U.S. Integrated Ocean Observing System (IOOS®)

Improving Aquatic Biological Observing Data Integration and Dissemination

A Multi-Agency, Multi
Data/Metadata Stand
Integrated Aquatic

Hassan Mous

NOAA EDMC

US IOOS: Program overview

IOOS BDP and Bio. Core Var.

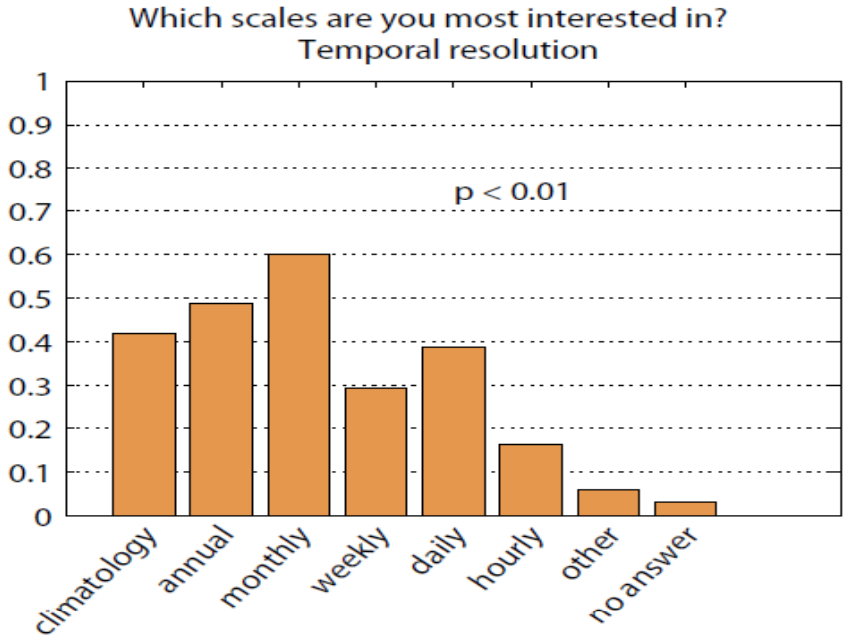
ATN Obs
- TOPP to Ocean models
- POST to Fisheries models

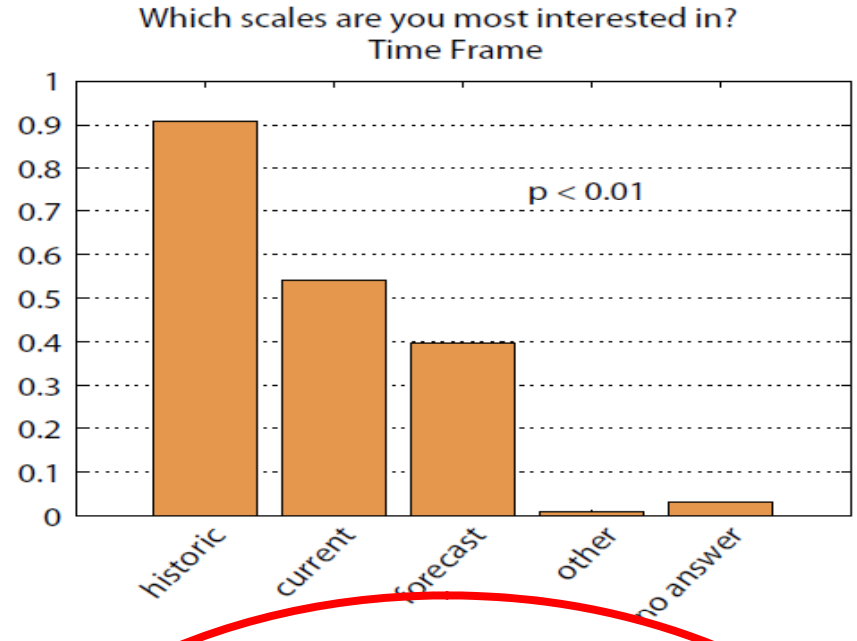
Acoustic Obs

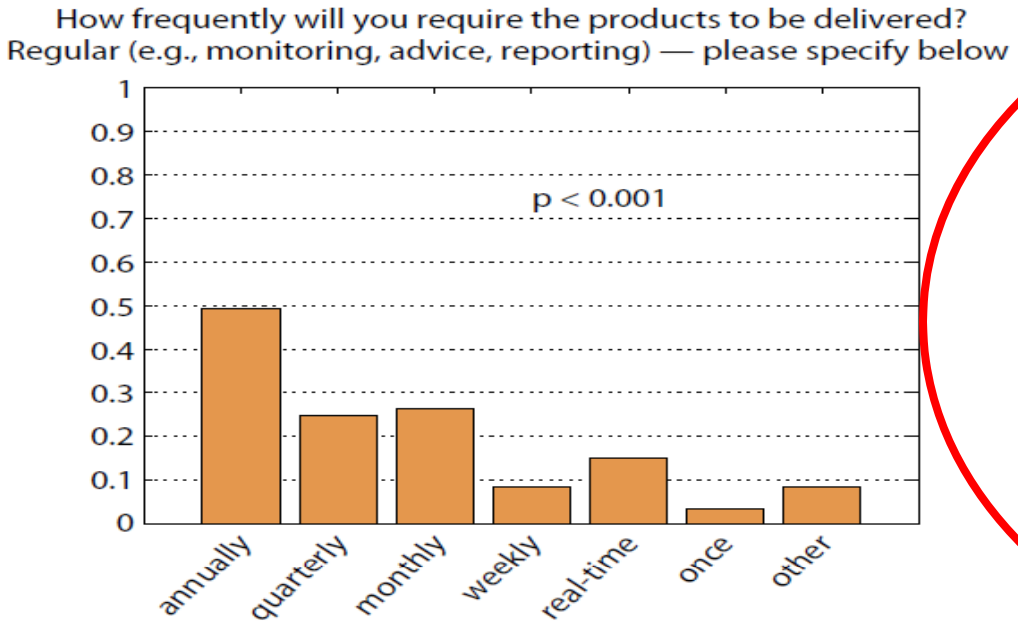
Take Home Points

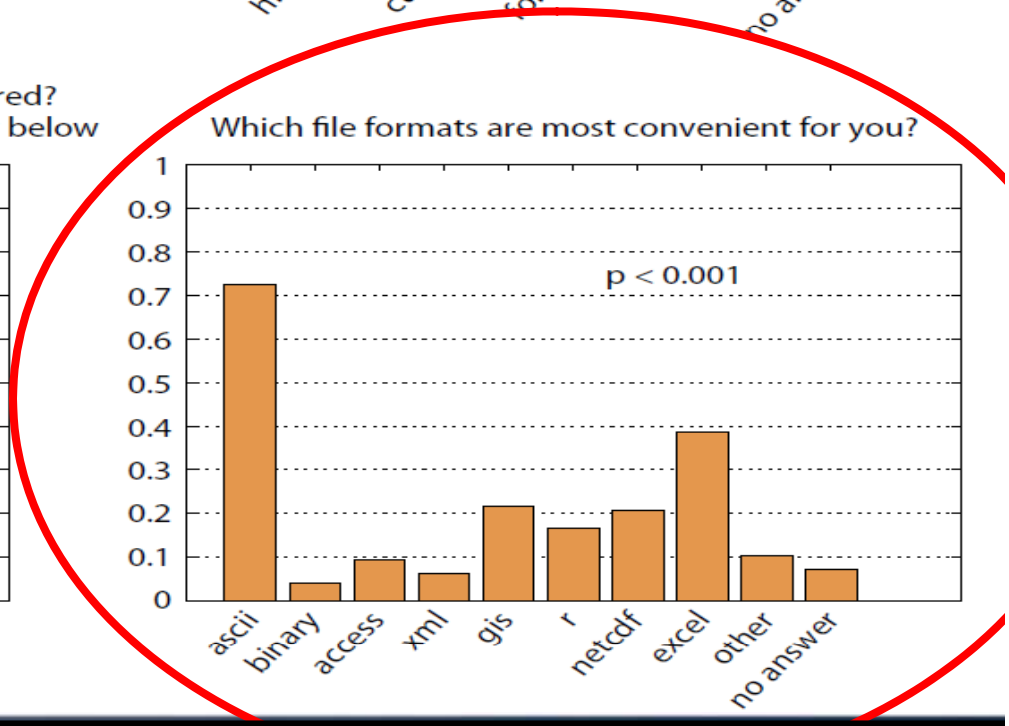
Temperature
Salinity
Water level
Currents
Surface Waves
Surface Winds
Ocean color
Dissolved nutrients
Optical properties
Total suspended matter
Dissolved oxygen
pH
Colored dissolved organic matter
Species occurence
Species abundance

U.S. Integrated Ocean Observing System (IOOS)


Which scales are you most interested in?
Temporal resolution
p < 0.01
climatology
annual
monthly
weekly
daily
hourly
other
no answer


Which scales are you most interested in?
Time Frame
p < 0.01
historic
current
forecast
other
no answer


How frequently will you require the products to be delivered?
Regular (e.g., monitoring, advice, reporting) — please specify below
p < 0.001
annually
quarterly
monthly
weekly
real-time
once
other


Which file formats are most convenient for you?
p < 0.001
ascii
binary
access
xml
gis
r
netcdf
excel
other
no answer

CURRENT METHODS FOR DATA INTEGRATION STORING AND ACCESSING BIO. / ECO. DATA

Data integration: **Jones et al 2006 (NCEAS).**
Matching up and combining information from different sources, ideally in ways that are meaningful and useful` ``

Spreadsheets Soft **Relational database Systems** such as MSAccess or Statistical package such as SAS, R Note: Difficult to share databases and spreadsheets with colleagues	**Highly structured, vertically integrated data (Federated Data schema)** (e.g. GenBank, VegBank, and TreeBase, OBIS.. **Network distributed access to data from a number of compatible servers (e.g.** GBIF etc.)	**Metadata-Driven Data (Data Collections)-** Loosely structured collection of project-specific data sets accompanied by structured metadata about each of the data sets. (metadata content specifications e.g. Ecological Metadata Language (EML), NBII Biological Data Profile (BDP), etc.)

Data collections created to promote data sharing and long-term preservation of ecology and biology data

Jones et al. 200

Collection name	Standards supported	Archives data
Knowledge Network for Biocomplexity	EML, BDP, others	yes
NBII Metadata Clearinghouse	BDP	no[a]
NSDI Metadata Clearinghouse	CSDGM	no[a]
GBIF Taxonomic Collections	Darwin Core	no[a]
TOPP	EML	yes
Kruger National Park/SAEON	EML	yes
Open-Source Project for a Network Data Access Protocol (OPeNDAP) / Integrated Ocean Observing System (IOOS)	—	yes
VegBank	US NVC	yes
TreeBase	—	yes
Storage Resource Broker	various	yes
ORNL DAAC	BDP, other	yes
Global Change Master Directory	DIF	no[a]
ESA Data Registry	EML	no[a]
Ecological Archives (data papers)	various	yes
Ocean Biogeographic Information System (OBIS)	Darwin Core[b]	no[a]
Global Population Dynamics Database (GPDD)	various	yes

Metadata Standards, Content and Access

Metadata: information used to document and interpret data

Metadata Contents

- *Purpose*
  - Identification / Attribution
  - Discovery
  - Understanding
  - Referral
  - History / QA
  - (Details of Mapping to Terminology)
- *Content Standard*

FGDC is the Primary Format
- ISO and ncISO Supported
- Multiple Interface Methods

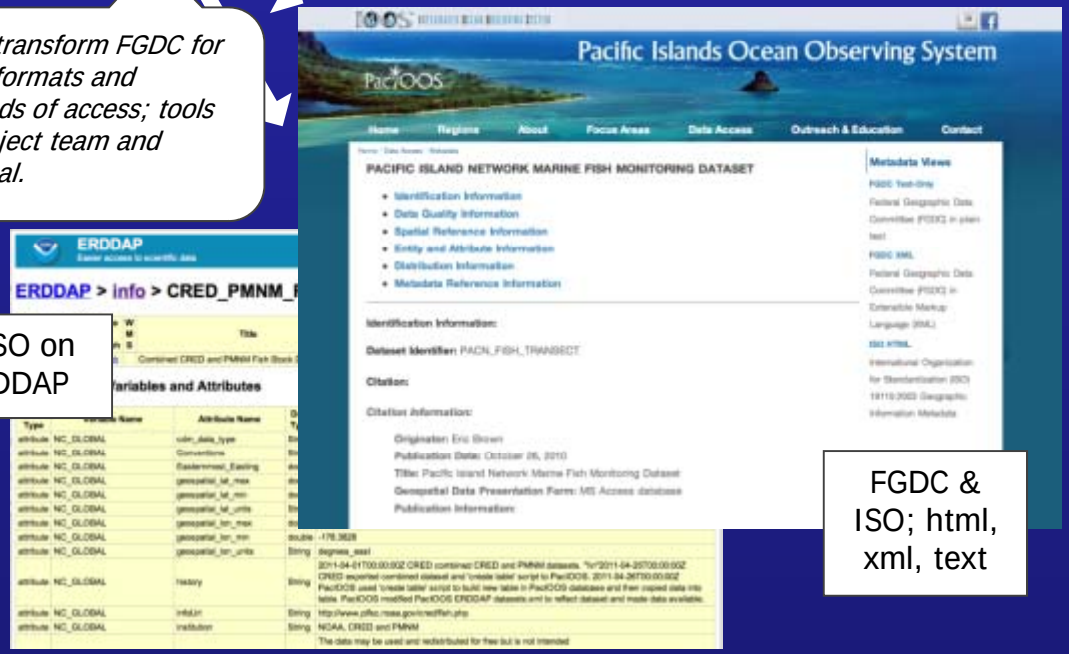
Metadata Access

FGDC Document of Record

Native FGDC xml via http://

Converted ISO xml via http://

Tools transform FGDC for other formats and methods of access; tools by project team and external.



ncISO on ERDDAP

FGDC & ISO; html, xml, text

Credit to OBIS-USA, PIFSC , NPS, PMNM and PacIOOS team

BDP data services and next plans?

- Data Terminology and exchange standards

- Data Model XML

http://www.ioos.gov/schema/ioosbiology/0.9/ioosbiologicalterminology_vpoint9.xsd

-Terminology Schema

http://www.ioos.gov/schema/ioosbiology/0.9/ioos_biological_terminology_vpoint9d.xml

-Data Access services

http://pacioos-mapserver2.ancl.hawaii.edu/erddap/index.html

Metadata Standards

http://oos.soest.hawaii.edu/cgi-bin/get_metadata.pl?id=PACN_FISH_TRANSECT&format=fgdc

asa
science. services. solutions.

São Paulo
Shanghai
Gold Coast
Perth

About ASA | Services | Software | News | Projects | Login & Support | Search AS

Software

- Software/Technology Solutions
- ⊞ OILMAP™
- SARMAP™
- CHEMMAP™
- Environmental Data Server™
- AIRMAP™
- HYDROMAP™
- MUDMAP™
- SIMAP™
- WQMAP™
- ⊞ ASA In-house Tools
- ⊞ ArcGIS® Tools

GIS & Data Integration

ASA Software Downloads

Software

Downloads : Freeware Catalogue

	Name	Size	Description
Download	**Environmental Data Connector MATLAB**	106 MB	The Environmental Data Connector (EDC) extension uses a Java-based browser to allow users to connect to THREDDS/SOS/ERDDAP data servers. The connector leverages existing components from the Unidata libraries so that users can filter large amounts of data in space and time. The data and metadata is then converted to into a convenient Matlab struct object and loaded into the workspace for analysis.
Download	**Environmental Data Connector R (Windows Package)**	86.1 MB	The Environmental Data Connector (EDC) extension uses a Java-based browser to allow users to connect to THREDDS/SOS/ERDDAP data servers. The connector leverages existing components from the Unidata libraries so that users can filter large amounts of data in space and time. The data and metadata is loaded into the R workspace as a named list for analysis.
Download	**Environmental Data Connector R (32-bit Macintosh Package)**	86.5 MB	The Environmental Data Connector (EDC) extension uses a Java-based browser to allow users to connect to THREDDS/SOS/ERDDAP data servers. The connector leverages existing components from the Unidata libraries so that users can filter large amounts of data in space and time. The data and

http://www.ioos.gov/biological_observations/welcome.html

INTEGRATED OCEAN OBSERVING SYSTEM

GLOBAL | NATIONAL | REGIONAL | Search

Home | IOOS in Action | About | Data | Observing Systems | Modeling | Themes | Education | Communications

HOME | Data | Biological Observations

Biological Observations Project

The IOOS Biological Observations Project is addressing the Data Managementand Communications (DMAC) requirements that pertain to biological observations standards and interoperability applicable to U.S. IOOS and to various observing systems. Biological observations are highly heterogeneous and the variety of formats, logical structures, and sampling methods create significant challenges.

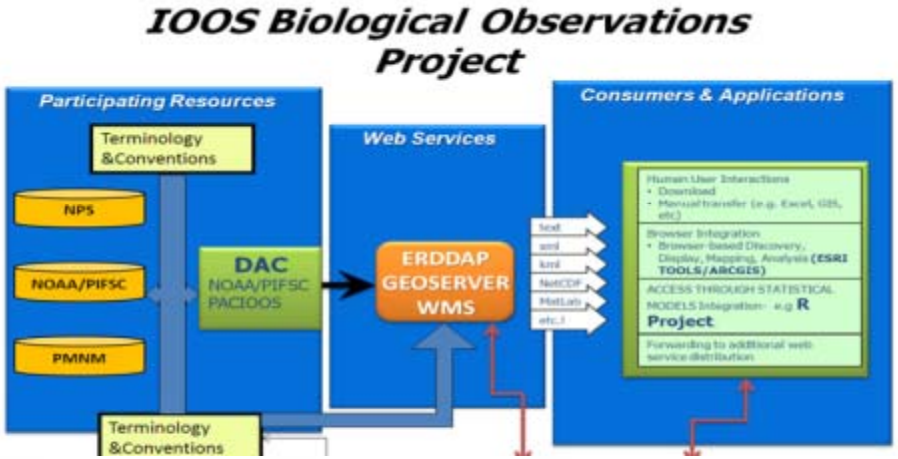
The objective of the IOOS Biological Observations Projects is to develop an efficient and effective information infrastructure for biological observations, adding components and links as necessary to serve end-users.

The U.S. IOOS Biological Observations Data Project, (IOOS BDP)

About | Details | Meetings

Project Details

System Concept and Schema/Terminology



The U.S. IOOS Biological Observations Project: Organizations and Partners Involved

This effort is comprised of a wide collaboration of organizations and partners, including:

- U.S. Integrated Ocean Observing System (IOOS)
- NOAA Pacific Islands Fisheries Science Center (PIFSC)/Coral Reef Ecosystem Division (CRED)
- Pacific Islands Ocean Observing System (PacIOOS)
- NOAA Papahānaumokuākea Marine National Monument (PMNM)
- National Park Services (NPS) Pacific Island Network (PACN)
- U.S. Geological Survey /Ocean Biogeographic Information System (USGS/OBIS-USA)
- Hawaii Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawaii, Manoa (UH/HIMB)
- U.S. Census of Marine Life (U.S CoML)
- Southwest Fisheries Science Center/ Environmental Research Division (SWFSC/ ERD)
- NOAA Fisheries/Office of Science and Technology (OST)

Biological Observations Project Flyer

Download pdf

Contact Information

Project Manager, Dr. Hassan

Next IOOS BDP plan

FY12- implementation in SECOORA and GCOOS-RA.

FY13-14- Implement in all IOOS RAs

A True Partnership effort

- U.S. Census of Marine Life ***(Michael Feldman)***

- NOAA Fisheries
  - o Pacific Islands Fisheries Science Ctr. ***(Rusty Brainard , Anntte DesRocehrs, Troy Kanemura et al)***
  - o Southwest Fisheries Science Ctr. ***(Roy Mendelssohn)***
  - o Office of Science and Technlgy. ***(Jim Sargent, Becky Shuford)***

- NOAA Sanctuaries/Monument ***(Randy Kosaki, David Graham, Steve Turner)***

- University of Hawai'i
  - o PacIOOS ***(Jim Potemra, Chris Ostrander)***
  - o HIMB ***(Jo-Ann Leong, Erik Franklin)***

- USGS and USA-OBIS ***(Mark Fornwall, Phillip Goldstein***)

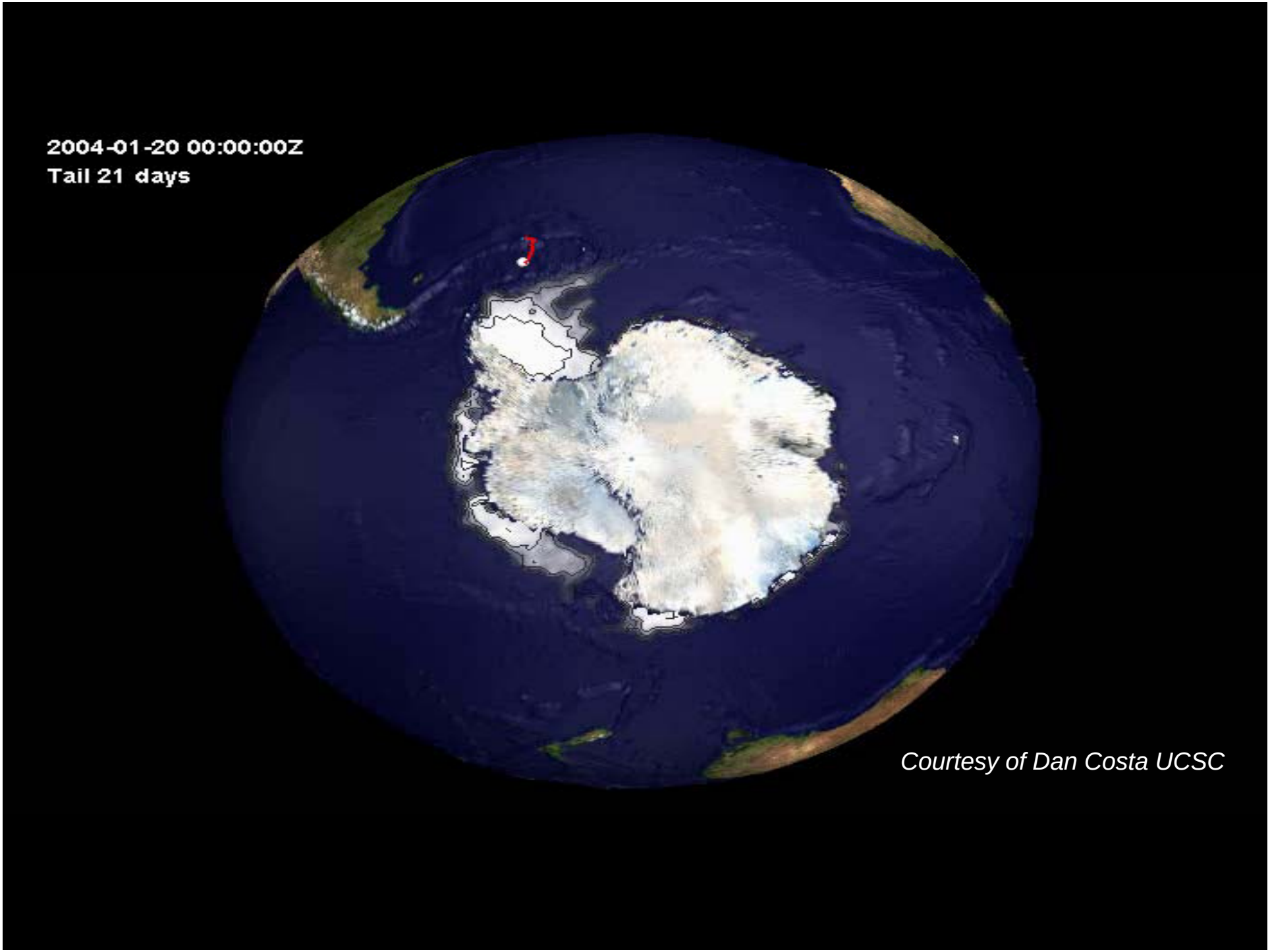
- NOAA IOOS Program ***(Charly Alexander, Hassan Moustahfid)***

Animal Telemetry Obs

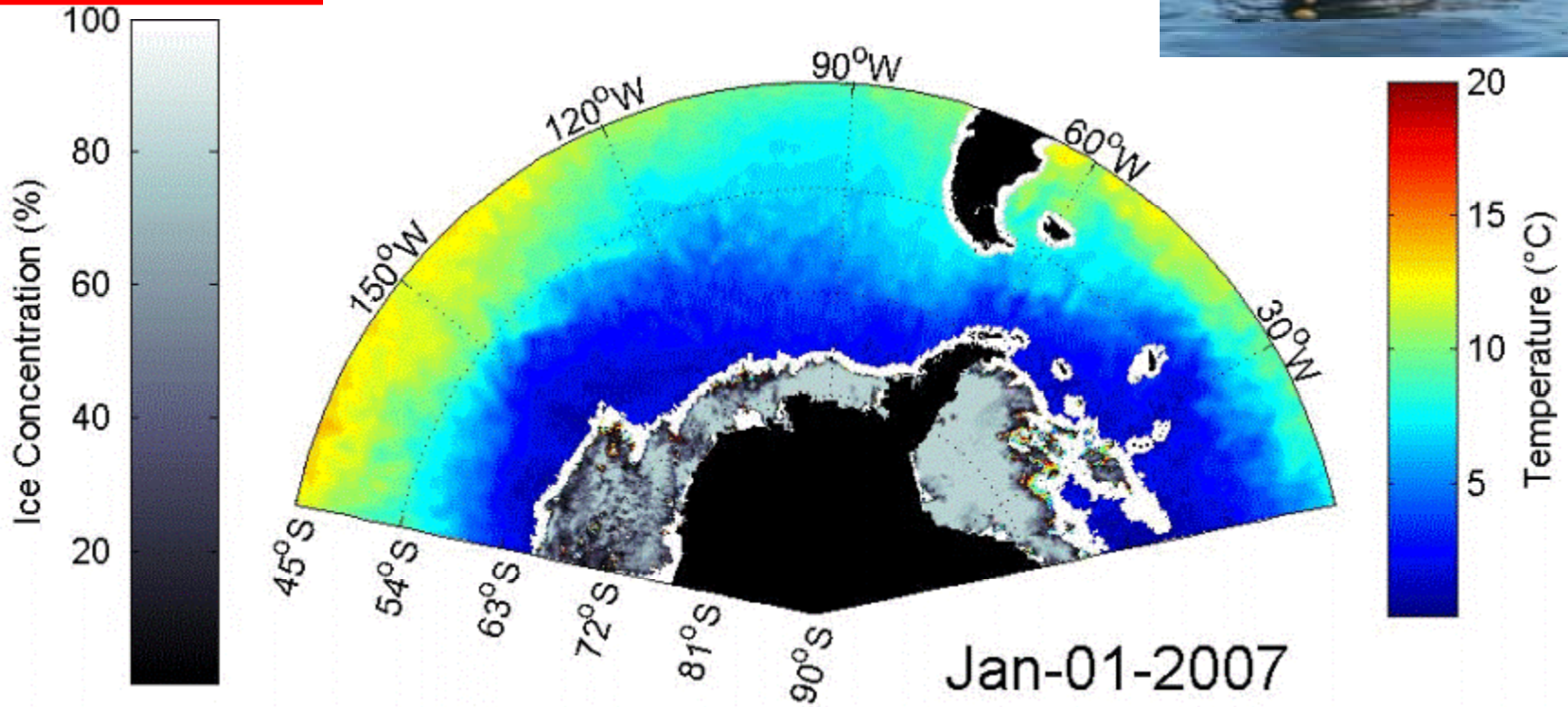
Two projects to reconcile animal telemetry data standards

- ***Project 1:*** partnering with TOPP and ONR to enable ocean modelers access to oceanographic observations collected from sensors on marine animals (e.g. Seals and Sharks);

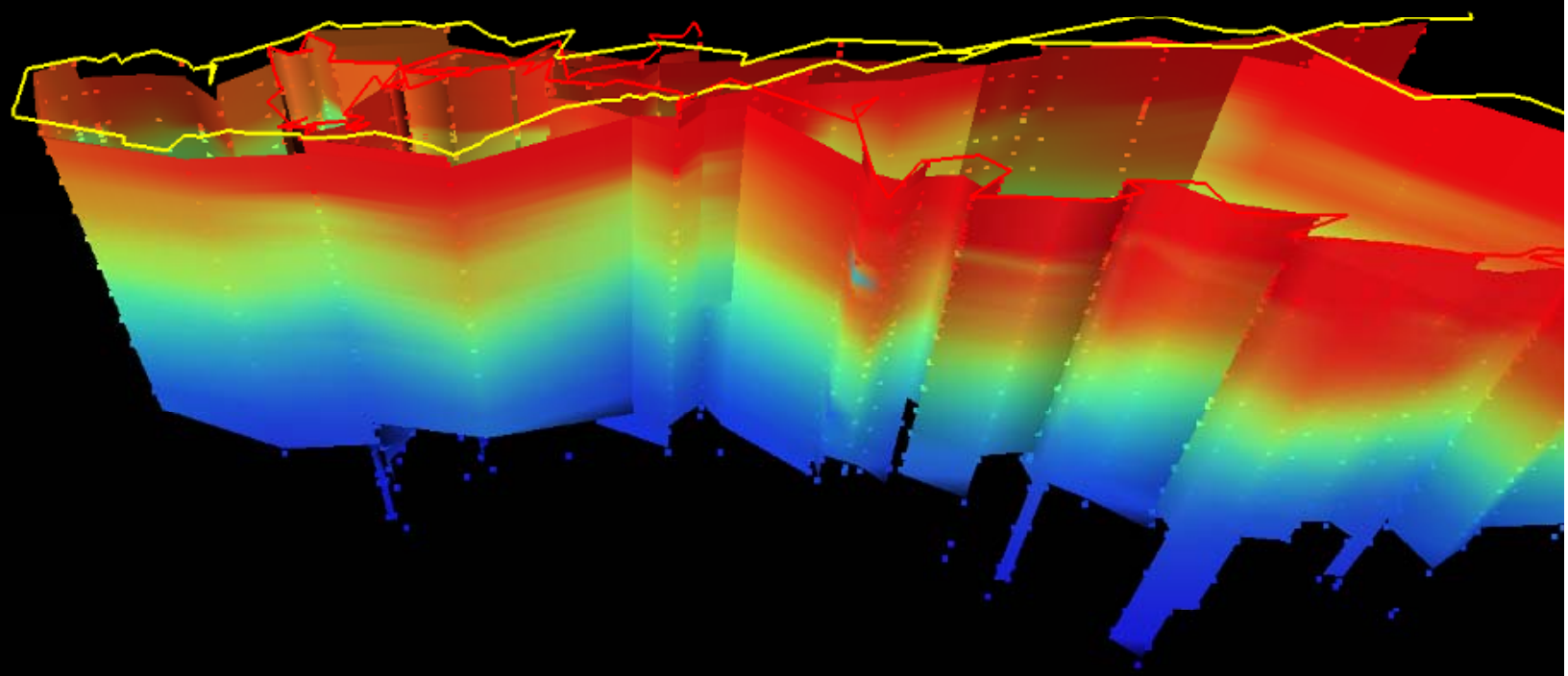
– ***Project 2:*** Partnering with POST and NOAA Fisheries etc+ IOOS RAs to establish a community standards for representation and access for animal acoustic telemetry data.



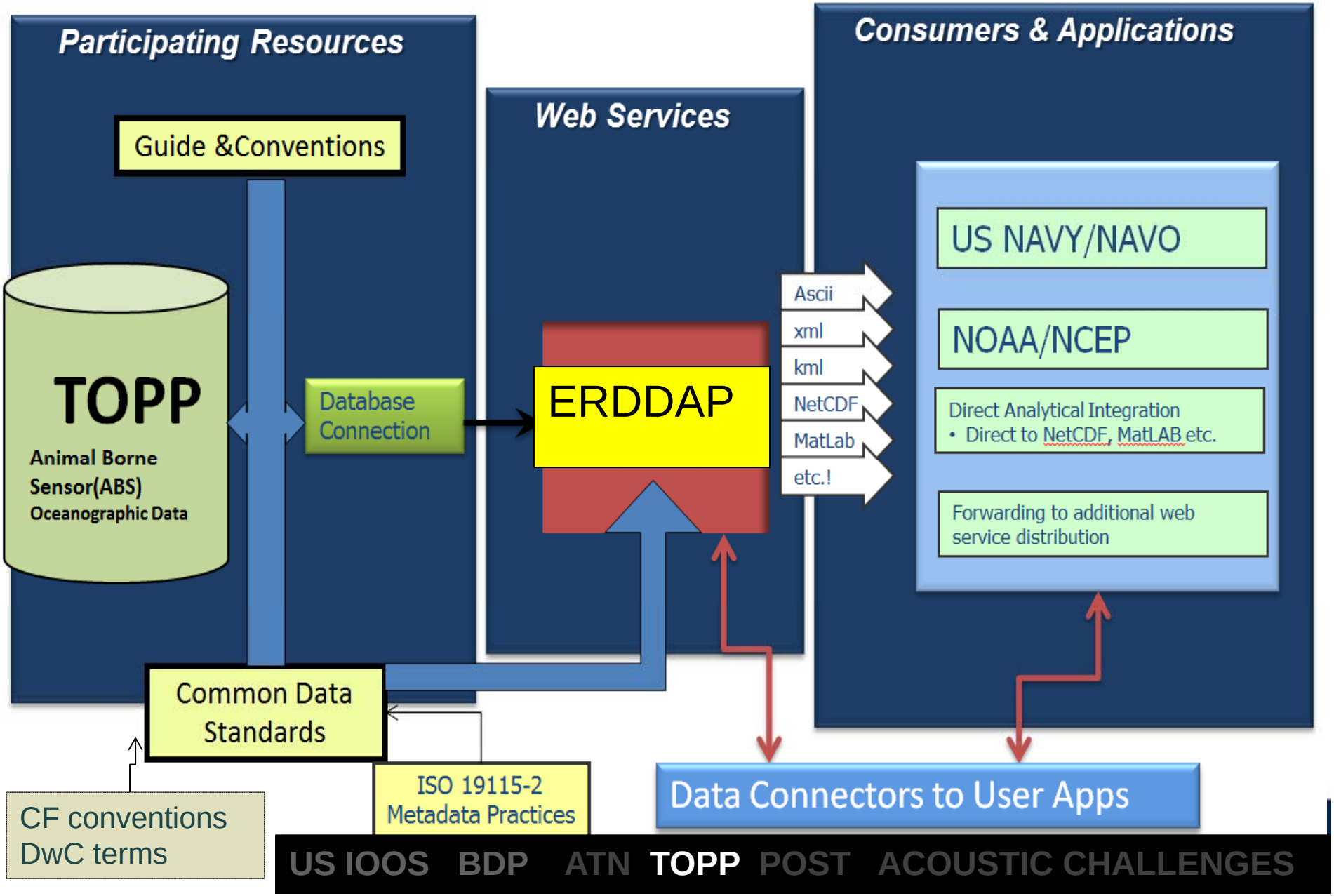
Courtesy of Dan Costa UCSC



Courtesy of Dan Costa, UCSC



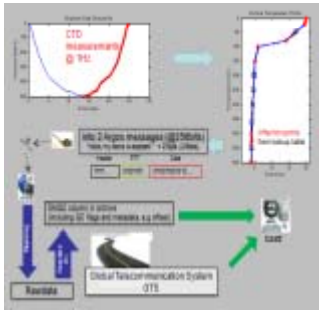
Animal Borne Sensors to Ocean Models



TOPP Real Time Data System

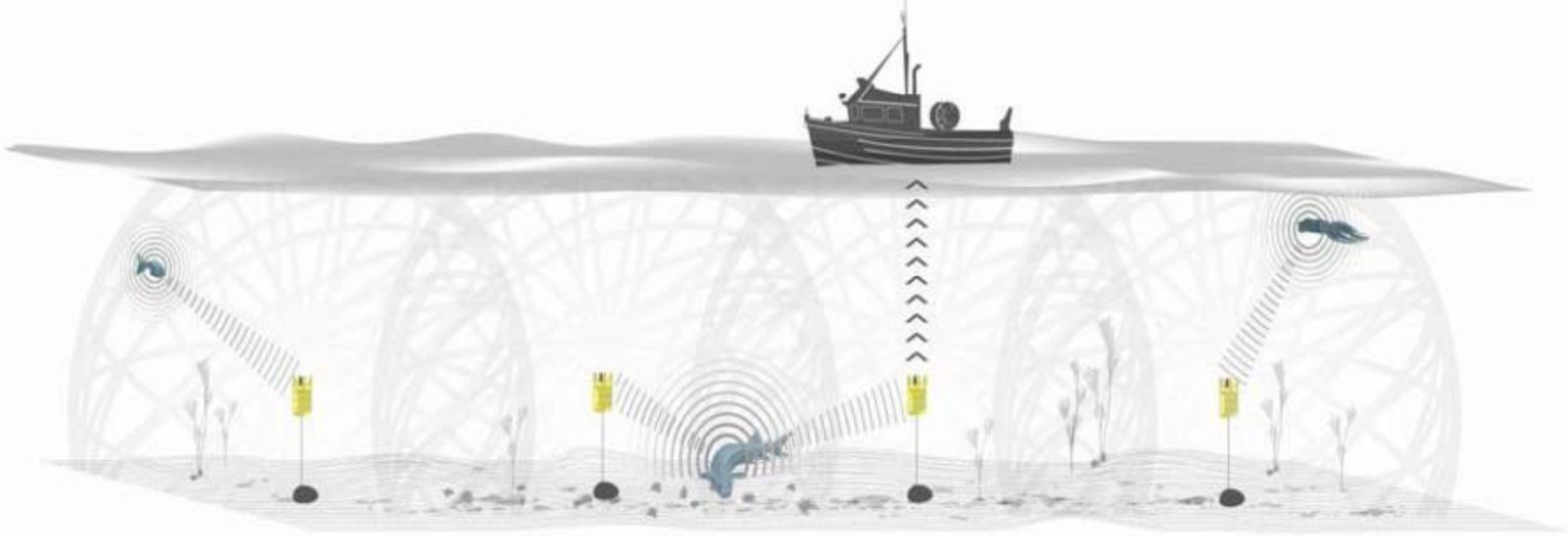
NOAA/NCEP
US NAVY/NAVOCEAN

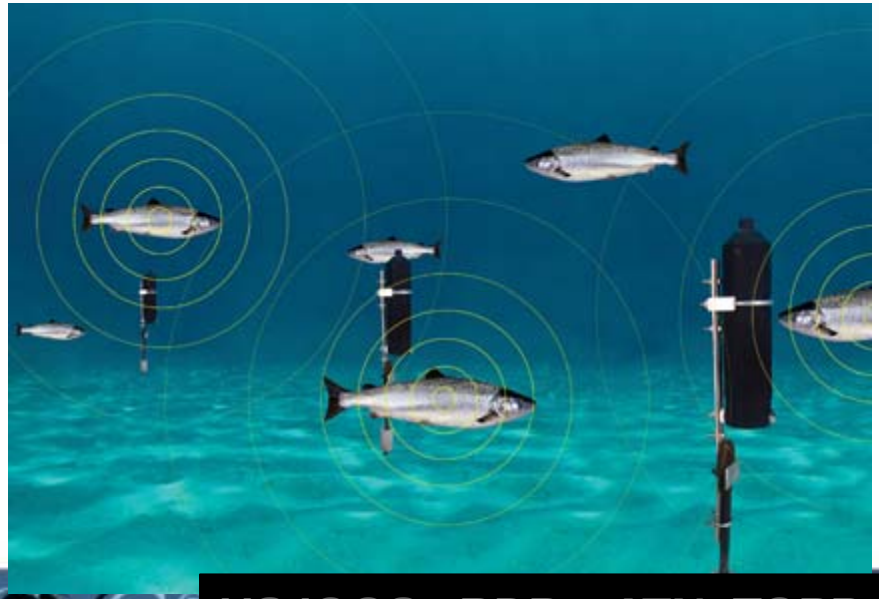
Global Telecommunication System (GTS) WMO Format



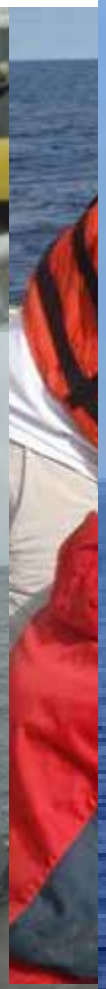
TOPP DataBase

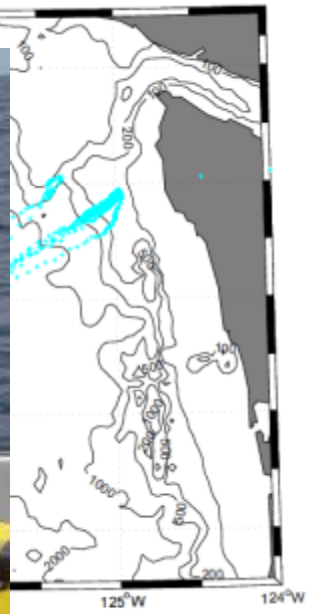
Animal Acoustic Sensors to Fisheries models





Seagliders: for testing, targeted searches, investigating temporary phenomena





Bioprobes

Ecosystem-based Management

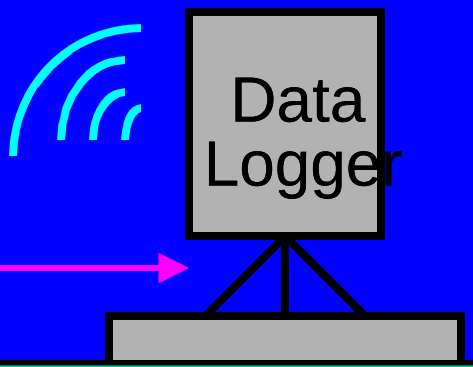
Fully Integrated Tags

- Send/Receive POST codes
- Geolocation
- Physics/Chemistry
- Acoustic download

1) I'm here . . .

3) I've been to . . .
4) It was cold…
5) I met…
6) I ate…

2) Where have you been?

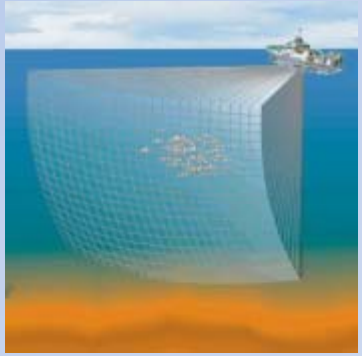


Holland, www.soest.hawaii.edu/PFRP/biology/dagorn_business_tags.html

Active Acoustic Data

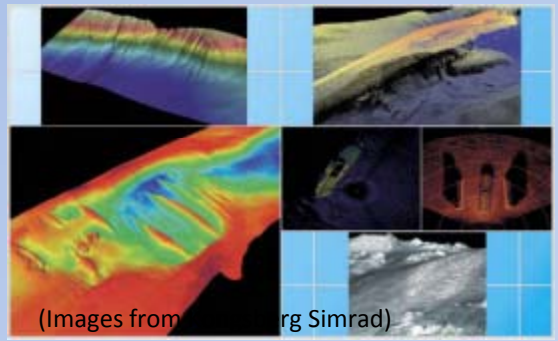
- A single observation can range from cm's to km's in space and tenths to tens of seconds in time

Kongsberg Simrad ME70 Multibeam Sonar
What makes this novel is the capability to simultaneously collect water column and seafloor calibrated backscatter

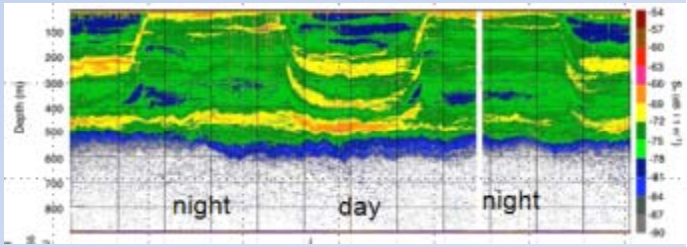


Kongsberg Simrad MS70 Multibeam Sonar

- Multiple observations can range over 1000's of km and months



Multibeam sonar and side scan images of the sea bed and objects on the sea bed

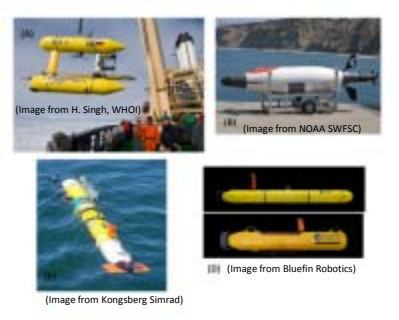


(Horne et al. 2010) 48-hour echogram with time on the x-axis, left to right vs. depth (900 m).

- Acoustic measurements are made from many platforms
  - Vessels, AUVs, ROVs, buoys

NOAA FRV Delaware II

Retractable keel with EK60 transducers on NOAA FSV HB Bigelow



Verification and Biological Sampling:
- Midwater Trawl
- Tow speeds 4 – 5 knots

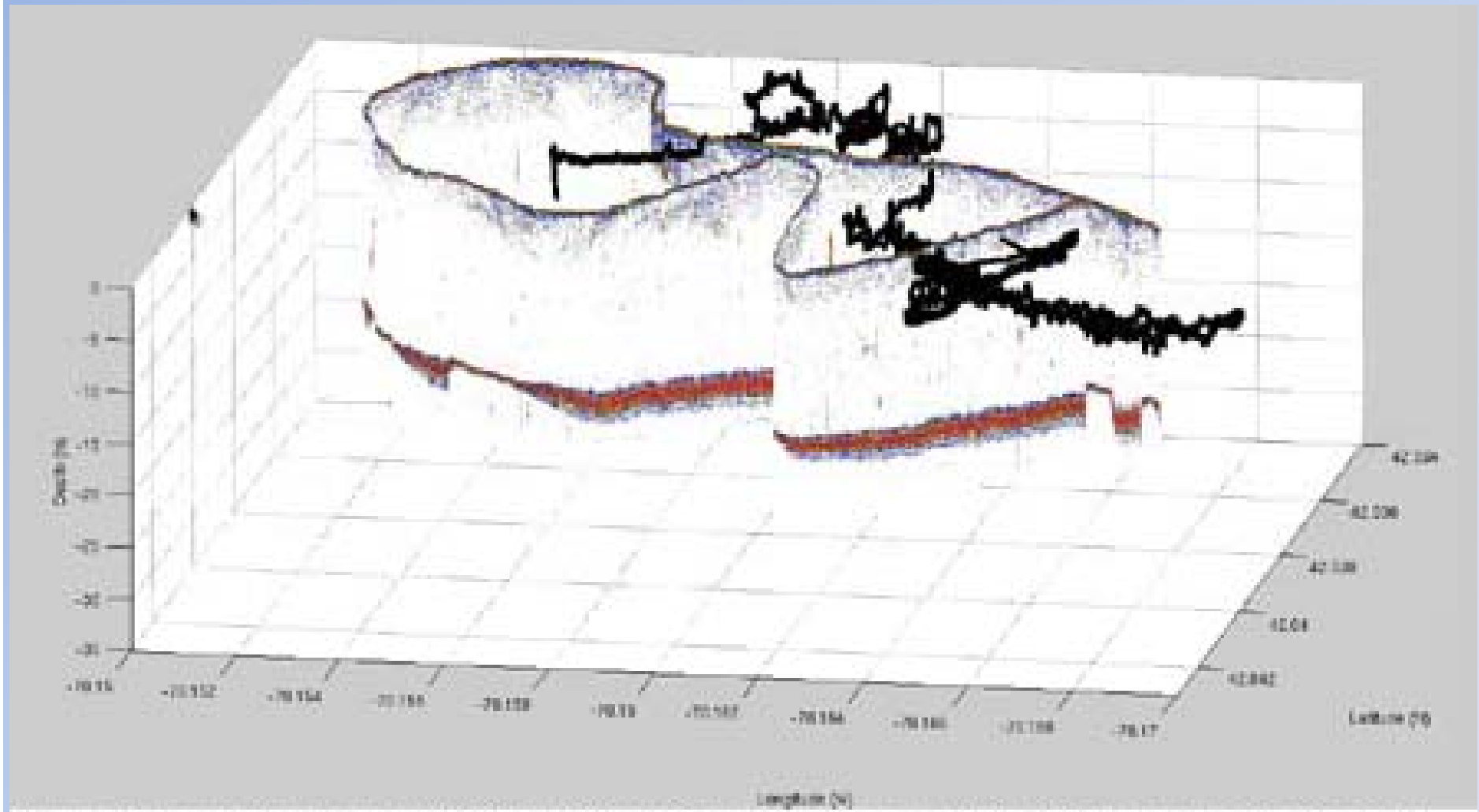
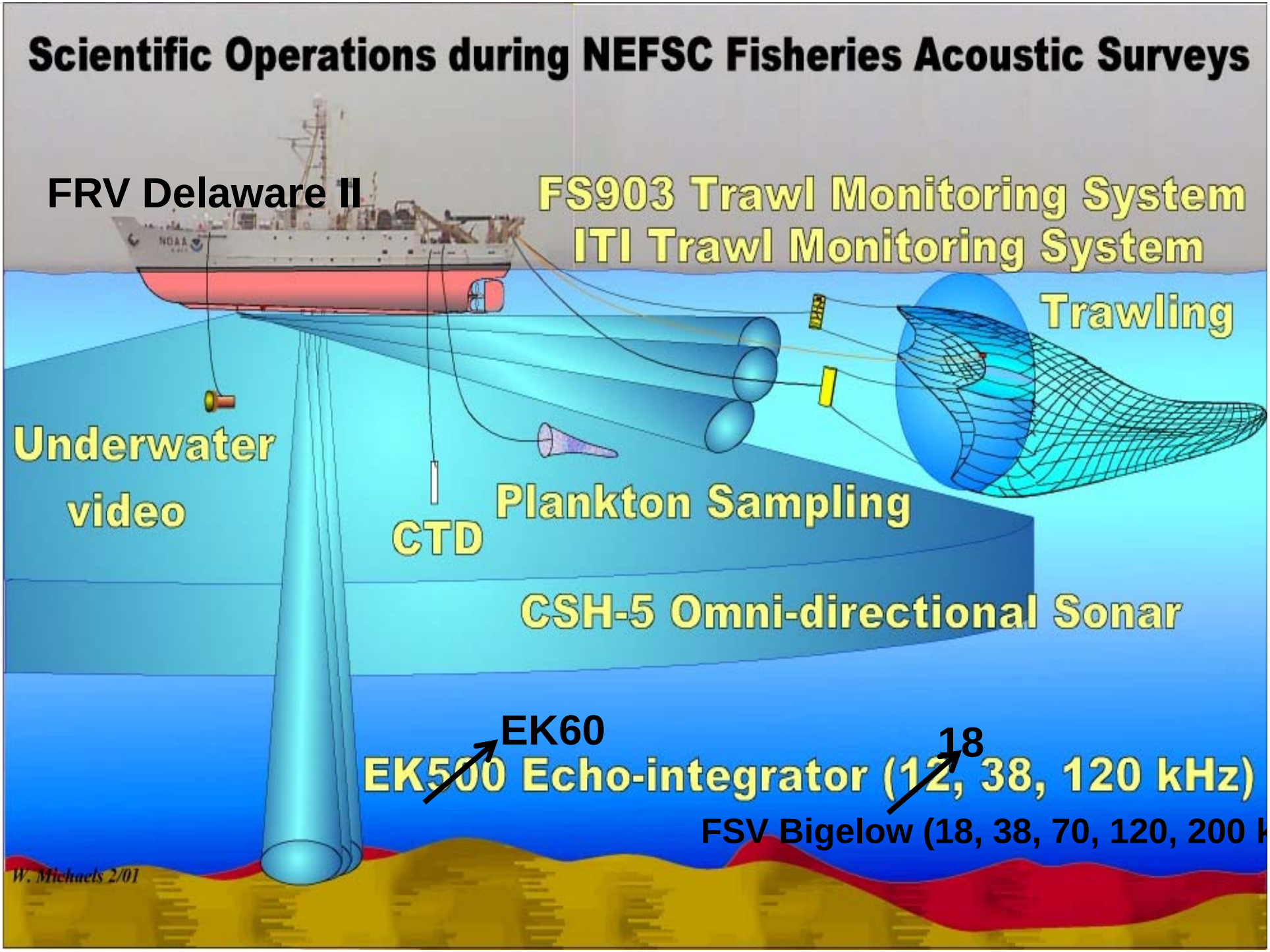


Figure 1. Acoustic echogram (710 kHz) from Cape Cod Bay overlaid with tagged path of a right whale (black line). The right whale spent nearly all its time within the upper 5 m of the water column. Blue areas in the echogram represent patches of calanoid copepods with numerical densities > 1000 animals / m3 while the red line is the seafloor. The track of the tagged right whale is psuedo-georeferenced so the horizontal uncertainty is on the order of 50 m.

Courtesy of Joseph D. Warren
(Stony Brook University)


Scientific Operations during NEFSC Fisheries Acoustic Surveys
FRV Delaware II
FS903 Trawl Monitoring System
ITI Trawl Monitoring System
Trawling
Underwater video
CTD
Plankton Sampling
CSH-5 Omni-directional Sonar
EK60
18
EK500 Echo-integrator (12, 38, 120 kHz)
FSV Bigelow (18, 38, 70, 120, 200 k
W. Michaels 2/01

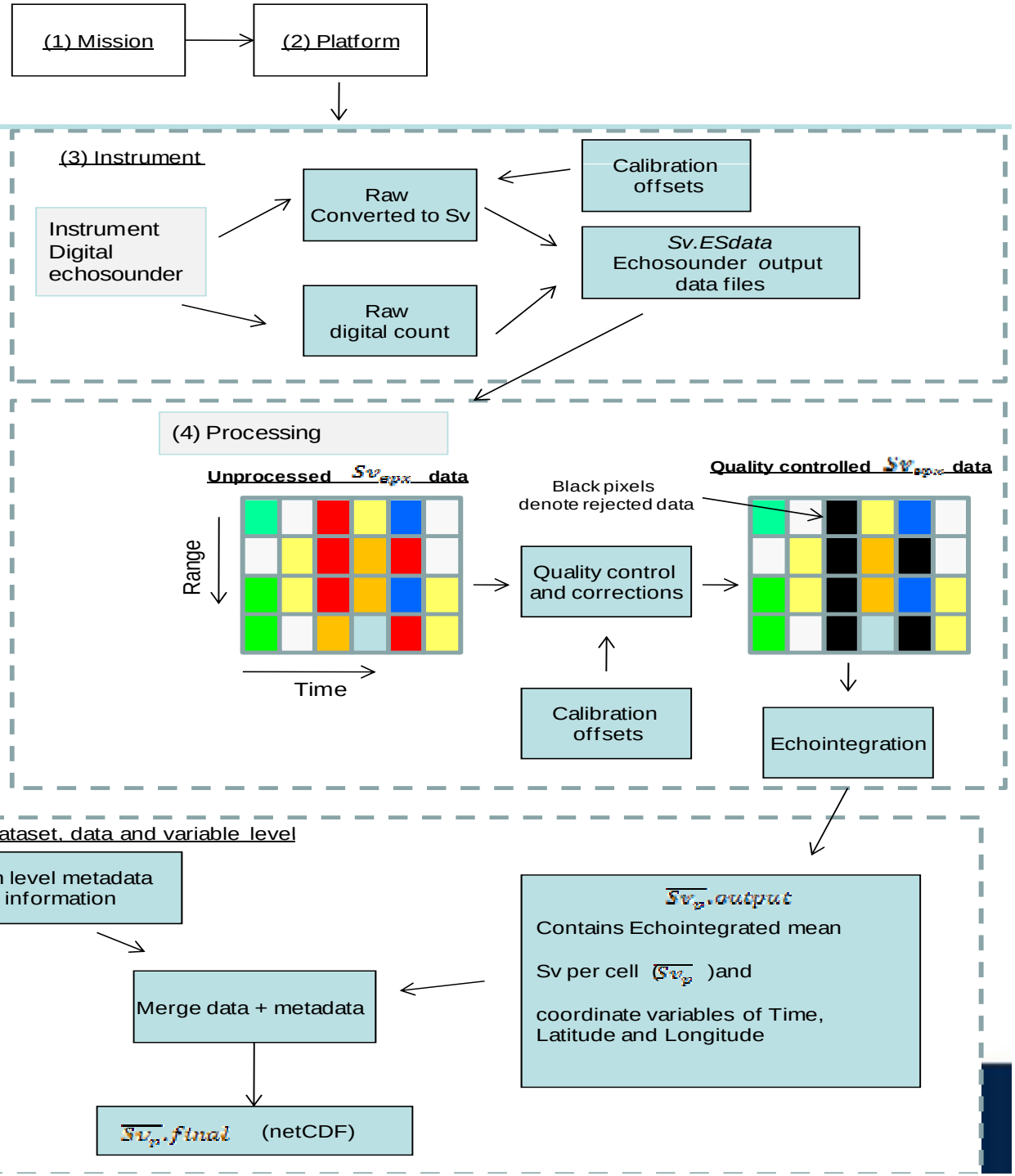
Which data model fit this type of data?

NetCDF (Network Common Data Form) may be a better format for storage of final acoustic output data.

Sv (Acoustic Volume Backscatter) final data file is well suited to array-oriented acoustic data.

Further, metadata contained in the NetCDF files can be automatically harvested to facilitate ready discovery of the data.

Data flow schema for production of mean Sv data with associated metadata levels

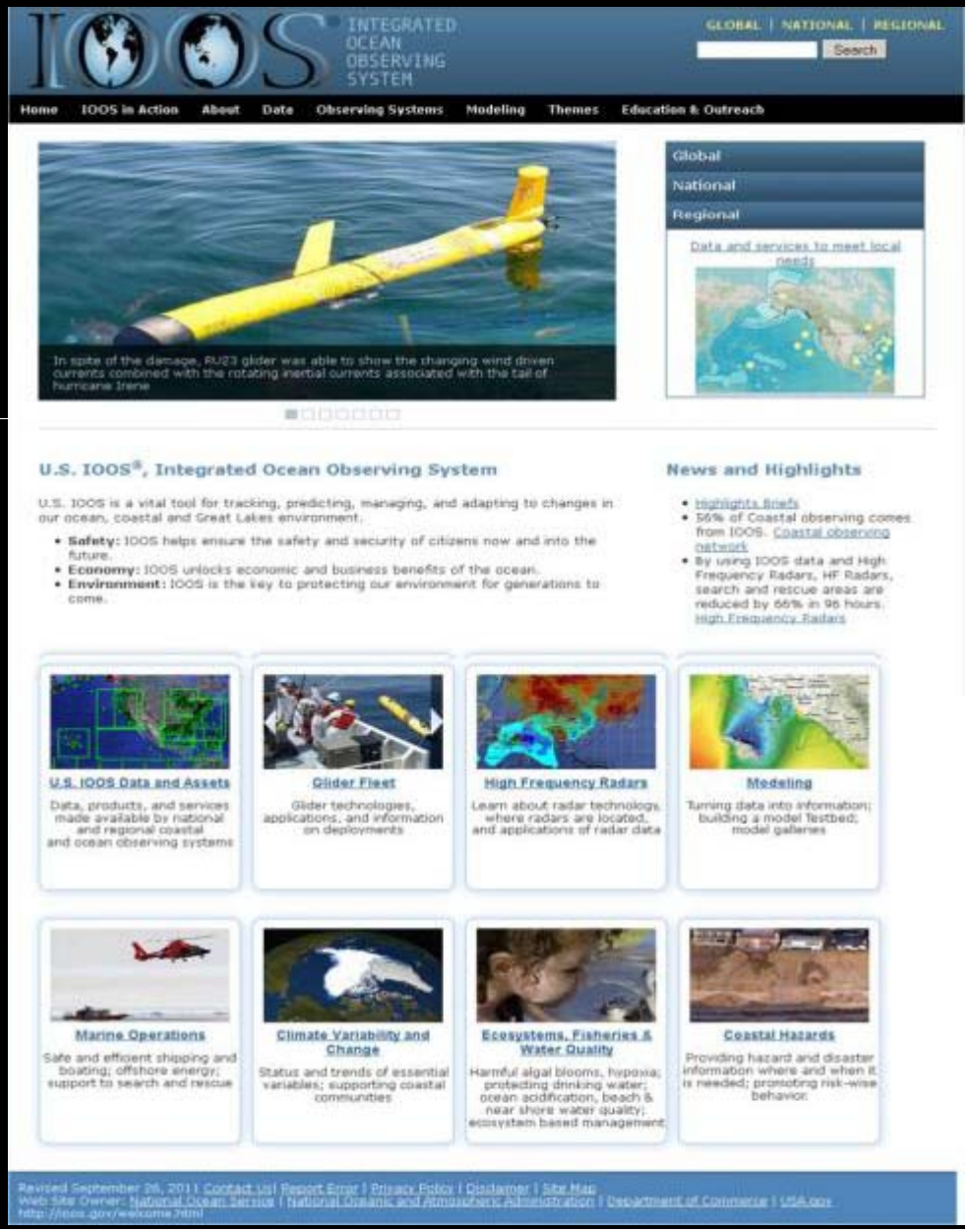


Example. Data variable for volume backscatter and ancillary data quality variable

```
Sv
    Size:      120x2154
    Dimensions: DEPTH,TIME
    Datatype:   double
    Attributes:
        name              = 'Sv'
        long_name         = 'mean_volume_backscatter_coefficient'
        units             = 'm-1'
        _FillValue        = 1e+04
        valid_min         = 0
        valid_max         = 1
        ancillary_variables = 'Sv_quality_control'
          Sv_quality_control
          Size:      120x2154
          Dimensions: DEPTH,TIME
          Datatype:   int8
    Attributes:
        long_name = 'quality flag for mean_volume_backscatter_coefficient'
        quality_control_set        = 1
        quality_control_conventions = 'IOOS standard flags'
        _FillValue                 = 99
        valid_min                  = 0
        valid_max                  = 9
        flag_values                = [0 1 2 3 4 5 6 7 8 9]
          flag_meanings            = 'No_QC_performed, Good_data, Probably_good_data,
          Bad_data_that_are_potentially_correctable, Bad_data Value_changed, Not_used, Not_used,
          Not_used, Missing_value
```

Three Take Home points

- US IOOS is a partnership across the US, that also operates in a global context
- Data Integration critical focus. IOOS BDP is establishing DMAC requirements for Biological data and related.
- Data integration requires engagement by all



Thank You

INTEGRATED OCEAN OBSERVING SYSTEM

Dr. Hassan Moustahfid
Hassan. Moustahfid@noaa.gov
301/427-2447
ioos.gov

IOOS Biological Data Extensibility

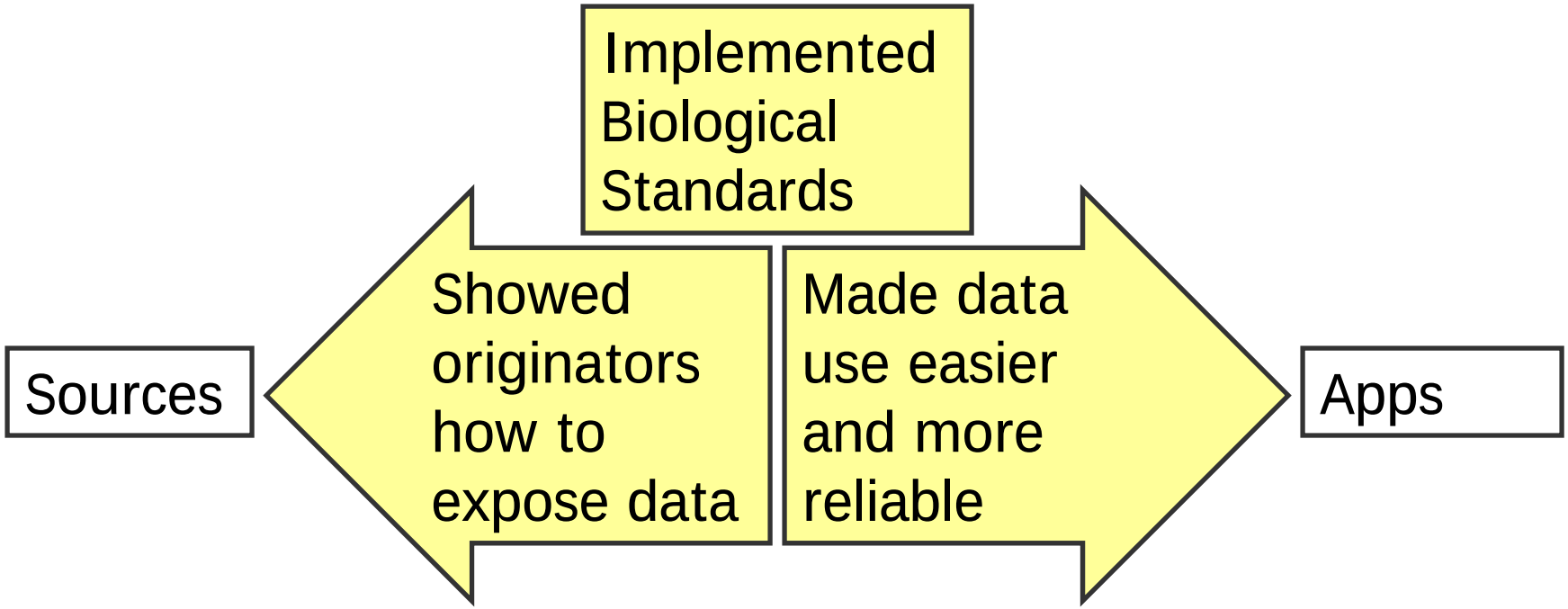
What Did We Solve? and How Did We Solve It?

General, Abstract	IOOS Biological Data	What did we solve? 1. Biogeography 2. Web Service Access 3. Taxonomic Resolution 4. Spatial-temporal resolution 5. Data Documentation, Quality & Richness 6. Interoperability	How did we solve it? 1.Technical Web Service Standards 2.Global Standards for Data and Metadata 3.Scientists' Requirements 4.IOOS-Specific Requirements
More specific but still general (!?)	Quantification	What did we solve? 1. Serve TDS Sampling Details	How did we solve it? • (same as above)
Very specific	Towed-Diver Survey	What did we solve? 1. Serve TDS Sampling Details	How did we solve it? • (same as above)

Credit to Philip Goldstein (OBIS-USA)

IOOS Biological Data Project

What Did We Solve? (Agency View)



Credit to Philip Goldstein (OBIS)

IOOS® Core Variables

1. **Temperature**
2. **Salinity**
3. **Water level**
4. **Currents**
5. **Surface Waves**
6. **Surface Winds**
7. **Ocean color**
8. Dissolved oxygen
9. pH
10. pCO^2
11. Heat flux
12. Bottom character
13. Pathogens
14. Bathymetry
15. Ice distribution
16. Contaminants
17. Stream flow
18. Dissolved nutrients
19. Optical properties
20. Total suspended matter
21. Colored dissolved organic matter
22. Fish species
23. Fish abundance
24. Zooplankton species
25. Phytoplankton species
26. Zooplankton abundance

list could be expanded based on identified needs...

Study A

Metadata (from EML)

Study A: White Mountains
Area column units: sq. meter
PIRU = *Picea rubens*
BEPA = *Betula papyrifera*

Data

Date	Site	Species	Area	Count
10/1/1993	N654	PIRU	2	26
10/3/1994	N654	PIRU	2	29
10/1/1993	N654	BEPA	1	3

Study B

Metadata (from EML)

Study B: Green Mountains
Area sampled: 1 sq. meter
picrub = *Picea rubens*
betpap = *Betula papyrifera*

Data

Date	Site	picrub	betpap
31 Oct 1993	1	13.5	1.6
14 Nov 1994	1	8.4	1.8

In order to combine the results from studies A and B, information contained in the metadata must be used to manipulate the data, since the two tables have different forms (schema). For example, the species columns from studies A and B use different structures and different annotations, but essentially describe the same organisms —a fact that would not be apparent without metadata.

Integrated Data

Study	Date	Site	Species	Density
A	10/1/1993	N654	*Picea rubens*	13.0
A	10/3/1994	N654	*Picea rubens*	14.5
A	10/1/1993	N654	*Betula papyrifera*	3.0
B	10/31/1993	1	*Picea rubens*	13.5
B	10/31/1993	1	*Betula papyrifera*	1.6
B	11/14/1994	1	*Picea rubens*	8.4
B	11/14/1994	1	*Betula papyrifera*	1.8

Metadata 'promoted' to become data

Format normalized using metadata

Species metadata from study B is now data (picrub/betpap column headings)

Density calculated using metadata

Jones et al 2006