



NODC Implementation of the OAIS Reference Model Functional Entities

The Lingua Franca of Digital Archives

Kenneth S. Casey
Technical Director
NOAA National Oceanographic Data Center



What is a Reference Model?

- A framework for understanding significant relationships between the entities of a system and to the external environment
- A *lingua franca* for discussion, development, implementation, management, and use of that system

A reference model is NOT an implementation!!



Why the OAIS Matters to NOAA

- Provides a framework for consideration of roles and responsibilities of NNCDs and CLASS to ensure fundamental archive responsibilities are accomplished
- Facilitates strategic planning of archive evolution
- Assists with the education of our customers and stakeholders on essential information preservation activities



Why the OAIS Matters to NODC

- It gives a common language for us to use as we discuss both our internal archive process and our cross-NOAA archive activities
- It serves as a useful tool for “gap analysis”, by providing a comprehensive suite of functions we should perform and responsibilities we must meet to be called a “capital A” Archive



The OAIS Reference Model

- The **Open Archival Information System Reference Model** (OAIS RM) is the CCSDS and ISO Standard (14721) for Digital Archives
- It applies to all organizations that need to preserve digital information for the long-term
- It does NOT specify any particular implementation
- An organization *conforms* to the OAIS RM by discharging a minimal set of responsibilities and supporting basic information concepts



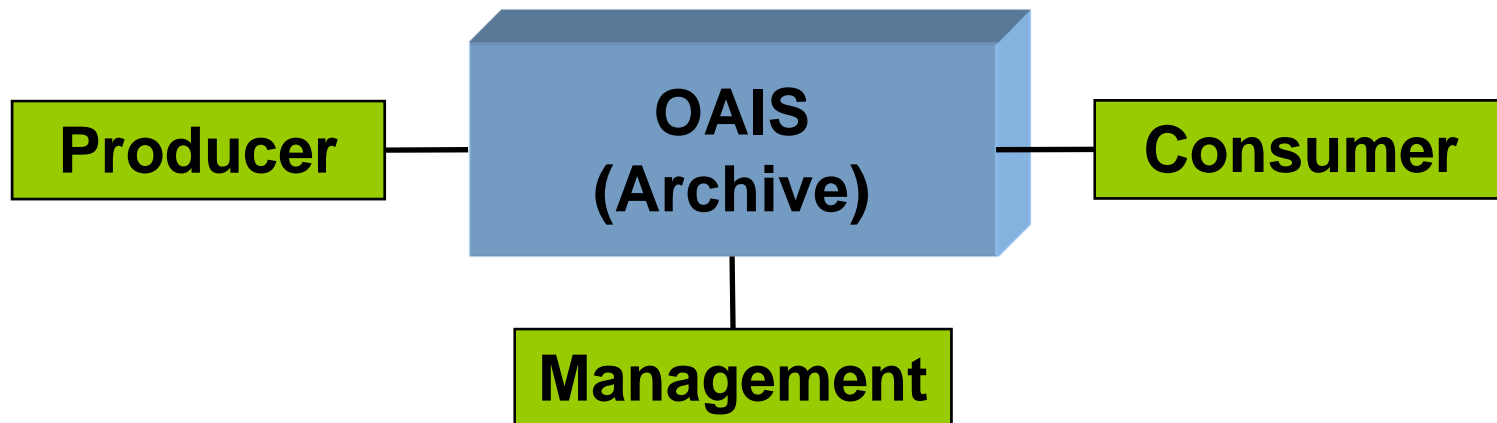
OAIS Responsibilities

- Negotiates and accepts information from **Producers**
- **Obtains sufficient control** to ensure long-term preservation
- Ensures the information to be preserved is **independently understandable** to identified **Designated Communities**
- **Follows** documented **policies and procedures** to insure information is preserved
- **Provides information** to the Designated Communities in understandable forms



The OAIS Environment

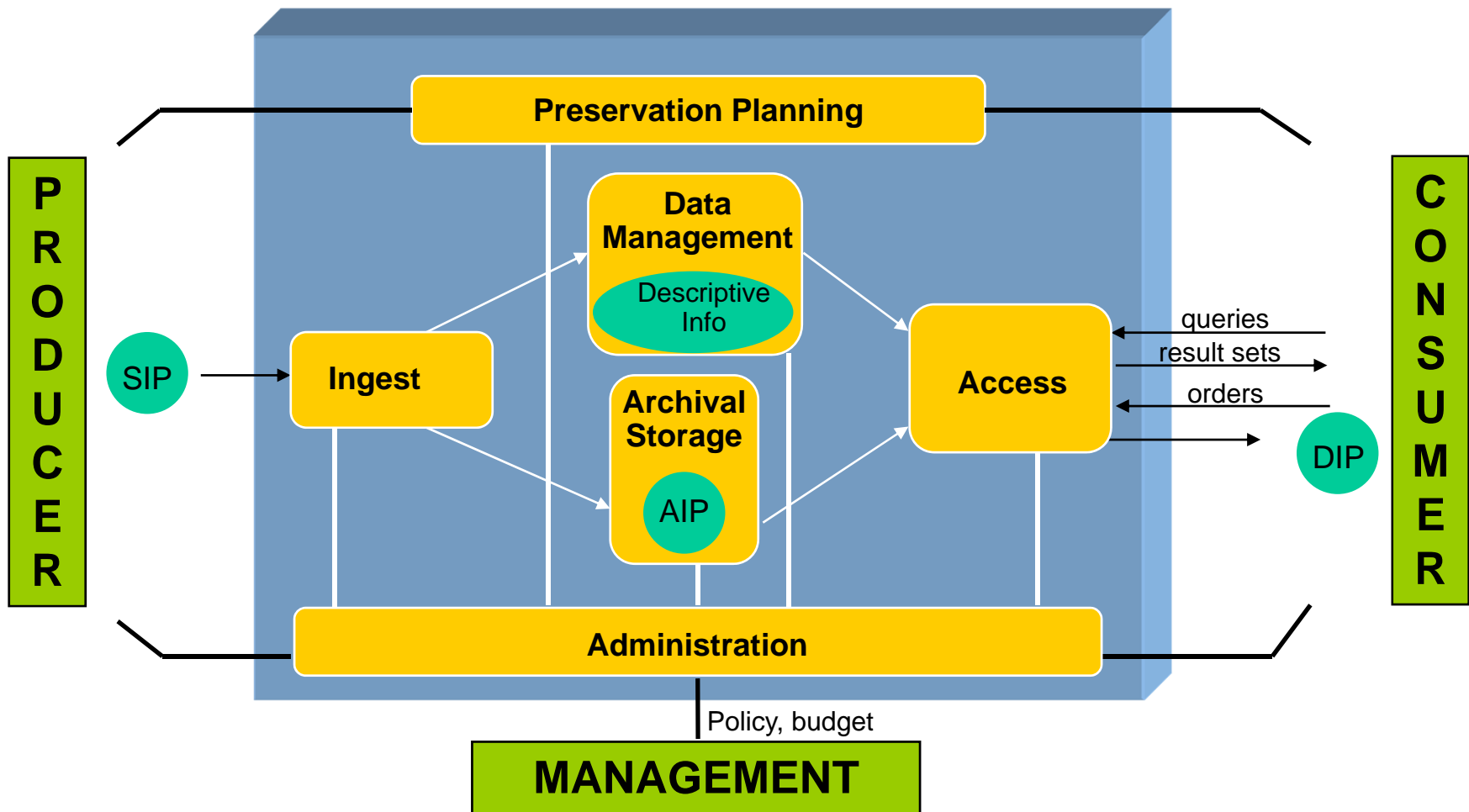
- ***Producer*** provides information to be preserved
- ***Management*** sets overall policy
- ***Consumer*** seeks and acquires preserved information



The OAIS Environment from 30,000 ft



OAIS Functional Entities



SIP = Submission Information Package

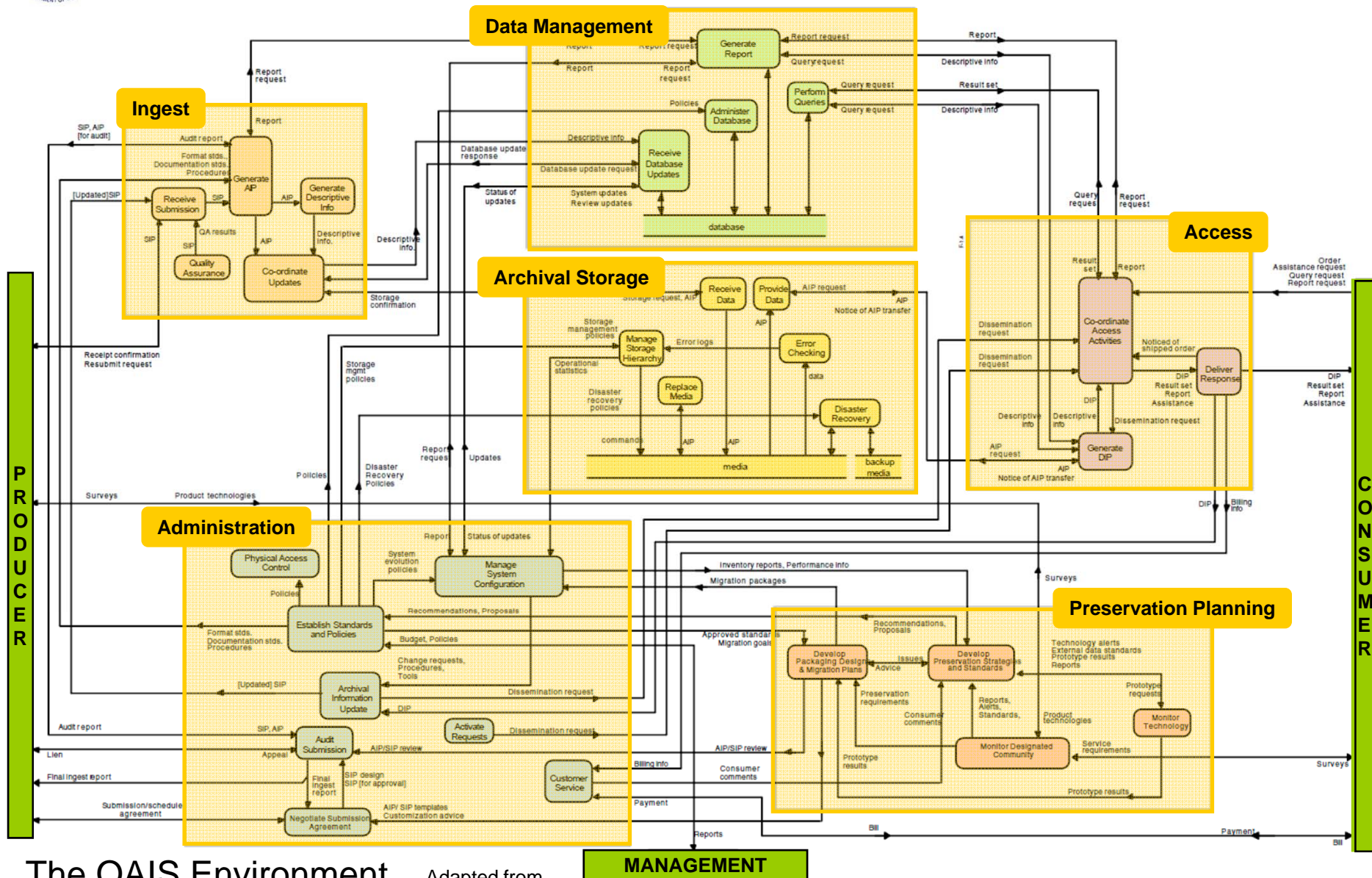
AIP = Archival Information Package

DIP = Dissemination Information Package

The OAIS Environment
from 10,000 ft



OAIS Functional Entities



The OAIS Environment
from Sea Level

Adapted from

Figure F-1: Composite of Functional Entities



So... How has NODC evolved to meet the tenets of the OAIS RM?

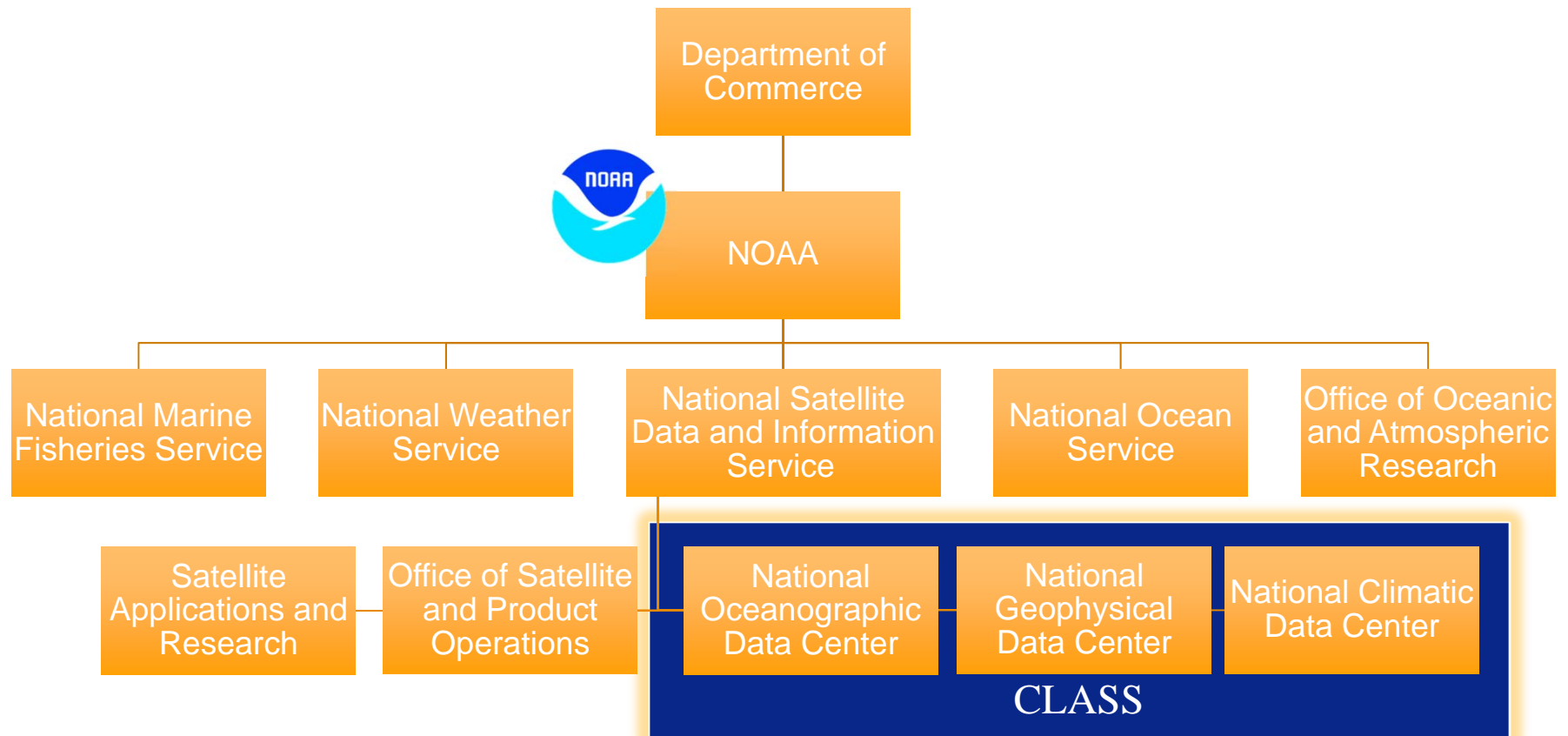
A little context to set the stage...



NAA is a Big Data Agency



NOAA and its National Data Centers

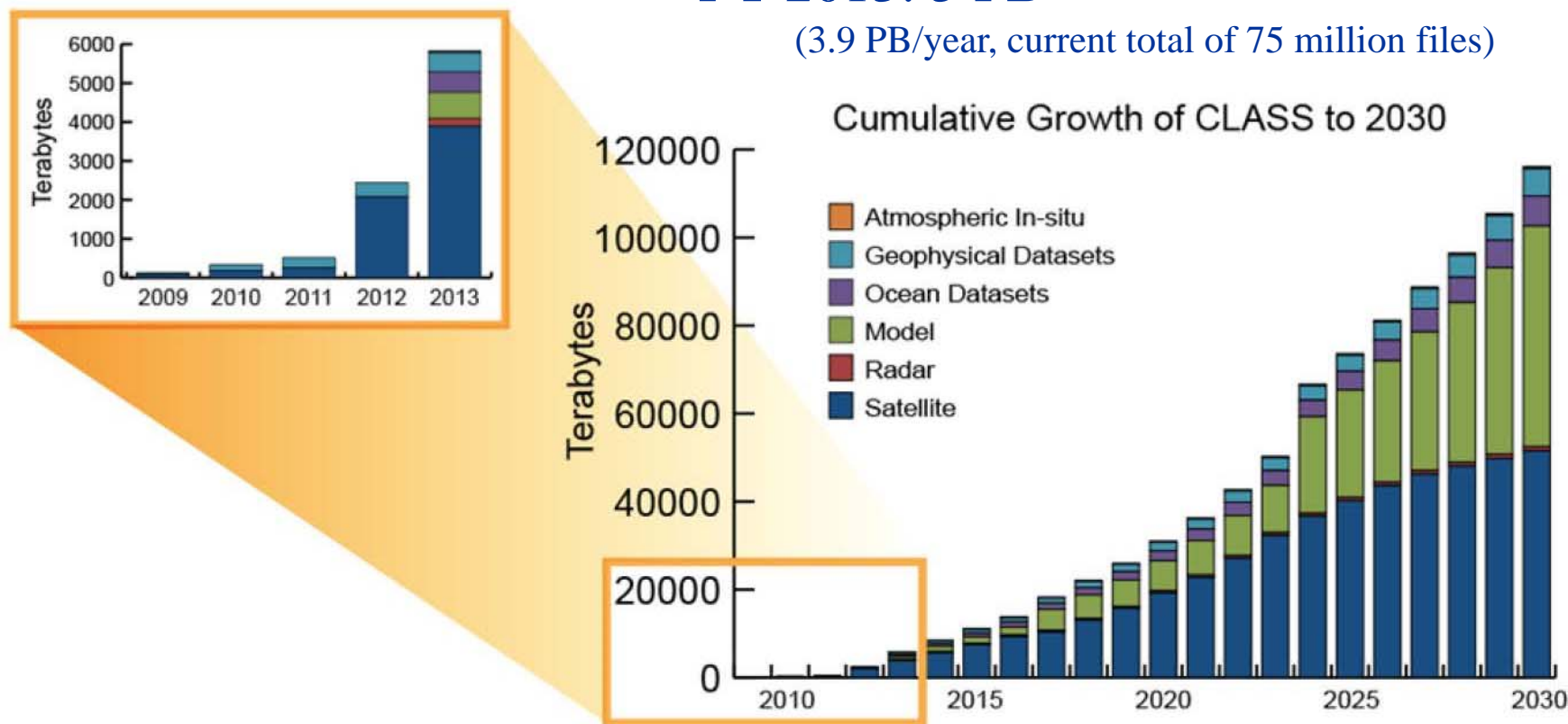


NOAA has three National Data Centers: NODC, NGDC, and NCDC
... but many parts of NOAA serve data!



NOAA Archive Data Volumes

FY 2013: 5 PB  FY 2030: 120 PB
(3.9 PB/year, current total of 75 million files)



Current and Projected Volumes in NOAA Archives

[Source: NOAA CLASS Program Office]



NOAA Archive User Accesses

- 🌐 **Using our shared storage and access system known as CLASS, the NOAA National Data Centers serve, each month...**
 - Over 7.7 million files
 - Over 550 Terabytes of data
- 🌐 **But each maintains its own local systems as well...**
 - Now look more closely at NODC's Big Data numbers...

Averages based on Apr-Jun 2013



NODC Archive Variety



Underway



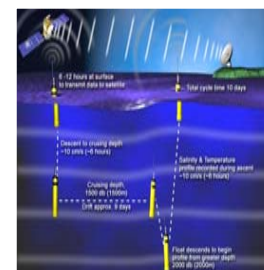
CTD/Niskin



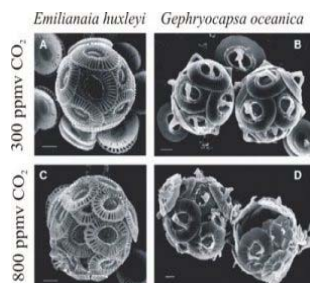
Buoys



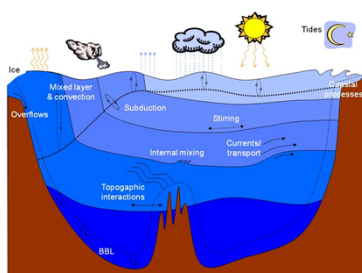
Net Tows



Argo Profilers



Experimental



Models



XBT



Satellite



AUVs



Aerial Surveys



Instrumented Animals



Wave Gliders

~100 Terabytes, in 27,000+ collections from 1500+ institutions, in hundreds of formats



NODC Archive Variety and Velocity

- 🌐 **Each month, NODC alone archives more than...**
 - 106 different data types, 300,000 files, in over 50 different data formats
- 🌐 **from...**
 - 22 distinct institutions on average, 37 automated data streams, dozens of individual submissions
- 🌐 **And provides more than 16 million files and 21 Terabytes of data to over 261,000 unique hosts**

Averages based on Apr-Jun 2013



S... We have the “classic” Big Data challenges of Volume, Velocity, and Variety...

What have we done to manage it and meet the tenets of the OAIS-RM?



NODC's *Big Data* Archive Paradigm

Human and machine
interfaces to

***Discover, Access, Use, and
Understand***

NODC's Archive Holdings

The Open Archival Information System (OAIS) Reference Model, the ISO standard for digital archives, specifies key functional entities of a “Capital A” Archive

AIP

AIP

AIP

AIP

AIP

The NODC Ocean Archive

In OAIS terminology, an archive preserves **Archival Information Packages (AIP)**. NODC refers to these AIPs as “Accessions”.



Understandability

Understanding

FGDC

ISO

Other standards

Ad hoc documentation

Metadata Views

Provenance

Checksums

Manifests

Value-Added Products

Understanding and *Preservation* are enabled through a focus on standardized descriptions of “contents”, their “containers”, and where they have been.

AIP

AIP

AIP

AIP

AIP

The NODC Ocean Archive

Better documentation enables better understanding and long term preservation.



Discoverability

Discovery

Human to Machine Interfaces

Google

Data.gov

Geo.Data.gov

CWIC

Geoportal Server Web App

Machine to Machine Interfaces

CSW

Geoportal Server REST API

OpenSearch

SRU/ISO23950

WAF

AIP

AIP

AIP

AIP

AIP

The NODC Ocean Archive

Discovery is enabled through numerous interfaces designed for both humans and their machine clients.

Human-to-machine interfaces include government-mandated generalized portals like Data.gov and international portals from CEOS, GEOSS, WMO, etc.

Discovery services are available for ALL of the NODC Archive Information Packages (AIP), but better metadata supports better discovery!



Accessibility – Pilot Projects

OPeNDAP Hyrax, Pathfinder... EC2

Hyrax, AVHRR Pathfinder, and other applications running in Amazon EC2 provide test environment for associating services and computational resources with the data in S3

<http://ocean-archive.data.nodc.noaa.gov/index.xml>

S3

Basic HTTP access provided by Amazon S3 provides test environment for cost analysis and technical connectivity

AIP

AIP

AIP

AIP

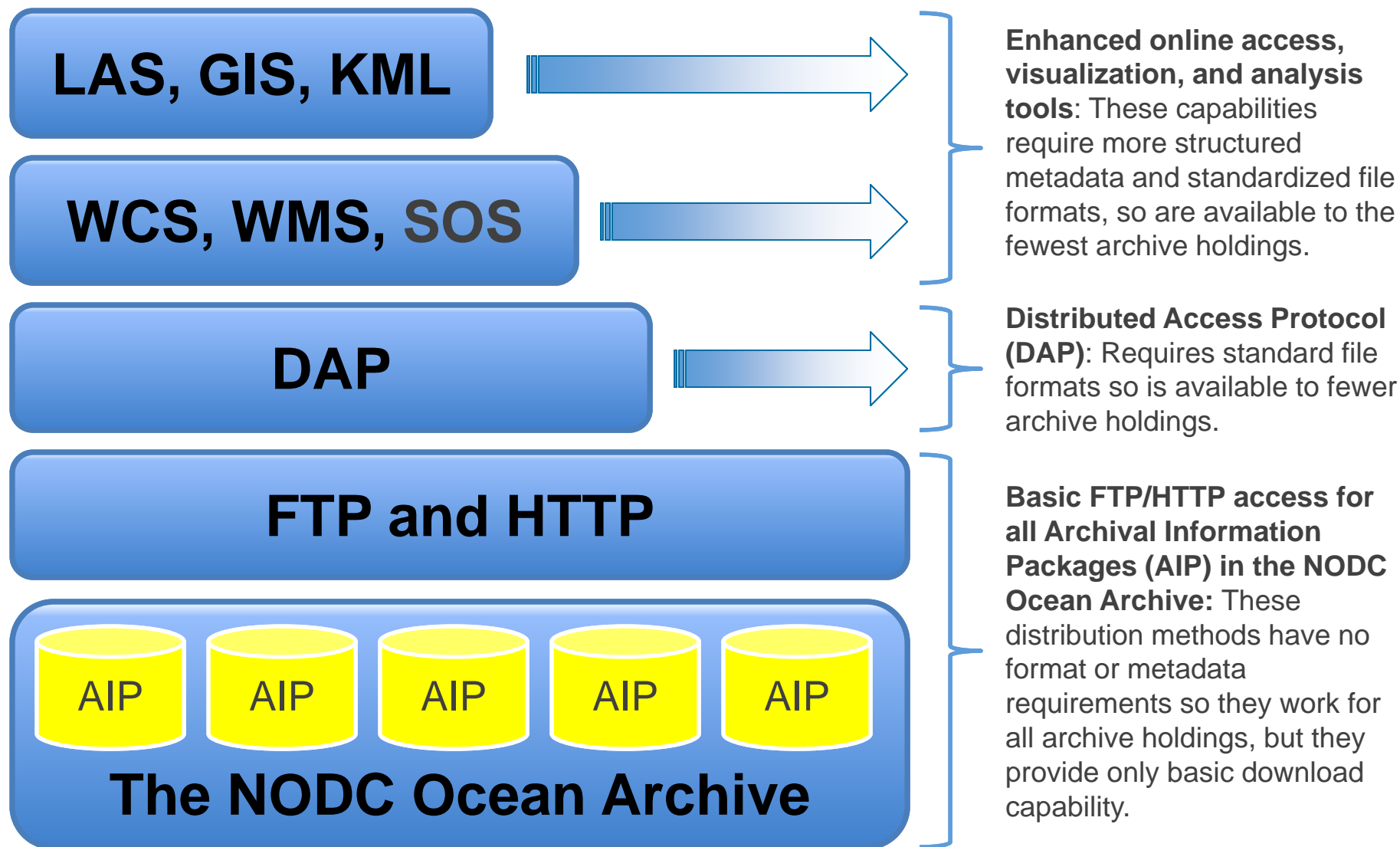
AIP

The NODC Ocean Archive

A subset of the NODC archive, our ocean datasets in CLASS, and data from NGDC and NCDC were uploaded to Amazon's S3 cloud.



Accessibility- Here and Now





Geoportal Server as Service Integrator

Linking collection level discovery to granule level discovery.

UAF HOME SEARCH BROWSE SEARCH TIPS

Search

Search metadata content, e.g. title:SST; use + to require keywords, e.g. +water +temperature;
use "" to search for an exact phrase, e.g. "water temperature"

fileIdentifier:REMSS-L2P_GRIDDED_25-WSAT

Additional Options

WHEN

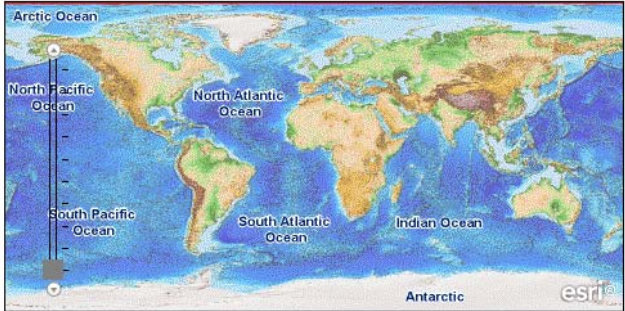
☒ Dates overlap range ☐ Dates within range

From: (yyyyymmdd)
To: (yyyyymmdd)

WHERE

Zoom the map to desired area and choose "intersecting" or "fully within"
You can zoom the map by shift-click-dragging a bounding box

☒ Anywhere ☐ Intersecting ☐ Fully within



Results 1-100 of 3080 record(s)


☐ Expand results [Zoom To Results](#) [Zoom To Searched Area](#)

- [20030626-WSAT-REMSS-L2P_GRIDDED_25-wsat_20030626v7-v01.nc.gz](#)
- [20030627-WSAT-REMSS-L2P_GRIDDED_25-wsat_20030627v7-v01.nc.gz](#)
- [20060217-WSAT-REMSS-L2P_GRIDDED_25-wsat_20060217v7-v01.nc.gz](#)
- [20060218-WSAT-REMSS-L2P_GRIDDED_25-wsat_20060218v7-v01.nc.gz](#)
- [20060219-WSAT-REMSS-L2P_GRIDDED_25-wsat_20060219v7-v01.nc.gz](#)
- [20060220-WSAT-REMSS-L2P_GRIDDED_25-wsat_20060220v7-v01.nc.gz](#)
- [20060221-WSAT-REMSS-L2P_GRIDDED_25-wsat_20060221v7-v01.nc.gz](#)
- [20060222-WSAT-REMSS-L2P_GRIDDED_25-wsat_20060222v7-v01.nc.gz](#)
- [20060223-WSAT-REMSS-L2P_GRIDDED_25-wsat_20060223v7-v01.nc.gz](#)
- [20060224-WSAT-REMSS-L2P_GRIDDED_25-wsat_20060224v7-v01.nc.gz](#)
- [20060225-WSAT-REMSS-L2P_GRIDDED_25-wsat_20060225v7-v01.nc.gz](#)
- [20060226-WSAT-REMSS-L2P_GRIDDED_25-wsat_20060226v7-v01.nc.gz](#)

calibrated passive microwave radiometer, similar to SSM/I, that contains lower frequency channels required for sea surface temperature (SST) retrievals. The TRMM is



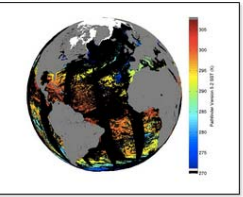
Sidebar: NOAA DOIs for Data

**NOAA** NATIONAL OCEANOGRAPHIC DATA CENTER (NODC)
UNITED STATES DEPARTMENT OF COMMERCE

NOAA Satellite and Information Service

AVHRR Pathfinder Version 5.2 Level 3 Collated (L3C) Global 4km Sea Surface Temperature for 1981-2011

AVHRR_Pathfinder-NODC-L3C-v5.2



The AVHRR Pathfinder Version 5.2 Sea Surface Temperature data set (PFV52) is a collection of global, twice-daily 4km sea surface temperature data produced in a partnership by the NOAA National Oceanographic Data Center and the University of Miami's Rosenstiel School of Marine and Atmospheric Science. PFV52 was computed from data from the AVHRR instruments on board NOAA's polar orbiting satellite series using an entirely modernized system based on SeaDAS. This system incorporates several key changes from Versions 5.0 and 5.1 of Pathfinder, including the use of an entirely new land mask, a modified grid, and the inclusion of sea ice, wind speed, and aerosol ancillary data to support the use of the SST data. Importantly, PFV52 data are provided in netCDF-4 (classic model, with internal compression and chunking) and are nearly 100% compliant with the GHRST Data Specification Version 2.0 for L3C products. These data deviate from that standard only in that sses_bias, sses_standard_deviation, and sst_dtime variables are empty. PFV52 data were collected through the operational periods of the NOAA-7 through NOAA-19 Polar Operational Environmental Satellites (POES), and are available for 1981 through 2011.

Get the Data

Access

download
[NODC HTTP server](#)
Navigate to URL: <http://data.nodc.noaa.gov/pathfinder/Version5.2/> and begin browsing through the file hierarchy. Clicking on any of the files will prompt you to download that file or will launch any application associated with netCDF files.

download
[NODC FTP server](#)
Navigate to URL: <ftp://ftp.nodc.noaa.gov/pub/data.nodc/pathfinder/Version5.2/> to download.

download
[NODC THREDDS server](#)
web services for data access, including OPeNDAP, HTTP, Web Coverage Service, Web Mapping Service, as well as NCML and ISO views to the data.

download
[NODC LAS server](#)
For data analysis and visualization

[Browse graphic](#)
Browse graphic for sample data view

Format(s)

netCDF
Format version: netCDF-4
Compression: Files are compressed internally using netCDF4.

Distributor(s) / Contact Info

[DOC/NOAA/NODC > National Oceanographic Data Center, NOAA, U.S. Department of Commerce](#)
Phone/E-mail/letter, NODC provides data free of charge.
(301) 713-3277
9:00 - 5:00 EST

Instructions / Constraints

Fees
For digital delivery, free for all users. For delivery on physical media, users are responsible for the cost of the physical media and production and delivery of the product.


Use Limitation
Cite as: Casey, Kenneth S., Evans, Robert H., Baringer, Warner, Kilpatrick, Katherine A., Podesta, Guillermo P., Walsh, Susan, Williams, Elizabeth, Brandon, Tees B., Byrne, Deirdre A., Foti, Gregg, Li, Yuanjie, Phillips, Sheri A., Zhang, Dexin, and Zhang, Yongsheng (2011): AVHRR Pathfinder Version 5.2 Level 3 Collated (L3C) Global 4km Sea Surface Temperature for 1981-2011. National Oceanographic Data Center, NOAA. Data set doi:10.7289/V5WD3XHB [access date]

Legal Constraints
Other

Access Data - Submit Data - Intended Use of the Data? - Online Store - Customer Service

Dept. of Commerce - NOAA - NESDIS - NODC
Offsite Link Notification

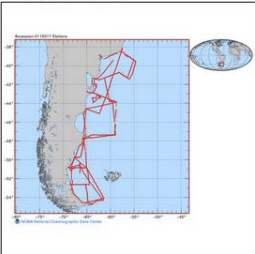
NODC.Webmaster@noaa.gov
Like us on Facebook | Follow us on Twitter
Privacy Policy - Disclaimer - Information Quality
Freedom of Information Act (FOIA)
USA.gov - The U.S. Government's Web Portal

**NOAA** NATIONAL OCEANOGRAPHIC DATA CENTER (NODC)
UNITED STATES DEPARTMENT OF COMMERCE

NOAA Satellite and Information Service

Surface and Oceanographic profile temperature, salinity and other measurements using CTD and thermosalinograph, taken from the PUERTO DESEADO in the Coastal South Atlantic, South Atlantic and other locations from 2005 to 2006 (NODC Accession 0110317)

NODC Accession 0110317



These data were collected as part of the scientific agenda of the United Nations Development Program (UNDP) and sponsored by the Secretary of Environment and Sustainable Development of Argentina (<http://www.ambiente.gov.ar>). The project Coastal Pollution Prevention and Marine Biodiversity Management (Project 28385-AR-UNDP-ARG/02/018) also referred to as GEF Patagonia was partially funded by the Global Environment Facility (GEF), with additional funding from the Inter-American Institute for Global Change Research (CRN061 and 2076). The project goal was to determine the seasonal variability of physical-chemical properties of highly productive regions of the southwest South Atlantic continental shelf.

The GEF Patagonia field experiments consisted of three oceanographic cruises carried out on board RV ARA PUERTO DESEADO, in October 2005 (PD GEFPAT_1), and March (PD GEFPAT_2) and September 2006 (PD GEFPAT_3). In each cruise, hydrographic stations were occupied along cross-shelf sections spanning the shelf and extending to the onshore edge of the western boundary currents between 38° and 55°S. This submission reports the CTD and thermosalinograph data collected during the three GEF Patagonia cruises.

Get the Data

Access

download
[NODC HTTP server](#)
Navigate directly to URL for data access and direct download.

download
[NODC FTP server](#)
Direct FTP access: Navigate to <ftp://ftp.nodc.noaa.gov/nodc/archive/larc0059/0110317/1.1/> using any FTP client to begin downloading data.

Format(s)

Seabird ascii format
Format version: 2006

Distributor(s) / Contact Info

[DOC/NOAA/NODC > National Oceanographic Data Center, NOAA, U.S. Department of Commerce](#)
Phone/E-mail/letter, NODC provides data free of charge.
(301) 713-3277
9:00 - 5:00 EST

Instructions / Constraints

Fees
For digital delivery, free for all users. For delivery on physical media, users are responsible for the cost of the physical media and production and delivery of the product.

Use Limitation
Cite as: Charo, Marcela and Piola, Alberto R. (2013): Surface and Oceanographic profile temperature, salinity and other measurements using CTD and thermosalinograph, taken from the PUERTO DESEADO in the Coastal S Atlantic, South Atlantic and other locations from 2005 to 2006 (NODC Accession 0110317). National Oceanographic Data Center, NOAA. doi:10.7289/V5RN35S0 [access date]

Legal Constraints
Other

Access Data - Submit Data - Intended Use of the Data? - Online Store - Customer Service

Dept. of Commerce - NOAA - NESDIS - NODC
Offsite Link Notification

NODC.Webmaster@noaa.gov
Like us on Facebook | Follow us on Twitter
Privacy Policy - Disclaimer - Information Quality
Freedom of Information Act (FOIA)
USA.gov - The U.S. Government's Web Portal



NODC Big Ocean Data



Data as Infrastructure Data as a Platform





Remaining slides for reference

🌐 Questions?



OAIS Common Services

- Operating System Services
- Network Services
- Security Services

These pervasive functions link together the other 6 OAIS Functional Entities...



OAIS Functional Entities

- **Ingest:** This entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers and prepare the contents for storage and management within the OAIS
 - Receive Submission
 - Quality Assurance
 - Generate AIP
 - Generate Descriptive Information
 - Coordinate Updates



OAIS Functional Entities

- **Archival Storage:** This entity provides the services and functions for the storage, maintenance and retrieval of Archival Information Packages (AIPs)
 - Receive Data
 - Manage Storage Hierarchy
 - Replace Media
 - Error Checking
 - Disaster recovery
 - Provide Data



OAIS Functional Entities

- **Data Management:** This entity provides the services and functions for populating, maintaining, and accessing both descriptive information that identifies and documents OAIS holdings and internal OAIS administrative data
 - Administer Database
 - Perform Queries
 - Generate Report
 - Receive Database Updates



OAIS Functional Entities

- **Access:** This entity supports Consumers in determining the existence, description, location and availability of information stored in the OAIS and allows Consumers to request and receive Dissemination Information Packages (DIPs)
 - Coordinate Access Activities
 - Generate DIP
 - Deliver Response



OAIS Functional Entities

- **Preservation Planning:** This entity monitors the environment of the OAIS and provides recommendations to ensure that the information stored in the OAIS remain accessible to the Designated Community over the long term
 - Monitor Designated Community
 - Monitor Technology
 - Develop Preservation Strategies and Standards
 - Develop Packaging Designs and Migration Plans



OAIS Functional Entities

- **Administration:** This entity manages the overall operation of the OAIS
 - Negotiate Submission Agreement
 - Manage System Configuration
 - Archival Information Update
 - Physical Access Control
 - Establish Standards and Policies
 - Audit Submission
 - Activate Requests
 - Customer Service