



IOOS Regional Association Archive/Metadata Training Summary

Ted Habermann
NOAA National Geophysical Data Center

Kenneth S. Casey
NOAA National Oceanographic Data Center

Kenneth.Casey@noaa.gov
Ted.Habermann@noaa.gov



Summary of Archive Training



http://en.wikipedia.org/wiki/Mad_scientist

An experiment of sorts...

- 7 RA data managers and NODC, NGDC, NDBC, and IOOS Office
- **Goal:**
 - Show how to submit data to NODC for documentation, preservation, and access
- **Approach:**
 - Emerging netCDF templates and automation elements, coupled with good ISO enable discovery and use, AND preservation and understanding;
 - Producer-Archive relationships



Summary of Archive Training

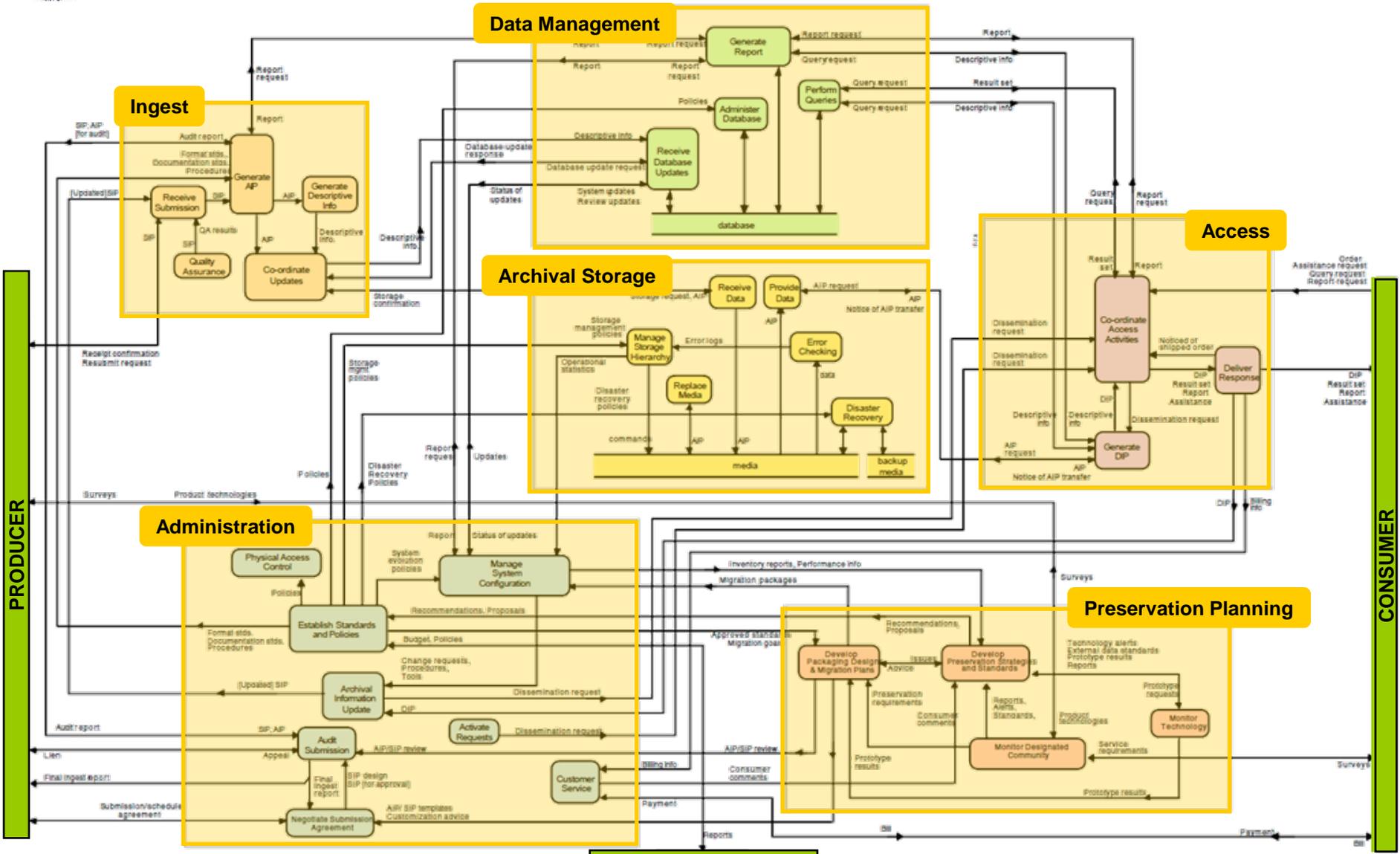


http://en.wikipedia.org/wiki/Mad_scientist

- 1. Background on “Capital A” Archives**
- 2. netCDF templates**
- 3. ISO 19115**
- 4. Archive automation**
- 5. Producer-Archive relationships**
- 6. NDBC Case Study**



OAIS Functional Entities



The OAIS Environment from Sea Level

Adapted from **Figure F-1: Composite of Functional Entities**



NODC *Archive* Paradigm

Human and machine
interfaces to
*Discover, Use, Preserve, and
Understand*
NODC Archive Holdings

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



The NODC Ocean Archive

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



Discovery

Discovery

Human to Machine Interfaces

Google

Data.gov

GOS

OAS

Geoportal Server Web App

Machine to Machine Interfaces

CSW

Geoportal Server REST API

OpenSearch

SRU/ISO23950

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 Geospatial One-Stop (GOS)

AIP

AIP

AIP

AIP

AIP

The NODC Ocean Archive

Discovery services are available for ALL of the NODC Archive holdings, but better metadata supports better discovery!



Access and Use

LAS, GIS, KML



WCS, WMS, SOS



DAP



FTP and HTTP



The NODC Ocean Archive

Enhanced online access, visualization, and analysis tools: These capabilities require more structured metadata and standardized file formats, so are available to the fewest archive holdings.

Distributed Access Protocol (DAP): Requires standard file formats so is available to fewer archive holdings.

Basic FTP/HTTP access for all Archival Information Packages (AIP) in the NODC Ocean Archive: These distribution methods have no format or metadata requirements so they work for all archive holdings, but they provide only basic download capability.



Understand and Preserve

Understand and Preserve

Metadata and Transformations

FGDC

ISO

Ad hoc docs

Other standards

Provenance

Checksums

Manifests

Multiple Views

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

As with Preservation, better documentation enables better understanding and long term preservation.



Archive Services and MORE!

- Not only does NODC provide long term preservation services, but we are also active partners in providing IOOS standard services
 - OPeNDAP's Data Access Protocol via Hyrax
 - OGC's WMS and WCS via THREDDS Data Server
 - THREDDS Catalog Services and Aggregation
 - Coming soon OGC's Catalog Search for the Web (CSW) and Search/Retrieval via URL (SRU), and someday Sensor Observation Service (SOS)
 - Interactive web mapping applications
 - Also Live Access Server (LAS) and ArcGIS Server



In other words, an Archive is a whole lot more than a backup copy of data on your shelf

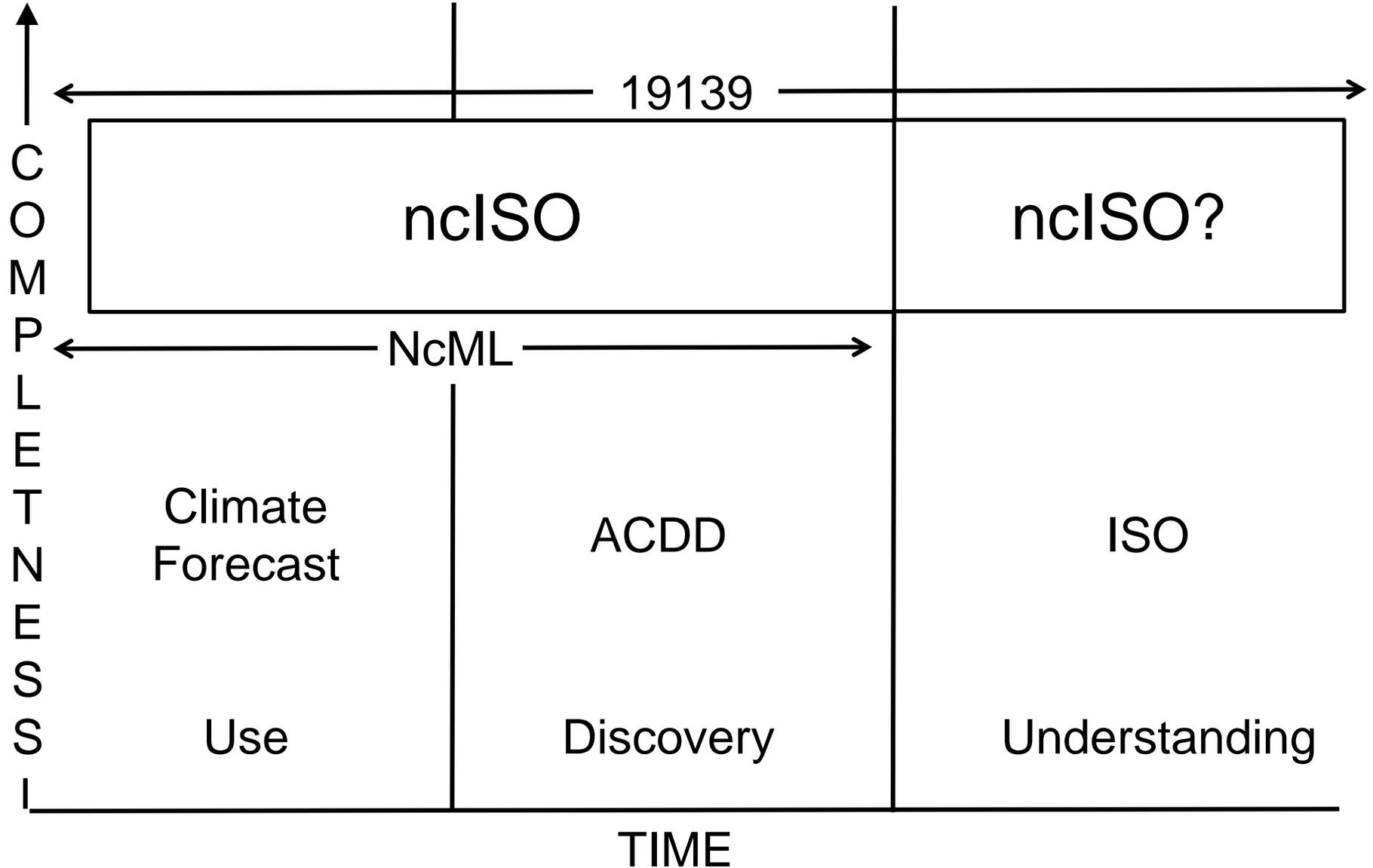


Archive Training

- netCDF templates
 - CF and ACDD, with best practices, for each featureType, starting with trajectory and profile
- Connections with ISO 19115
 - via ncISO response of the TDS; enables *understanding*
- Establishing archive automation
 - Checksums, manifests, hand-shaking
- Building a relationship
 - *Producer and Archive* working together



Documentation Completeness





Spiral Development / Training: Rubrics

Creating Good Documentation - GEO-IDE Guidelines and Best Practices Wiki - Mozilla Firefox

Creating Good Documentation - GEO-IDE Guidelines and Best Practices Wiki - Mozilla Firefox

http://www.ngdc...8day_REPORT.html

NetCDF Attribute Convention for Dataset Discovery Report

The Unidata Attribute Convention for Data Discovery is a standard for describing those files using standard metadata searches. This report identifies ISO metadata elements described in spirals of documentation development described in [Creating Good Documentation](#). Together these spirals build a strong foundation for high-quality documentation. The ISO Standard includes a number of options for building on that foundation by addressing specific scientific needs. See [Use Cases to CRUD](#) for some examples.

The elements are listed by name and are followed by M, C, or O if they are Mandatory, Conditional or Optional. They are followed by UDD (attribute name) if they are included in the [NetCDF Attribute Convention for Dataset Discovery](#).

The ISO 19115 Standard recommends [Core Elements](#) for inclusion in metadata. This tool tests also conformance with those recommendations.

The Rubric at the top of the report summarizes the results. Each spiral is represented by a row in the rubric. The columns show the % of the elements in that spiral that exist in the record. Click the spiral name for more details.

This report is produced using this [stylesheet](#). Please contact [Ted Habermann](#) if you have questions or suggestions.

Title: Chlorophyll-a, Aqua MODIS

Total Score: 28/43

General File Characteristics

Number of Global Attributes 60
 Number of Variables 5
 Number of Variable Attributes 45
 Number of Standard Names 5

Spiral	None
Total	
Identification	
Text Search	
Extent Search	
Other Extent Information	
Creator	
Contributor	
Publisher	X
Other Attributes	

Identification Information

Done

Find: aname

http://www.ngdc...ew=isoRubricHTML

ISO 19115 Report

This report identifies ISO metadata elements described in spirals of documentation development described in [Creating Good Documentation](#). Together these spirals build a strong foundation for high-quality documentation. The ISO Standard includes a number of options for building on that foundation by addressing specific scientific needs. See [Use Cases to CRUD](#) for some examples.

The elements are listed by name and are followed by M, C, or O if they are Mandatory, Conditional or Optional. They are followed by UDD (attribute name) if they are included in the [NetCDF Attribute Convention for Dataset Discovery](#).

The ISO 19115 Standard recommends [Core Elements](#) for inclusion in metadata. This tool tests also conformance with those recommendations.

The Rubric at the top of the report summarizes the results. Each spiral is represented by a row in the rubric. The columns show the % of the elements in that spiral that exist in the record. Click the spiral name for more details.

This report is produced using this [stylesheet](#). Please contact [Ted Habermann](#) if you have questions or suggestions.

Title: Aerosol Optical Thickness (100 KM)

Total Spiral Score: 27/41

Spiral	None ☆☆☆☆	1-33% ★☆☆☆	34-66% ★★★☆	67-99% ★★★★★	All ★★★★★
Total Spiral					
Identification					
Extent					
Connection					
Distribution					
Description					
Content					
Lineage					
Acquisition Information					

Done

Example: Archiving NDBC Data at NODC

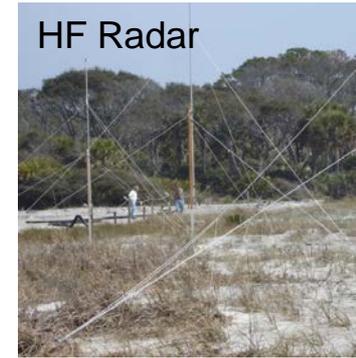
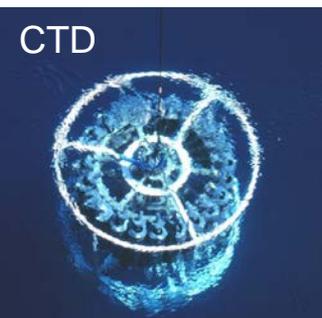
1. Modernize and automate ingest and archive services for:



- Weather Buoy and Coastal Automated Stations (C-MAN)
- Tropical Atmosphere-Ocean (TAO) array
- CTD casts taken during maintenance of TAO buoys

2. Examine and assess for possible inclusion in the NODC archive:

- Partner Buoys
- High-Frequency (HF) Radar
- OceanSITES
- Ocean Current Profiling at Oil and Gas Platforms (ADCP data)





Feedback from RAs

- You had me at hello
- Positive so far..
- Still widespread misunderstanding of *Archives*
- Emilio Mayorga