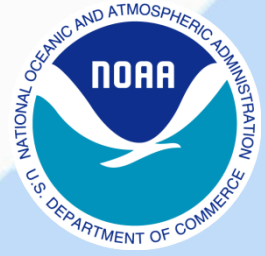


Levels of Archival Stewardship at the NOAA National Oceanographic Data Center: A Conceptual Model

Dr. Deirdre A. Byrne, lead, satellite oceanography team
with input from Mr. Donald Collins, lead archivist
2014-07-22



~~ARCHIVAL~~ SCIENTIFIC DATA STEWARDSHIP



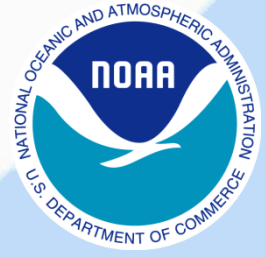
Q: What is scientific data stewardship?*

A: Application of an integrated suite of functions enabling the full scientific value of environmental data and information to be preserved and **exploited** over the long-term (decades).

*Much of this definition was adapted and condensed from the whitepaper outlining the goals and objectives of NOAA's Scientific Data Stewardship Program (attributed to Bates, Goldberg, *et al.*, 2005), and the corresponding review from the National Research Council (NAP, 2005).

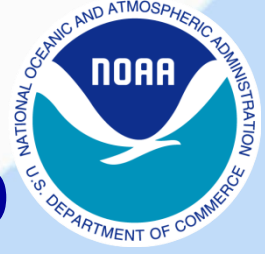


The five fundamental principles of Scientific Data Stewardship



1. Document the data and preserve it. Establish end-to-end accountability for its integrity. (Digital Preservation)
2. Provide IT support for the data to maintain its integrity and provide access that is flexible, adaptable and provides the capability perform computations (e.g., generate products or statistics). (Digital Preservation +)
3. Assess and ensure data quality. This mandates the detection of spurious observations and biases and requires subject matter expertise.
4. Engage with the scientific community to ensure the best understanding of the data is applied to any analyses or products. This also requires subject matter expertise.
5. Continuity: to be meaningful, stewardship must be provided over years and decades, spanning and adapting to scientific advances and changes in scientific understanding as well as those in information technology hardware and software.

Again adapted and condensed from the whitepaper attributed to Bates, Goldberg, *et al.* (2005).



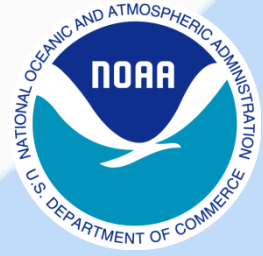
NODC Levels of Stewardship



Byrne, 2012



Levels of Scientific Data Stewardship: Level 1

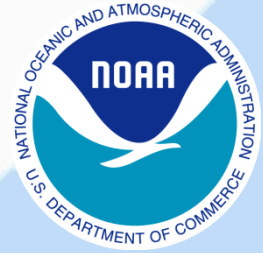


- 1. Long-term preservation and access:** Store and preserve data and metadata created by others and create (in an automated fashion or otherwise) brief metadata to describe it. For example: Ingest data or data products and associated metadata, confirm the accuracy of that ingest – that the bits and bytes have been accurately transferred into the archive. Ingest may be manual or automated. **By NDSA Levels of Digital Preservation:**

	Level 1 Protect	Level 2 Know	Level 3 Monitor	Level 4 Repair
Storage and Geographic Location	●	●	●	●
Data Integrity	●	●	●	●
InfoSec	●	●	●	●
Metadata	●	●	●	●
Formats	●	●	●	●

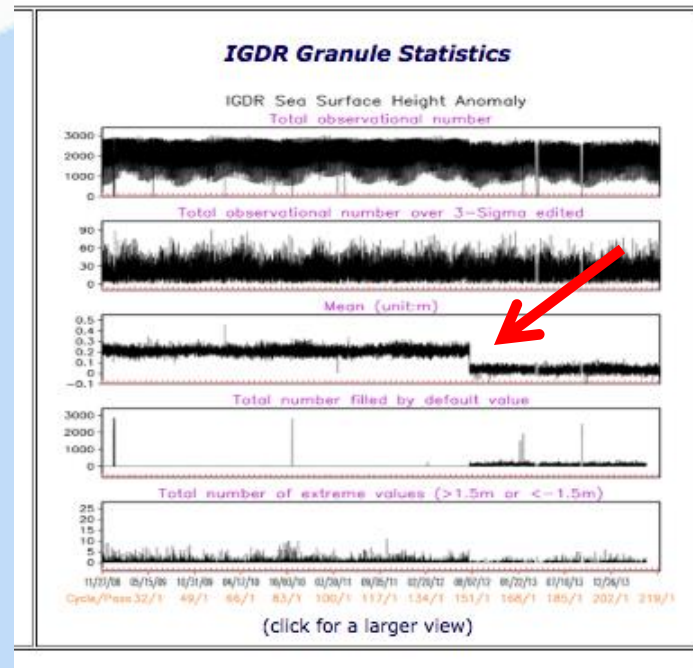
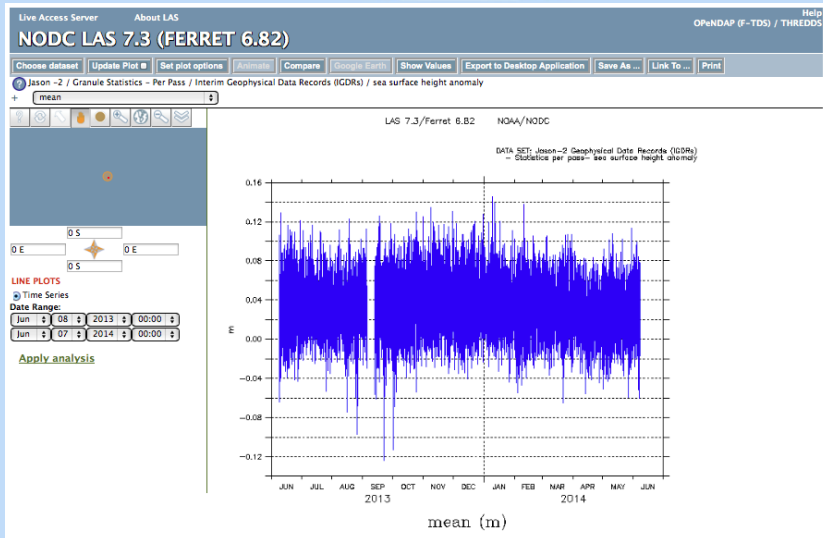


Levels of Scientific Data Stewardship: Level 2



- 2. **Tailored Access and Rich Metadata:** Describe in more detail (e.g., full ISO record), characterize qualitatively and quantitatively using automated systems, and provide enhanced access to data. For example: Data is converted to a common format for improved access and sub-sampling. Dissemination and sub-sampling, such as OPeNDAP, ArcGIS server, or Live Access Server.

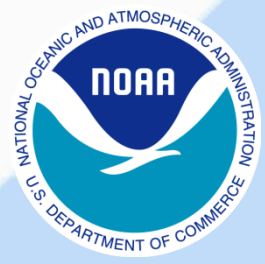
Quality Monitoring Interface



Live
 Automated
 Accessible
 Pro-active



Levels of Scientific Data Stewardship: Level 3



3. Scientific QC-Data Improvements: Improve data with scientific quality assessments and control, flagging, and corrections. Requires SME. For example: Assess the quality of, calibrate, reprocess, or otherwise improve the data. This might include calculations of aggregate means and standard deviations over a geographic region and the characterization of each measurement within that region in terms of those quantities.

Quality Control Cookbook for XBT Data*
(* Expendable Bathythermograph Data)
Version 1.1

R. Bailey, A. Gronell, H. Phillips, E. Tanner,
and G. Meyers

2.8 Sippican MK-9 Timing Delay Problem or Driver Error (DR)

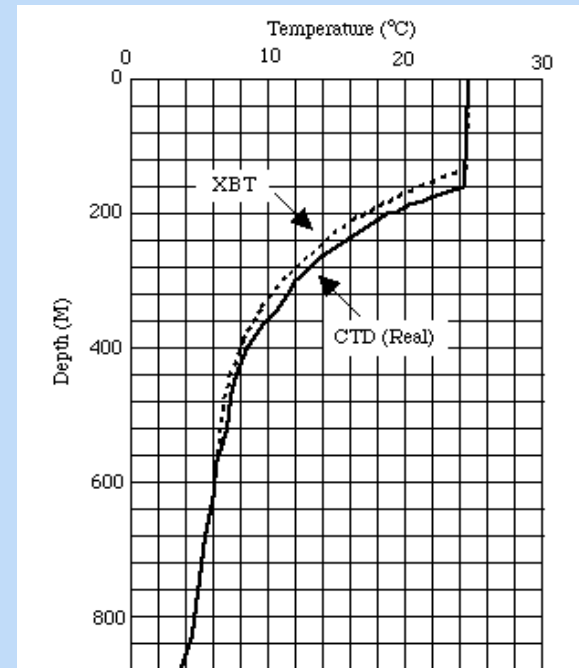
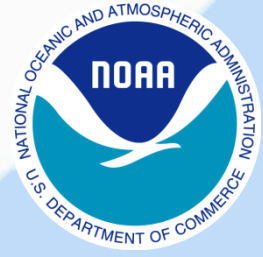


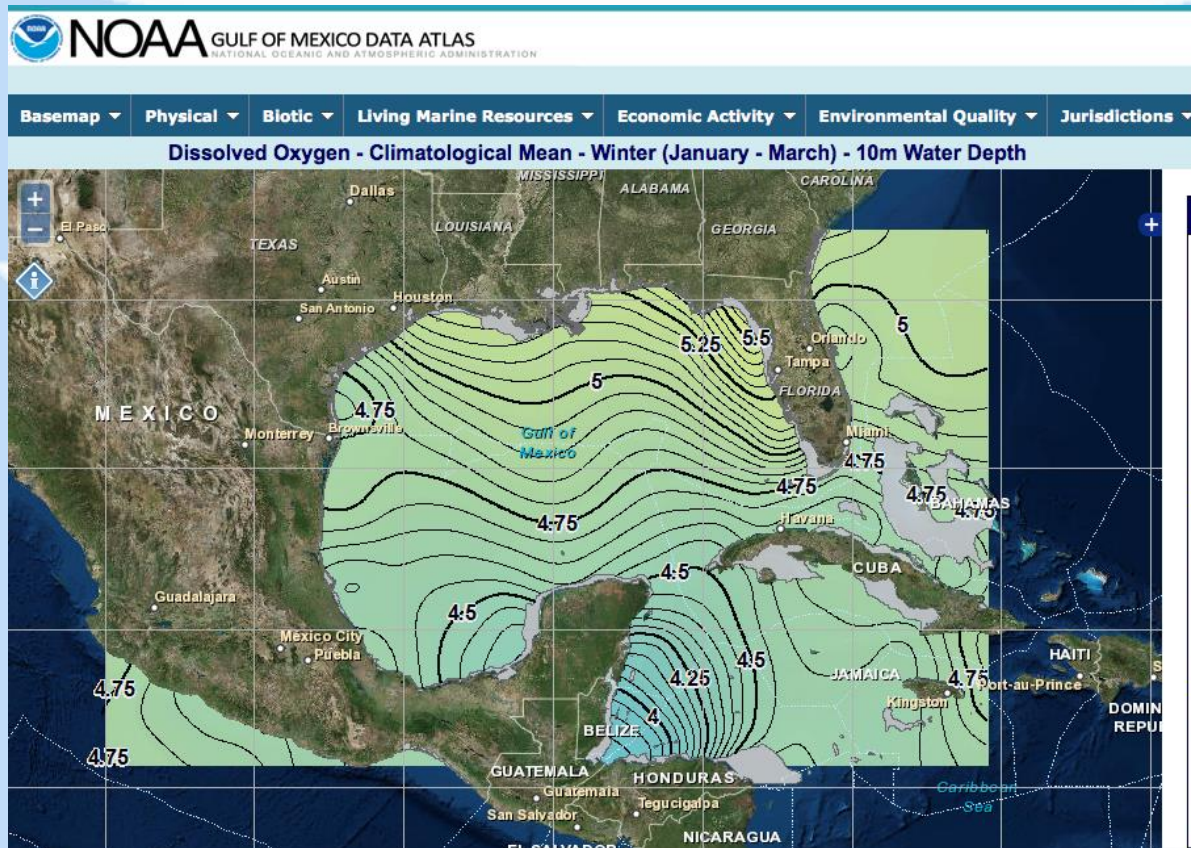
Figure 2.8 Timing Delay Problem or Driver Error (DR)



Levels of Scientific Data Stewardship: Level 4

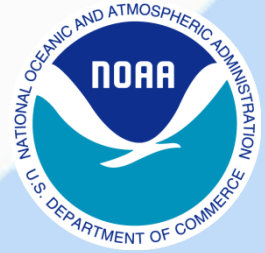


- 4. Aggregated Products:** Derive products from aggregations of data. Databases, climatologies, etc. In every case, information is augmented.

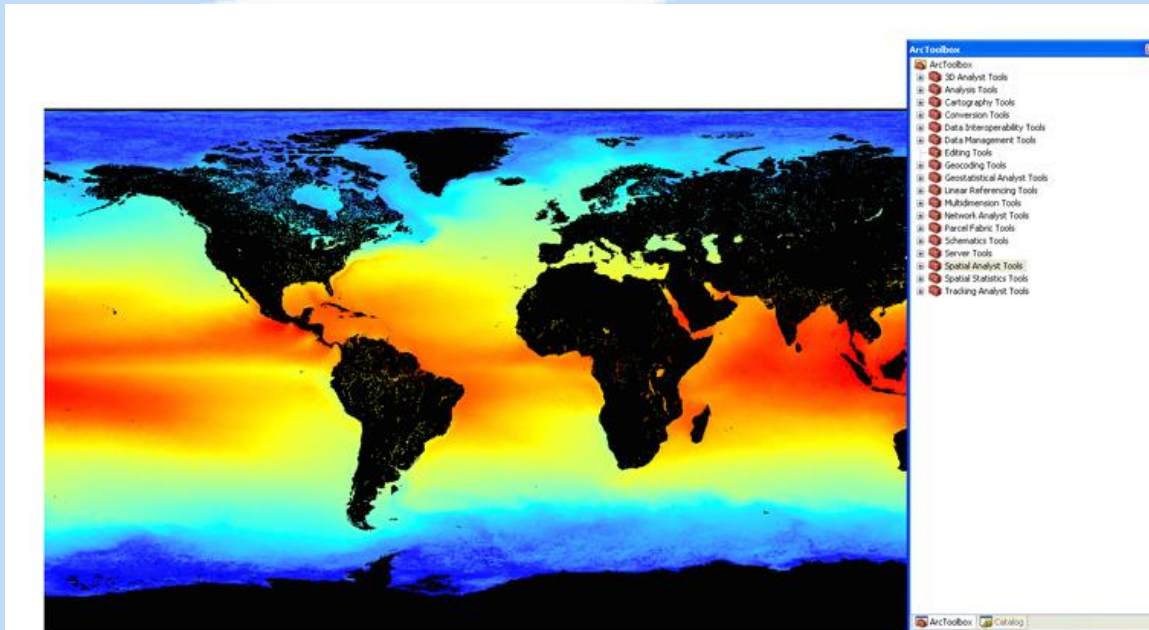




Levels of Scientific Data Stewardship: Level 5



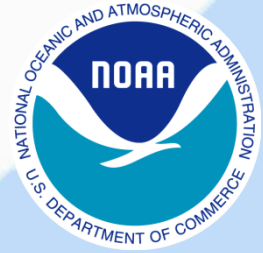
5. **Create Authoritative Records:** Recognized authority for a particular variable and/or product. For example, *de facto* acceptance of a product as a standard by scientists, resource managers, and the body public; official sanctioned climate data records; participation in cross-community intercalibration activities (e.g., XBT fall rate correction tables and reprocessed data, WOD, ocean heat content, and similar time series) and inter-calibration of multiple platforms.



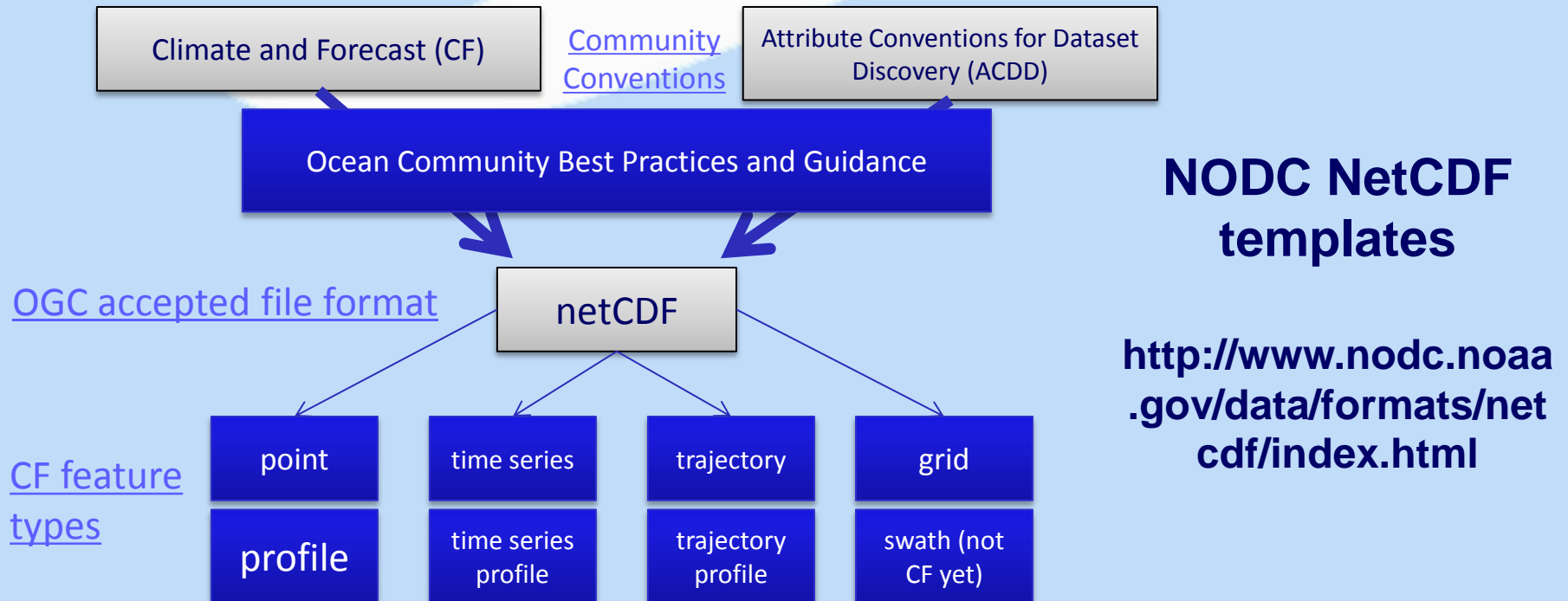
NOAA Climate Data Record: Pathfinder Sea Surface Temperature (SST) is based on the inter-calibration of AVHRR data from eight satellites



Levels of Scientific Data Stewardship: Level 6



6. National Services/Community Leadership: Provide service at a national level and/or lead the community in the development of standards or in the practice of scientific stewardship of ocean data. For example, lead the national and international oceanographic communities in coordinating or in implementing scientific stewardship activities, engaging in these activities in such a way that NODC personnel are setting a standard against which other institutes measure their activity.



**NODC NetCDF
templates**

<http://www.nodc.noaa.gov/data/formats/netcdf/index.html>

OGC accepted file format

CF feature types

Climate and Forecast (CF)

Community Conventions

Attribute Conventions for Dataset Discovery (ACDD)

Ocean Community Best Practices and Guidance

netCDF

point

time series

trajectory

grid

profile

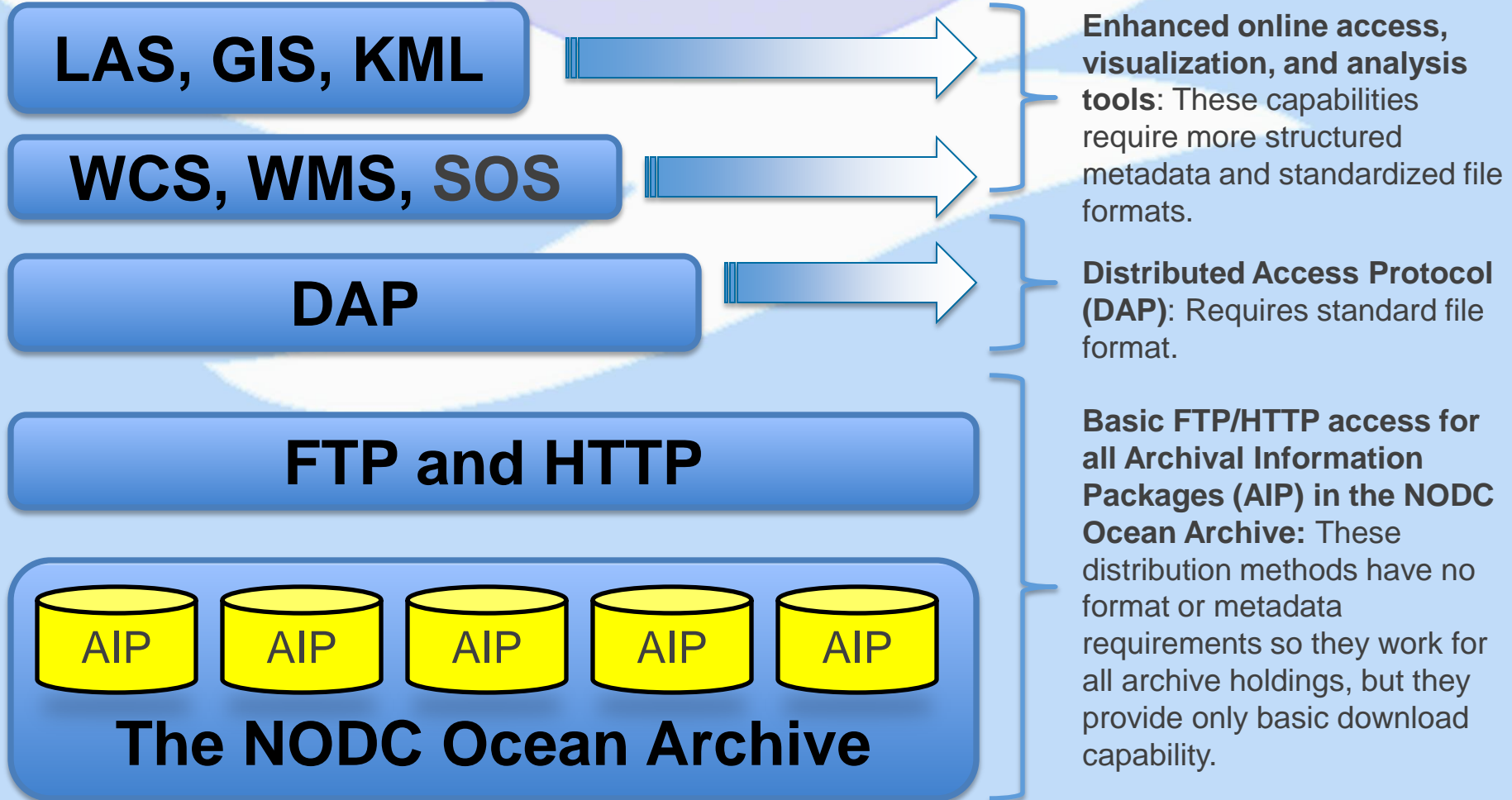
time series profile

trajectory profile

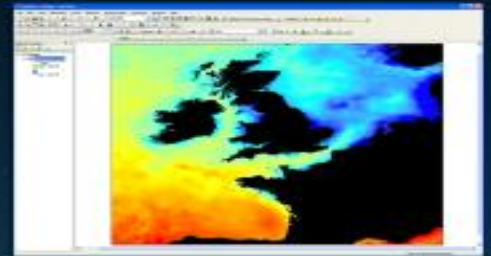
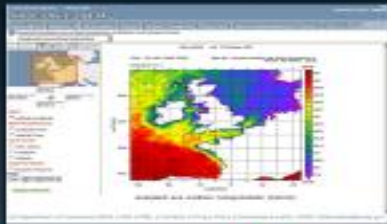
swath (not CF yet)



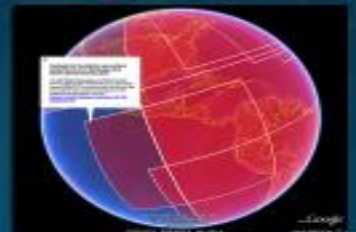
NODC standard protocols: enhancing Access and Use



Access: functions to discover, understand, and use data stored in the NODC Archive



Metadata of Service Based on ISO 19115	
Service Title	...
Service Identifier	...
Service Type	...
Service Version	...
Service Keywords	...
Service Provider	...
Service User Roles	...
Service Capabilities	...
Service Fees	...
Service Access Constraints	...
Service Contact Information	...



Metadata of Service Based on ISO 19115

Service Title: ...

Service Identifier: ...

Service Type: ...

Service Version: ...

Service Keywords: ...

Service Provider: ...

Service User Roles: ...

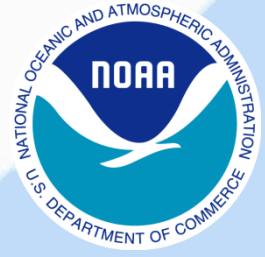
Service Capabilities: ...

Service Fees: ...

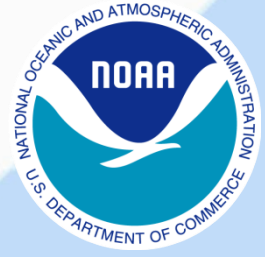
Service Access Constraints: ...

Service Contact Information: ...

Preserve. Discover. Access. Use.



Backup Slides

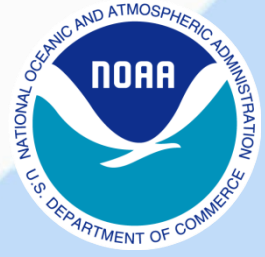


The OAIS Reference Model

- The Open Archival Information System Reference Model (**OAIS RM**) is the CCSDS and ISO Standard (14721) for Digital Archives
- Formulated with space-based data in mind, it applies to all organizations that need to preserve digital information for the long-term
- It does NOT specify a particular implementation
- The **NOAA National Oceanographic Data Center** (NODC) is an archive conforming to the OAIS-RM and the designated archive for NJGS data and information
- We also work to conform to guidance from the National Archives and Records Administration (NARA), and work *with* NARA and LOC on issues surrounding preservation of space data.



OAIS Responsibilities

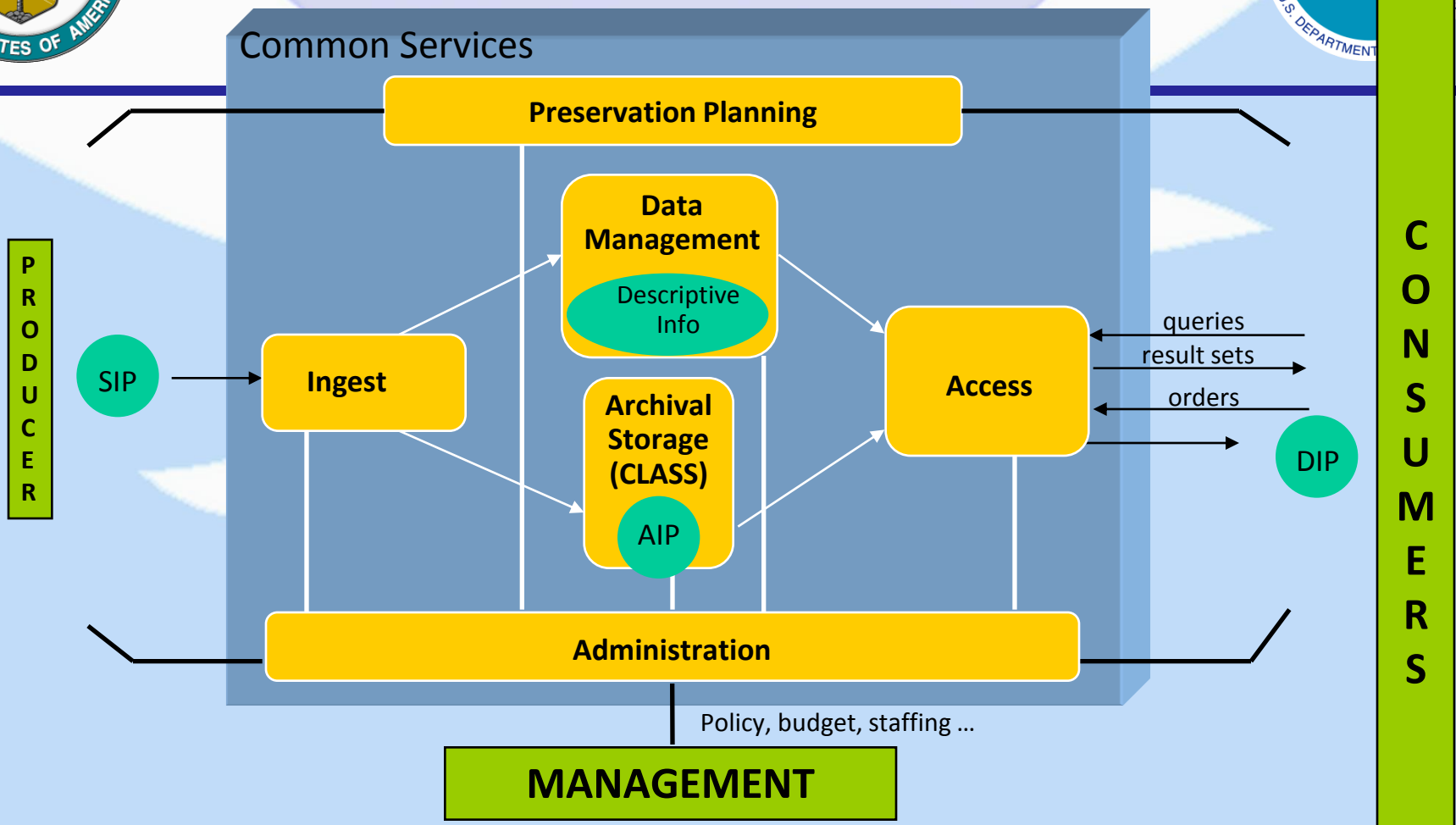
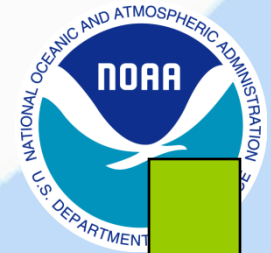


An Archive must ...

- Negotiate and accept information from **Data Producers**
- **Obtain sufficient control of the data** to ensure its long-term preservation
- Ensure the information to be preserved is **independently understandable - without expert assistance** to identified **Designated Communities**
- Develop, document and follow **standard policies and procedures** to insure information is preserved
- **Provide information** to the Designated Communities in understandable forms using standardized protocols
- The highest level of scientific stewardship includes provision of **value-added products and services**

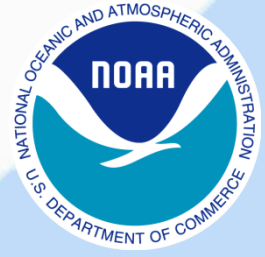


OAIS Functional Entities



SIP = Submission Information Package
 AIP = Archival Information Package
 DIP = Dissemination Information Package

The OAIS Environment
 from 10,000 ft



OAIS Functional Entities

Ingest: Accept Submission Information Packages (SIPs) from Producers and prepare the contents for storage and management

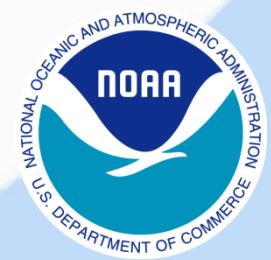
Archival Storage: Provides for the storage, maintenance and retrieval of Archival Information Packages (AIPs)

Data Management: Provide services and functions for populating, maintaining, and accessing both descriptive information that identifies and documents OAIS holdings and internal OAIS administrative data

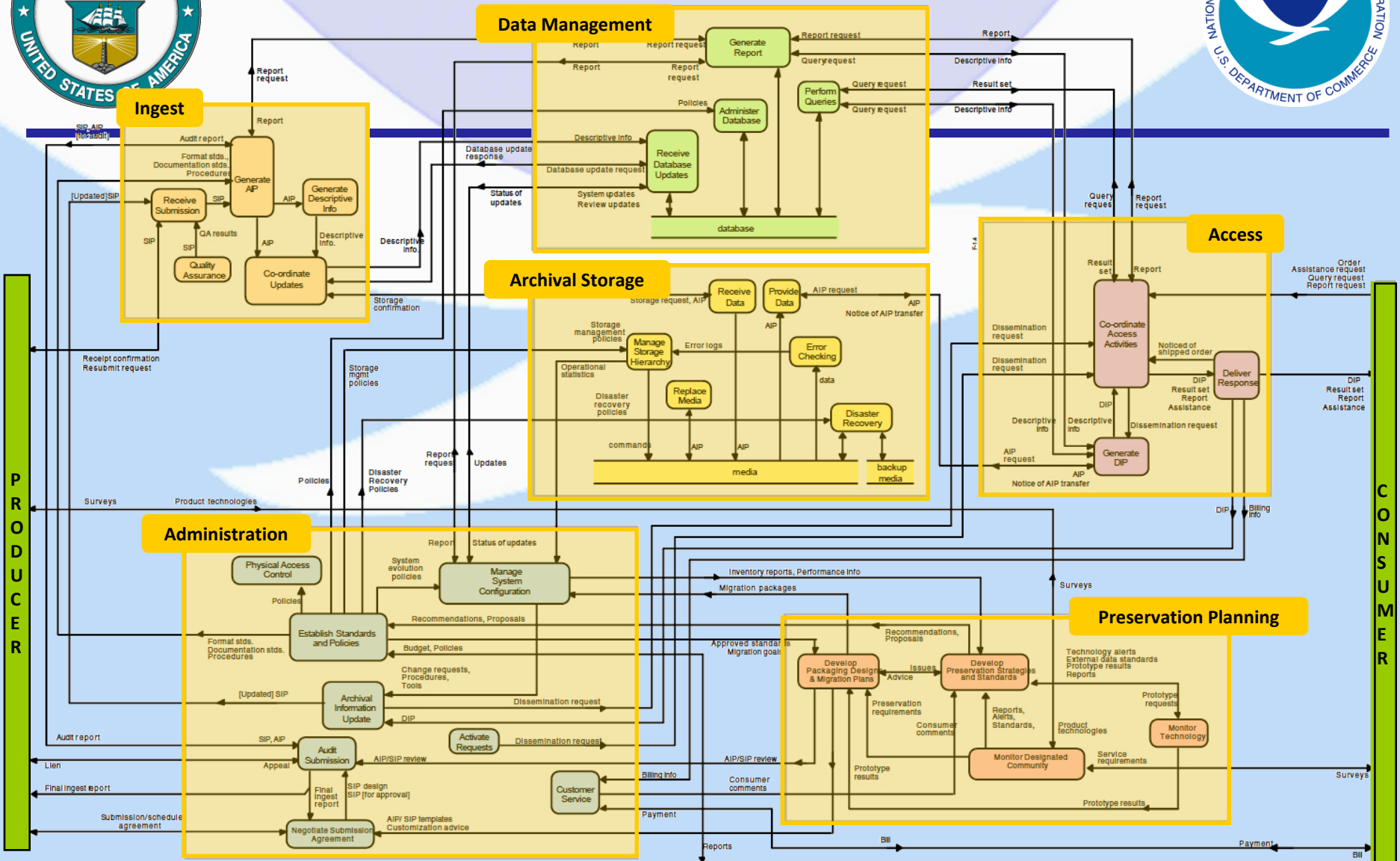
Access: Supports Data Discovery and Data Access - How users determine the existence, description, location and availability of information stored in the OAIS and allows Consumers to request and receive information and data.

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.

Administration: Manage the overall operation of the OAIS - Negotiate Submission Agreement, Manage Archive System Configuration and Archival Information Update. Provide Physical Access Control, Establish Standards and Policies, meet Requests for Data, provide Customer Service



OAIS Functional Entities

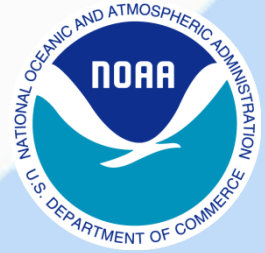


The OAIS Environment from Sea Level

Adapted from Figure F-1: Composite of Functional Entities



Linking Collections and Granules



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

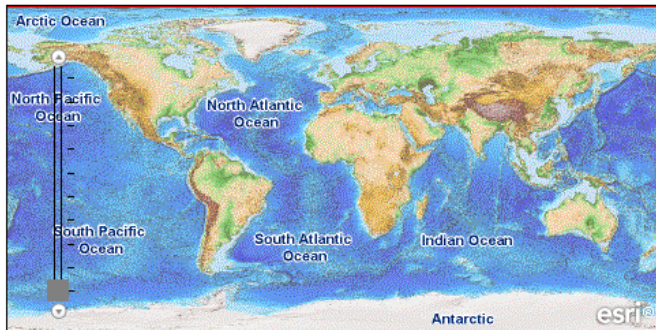
From: (yyyyymmdd)

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) [1](#) [2](#) [3](#) [4](#) [5](#) [>](#) [Last](#)

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](#)

