the NODC reviews metadata for all accessions on an ongoing basis. Other related points of contact for this data set include: Contact info: Agency: OREGON STATE UNIVERSITY PI: (1541) Hoyer, Dr. Adrianna PI: 1541 Address: address: OCEAN ADMINISTRATION BUILDING 104 city: CORVALLIS state: OR postal: 97331-5503 country: UNITED STATES
NOAA Supplemental:
Entry_ID: TIME SERIES CTD from FIXED PLATFORM : 19930924 to 19960903 (NODC Accession No: 0000001) - Mozilla
Sensor_Name:
Sensor_ID:
Source_Name:
Project:
Originator:
Storage:
Time_Period:
Time_Period:
Range_of:
Begin:
Ending:
Currentness:
Status:
Progress:
Maintenance:

TIME SERIES CTD from FIXED PLATFORM : 19930924 to 19960903 (NODC Accession No: 0000001)

Metadata also available as

Metadata:

- [Identification Information](#)
- [Entry and Attribute Information](#)
- [Distribution Information](#)
- [Metadata Reference Information](#)

Identification_Information:
Citation:
Originator: UNIVERSITY OF ALASKA, IMS, FAIRBANKS
Publication_Date: 20000128
Title:
TIME SERIES CTD from FIXED PLATFORM : 19930924 to 19960903 (NODC Accession No: 0000001)
Edition: One
Online_Linkage: <<http://www.nodc.noaa.gov/col/projects/access/nodc.dir.html>>
Description:
Abstract: None provided
Purpose: Basic Research
Supplemental_Information: NONE PROVIDED
Supplemental_Information: Other Contacts include UNIVERSITY OF ALASKA, IMS, FAIRBANKS
NOAA Supplemental:
Entry_ID: NONE PROVIDED
Sensor_Name: TIME SERIES CTD
Source_Name: BUOY FIXED PLATFORM
Project/Campaign: UNKNOWN



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

NOAA Satellites and Information
National Oceanographic Data Center

You are here: [NODC Home](#) > [Access Data](#) > Archive Management System

Archive Management System: Search Accessions

[Search Preferences](#) [Parameter Utilities](#) [Help](#)

List accessions matching:

NODC Accession No.:

Start date:

End date:

Contains datatypes:

Includes sea areas

ABSORPTION, SCATTERING, ATTENUATION COEFFICIENTS
ADENOSINE TRIPHOSPHATE (ATP)
AEROSOL OPTICAL THICKNESS
AIDS TO NAVIGATION
AIR PRESSURE
AIR TEMPERATURE
ALGAL SPECIES
ALKALINITY
ALKALINITY-TITRATION

for file nh0005_ctd.txt

File Header Information for CTD data set NH0005

- 1) Project: US GLOBEC, NE Pacific
- 2) Platform/Cruise: R/V New Horizon/NH0005
- 3) Cruise Dates: 28 May 2000 to 13 June 2000

2) Data Parameters and Units

Pressure (P): db
Temperature (T): deg. C
Salinity (S): psu
potential temperature (POT_T): deg. C
Sigma Theta (SIGMA_THETA): Kg/m3
Specific Vol. Anomaly (S_V_AN): CL/T
Dynamic Height (DYN_HT): J/KG
Fluorescence (FL): volts
Transmission (TRM): %
O2 current (O2C): volts
O2 temperature (O2T): deg. C
O2 saturation (O2S): ml/l
O2: ml/l

3) Data Files (NH0005)

cruiseid	year	station	station_name	day	month	time	depth_w	lat	lon	P	T	S	POT_T
NH0005	2000	2	CR3	31	05	0500	176	41.8900	-124.5200	1.0	11.105	32.487	11.105
NH0005	2000	2	CR3	31	05	0500	176	41.8900	-124.5200	2.0	11.107	32.487	11.107
NH0005	2000	2	CR3	31	05	0500	176	41.8900	-124.5200	3.0	11.114	32.486	11.114
NH0005	2000	2	CR3	31	05	0500	176	41.8900	-124.5200	4.0	11.114	32.486	11.113
NH0005	2000	2	CR3	31	05	0500	176	41.8900	-124.5200	5.0	11.115	32.486	11.114



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

The screenshot shows a web browser window with the URL <http://archive.nodc.noaa.gov/ams/prod/>. The page header includes 'NOAA Satellites and Information' and 'National Oceanographic Data Center'. There are links for 'Search NODC' and 'Contact NODC'. The main content area is titled 'NODC Archive Management System' and contains a list of links: 'Login to Archive Management System', 'Manage NODC Oceanographic Data Archives...', 'Create New Accession', 'AMS Job Status Pages', and 'NODC Data Officer Utility'. Below this is a section for 'NODC Authority Tables' with a paragraph explaining their use and a list of vocabulary links: 'Persons Vocabulary', 'Projects Vocabulary', 'Institutions Vocabulary', 'Platforms Vocabulary', 'Seanames Vocabulary', 'Data Types Vocabulary', 'Instrument Types Vocabulary', and 'Observation Types Vocabulary'. At the bottom, there is a navigation bar with links for 'Site Map', 'Access Data', 'Submit Data', 'Intended Use of the Data?', and 'Customer Service'. The footer includes the date 'Date last modified: Fri, 16-May-2003 10:45 EDT' and the text 'NOAA-NESDIS-National Oceanographic Data Center'.

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



Additional Resources and Acknowledgements

[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.

Consultative Committee on Space Data Systems, 2002, Reference Model for an Open Archival Information System. Last accessed online on 4/12/04 at <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>

Animated GIF image of GTSP 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, Anna Fiolek, Bob Gelfeld, 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 and the boys