



# BCO-DMO: MANAGING LARGE-SCALE COMMUNITY OCEANOGRAPHIC DATA

Cyndy Chandler

Woods Hole Oceanographic Institution

*EarthCube Ocean Omics Workshop*

Santa Catalina Island, CA , USA

21-23 August 2013



# NSF OCE FUNDED RESEARCH DATA



Biological & Chemical Oceanography Data Management Office

[bco-dmo.org](http://bco-dmo.org) for NSF OCE project data

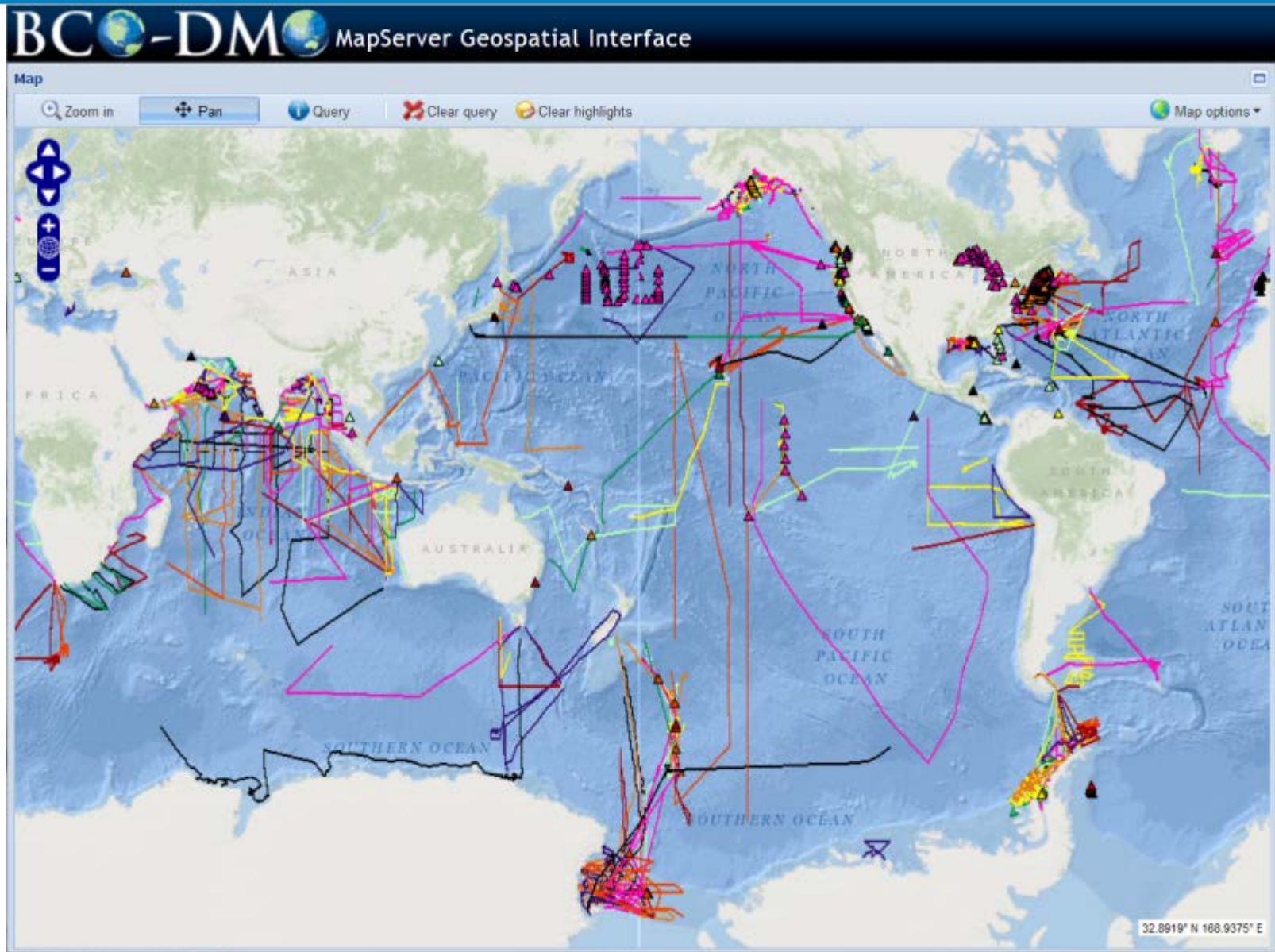
Current projects, and legacy data from large  
coordinated research programs  
(e.g. US GLOBEC and US JGOFS)



[bco-dmo.org](http://bco-dmo.org)



# NSF OCE RESEARCH PROJECT DATA



# CYBERINFRASTRUCTURE

"Cyberinfrastructure is the coordinated aggregate of software, hardware and other technologies, **as well as human expertise**, required to support current and future discoveries in science and engineering."

Definition from: Fran Berman (February 2005). "SBE/CISE Workshop on Cyberinfrastructure for the Social Sciences". San Diego Supercomputer Center, UCSD.

# BCO-DMO

4 project PIs: Peter Wiebe, Robert Groman,  
David Glover and Cyndy Chandler

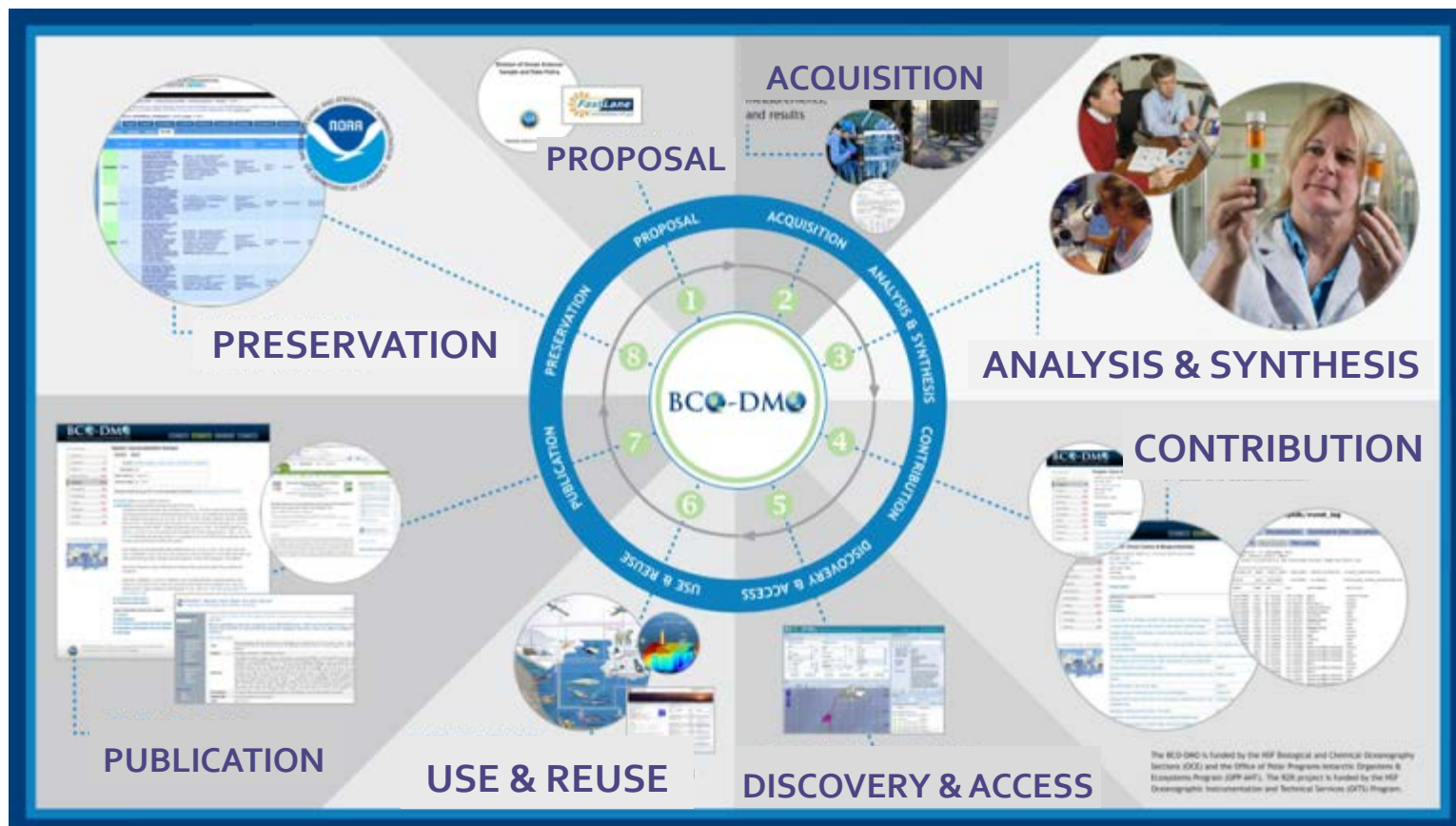
team of data managers / data specialists  
(4 FTE)

one full time programmer  
and additional support personnel as needed



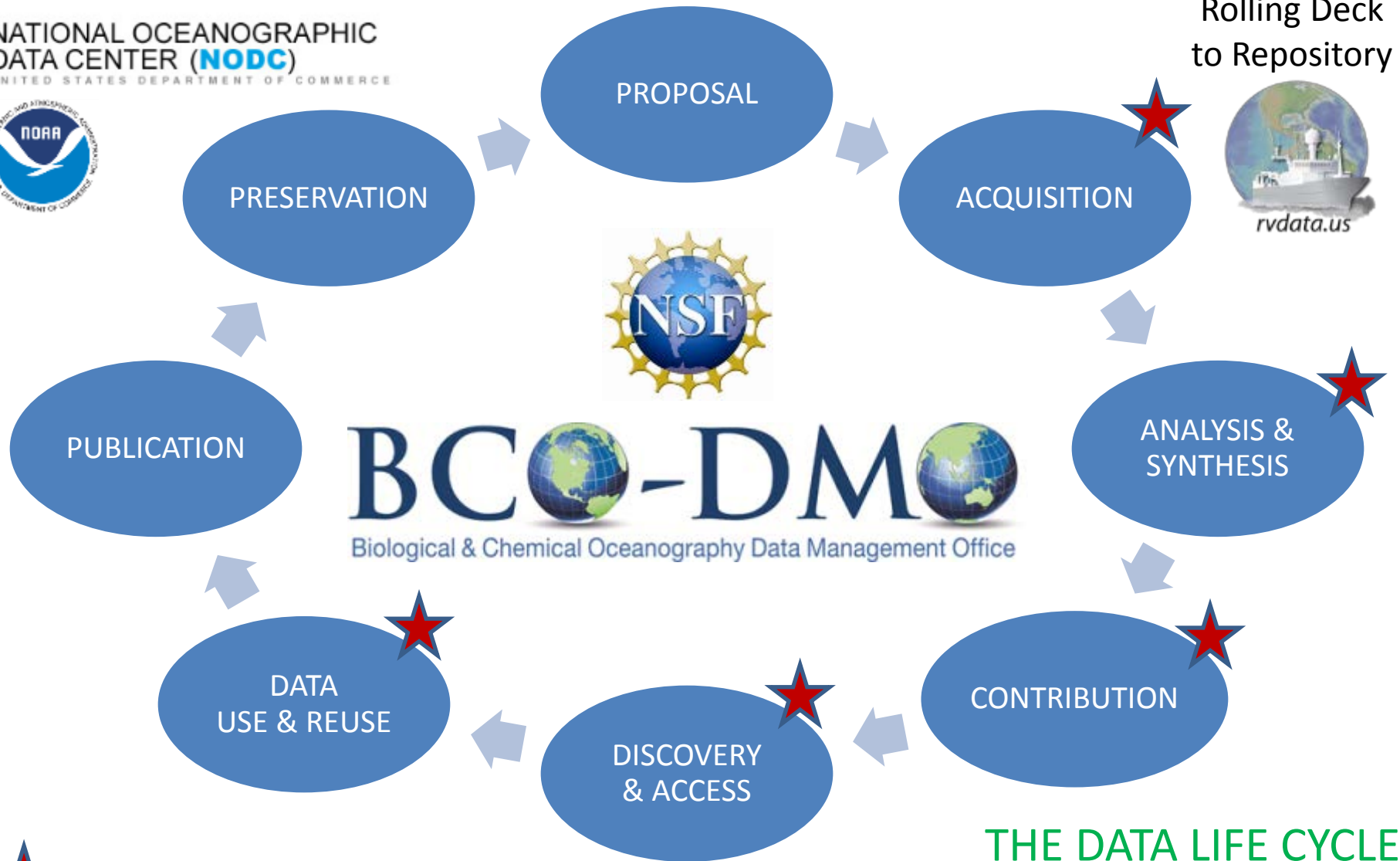
# DATA LIFE CYCLE

Good data management practices account for all aspects of the data cycle, featuring a stewardship philosophy in all phases from “proposal to preservation”.





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



★ Increased requirements for data integration

THE DATA LIFE CYCLE

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



PRESERVATION



PROPOSAL

ACQUISITION

Rolling Deck  
to Repository



GeoMICS

ANALYSIS &  
SYNTHESIS

PUBLICATION

GORDON AND BETTY  
MOORE  
FOUNDATION



CONTRIBUTION



DATA  
USE & REUSE

DISCOVERY  
& ACCESS

GeoMICS DATA LIFE CYCLE



Increased requirements for data integration





# CYBERINFRASTRUCTURE

**Cyberinfrastructure** consists of computing systems, data storage systems, advanced instruments and **data repositories**, visualization environments, **and people**, **all linked together** by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible.

Craig Stewart, Indiana University,  
<http://pti.iu.edu/ci/cyberinfrastructure-news?d=200703>

# EXAMPLE FROM BCO-DMO

## Project:

En-Gen: A Functional Genomics Approach to Organic Sulfur Cycling in the Ocean

**PI:** Mary Ann Moran (University of Georgia)

**Co-PIs:** Ronald Kiene (Dauphin Island Sea Lab)  
William Whitman (University of Georgia)

# LINKS TO DISTRIBUTED DATA

- NSF funded project is registered at BCO-DMO
- Data are available from BCO-DMO, NCBI GenBank and CAMERA



# EXAMPLE DATA FROM BCO-DMO

  
Biological & Chemical Oceanography Data Management Office

HOME | DATA | RESOURCES | ABOUT

**DATABASE**

- Welcome
- Programs 24
- Projects 222
- Deployments 1661
- Datasets 6260**
- Instruments 297
- Parameters 1299
- People 1284

**Dataset: bacterioplankton\_dmdA\_sequences**

Get Data | Map It

<b>Project:</b>	En-Gen: A Functional Genomics Approach to Organic Sulfur Cycling in the Ocean
<b>Principal Investigator:</b>	Mary Ann Moran (University of Georgia)
<b>Co-Principal Investigator:</b>	Ronald Kiene (Dauphin Island Sea Lab) William Whitman (University of Georgia)
<b>BCO-DMO Data Manager:</b>	Shannon Rauch (Woods Hole Oceanographic Institution)
<b>Validated:</b>	yes
<b>Data version:</b>	23 Nov 2012

Dataset Description: Links to published dmdA sequences from marine bacterioplankton from Sapelo Island, GA. Links are provided to published dimethylsulfoniopropionate (DMSP) demethylase protein, DmdA, sequence data from marine bacterioplankton that have been deposited in the NCBI GenBank Short Read Archive (SRA).

## BCO-DMO repository

[BROWSE map](#)
[KEYWORD search](#)
[ADVANCED search](#)
[Start over](#)

### Available programs

Name	#
ASCOS	1
CAMEO	1
CoML	91
Dimensions of Biodiver...	0

### Available projects

Name	#
En-Gen DMSP Cycling	3
En-Gen OA and E huxl...	0
ENAB	0
EqPac	7

### Available deployments

Name
DMSP_Dauphin_Island
DMSP_Sapelo_Island
DMSP_Sargasso_Sea

## Map

[Zoom in](#)
[Pan](#)
[Query](#)
[Clear query](#)
[Clear highlights](#)
[Map options](#)



## Visible deployments

Highlight selected deployments on map?

Select deployment(s) to view dataset(s). Right-click a

<input checked="" type="checkbox"/>	DMSP_Dauphin_Island
<input checked="" type="checkbox"/>	DMSP_Sapelo_Island
<input checked="" type="checkbox"/>	DMSP_Sargasso_Sea

Page 1 of 1 1 - 3 of

## Datasets

[Available datasets](#)
[Mapped datasets](#)

Group by: deployment [Remove](#)

Select dataset(s) to add to your map. Right-click a row

Dataset	Deployment
---------	------------

### Deployment: DMSP\_Dauphin\_Island

<input checked="" type="checkbox"/>	bacterial_production	DMSP_Dauphin_Isl...
<input checked="" type="checkbox"/>	chlorophyll_a	DMSP_Dauphin_Isl...
<input checked="" type="checkbox"/>	DICE_sequences	DMSP_Dauphin_Isl...
<input checked="" type="checkbox"/>	fluor_PAM	DMSP_Dauphin_Isl...
<input checked="" type="checkbox"/>	microcosm_summary	DMSP_Dauphin_Isl...





## BCO-DMO repository

[BROWSE map](#) [KEYWORD search](#) [ADVANCED search](#) [Start over](#) Quick find: enter a deployment name

### Available programs

Name	#
ASCOS	1
CAMEO	1
CoML	91
Dimensions of Biodiver...	0

### Available projects

Name	#
En-Gen DMSP Cycling	3
En-Gen OA and E huxl...	0
ENAB	0
EqPac	7

### Available deployments

Name
DMSP_Dauphin_Island
DMSP_Sapelo_Island
DMSP_Sargasso_Sea

## Visible deployments

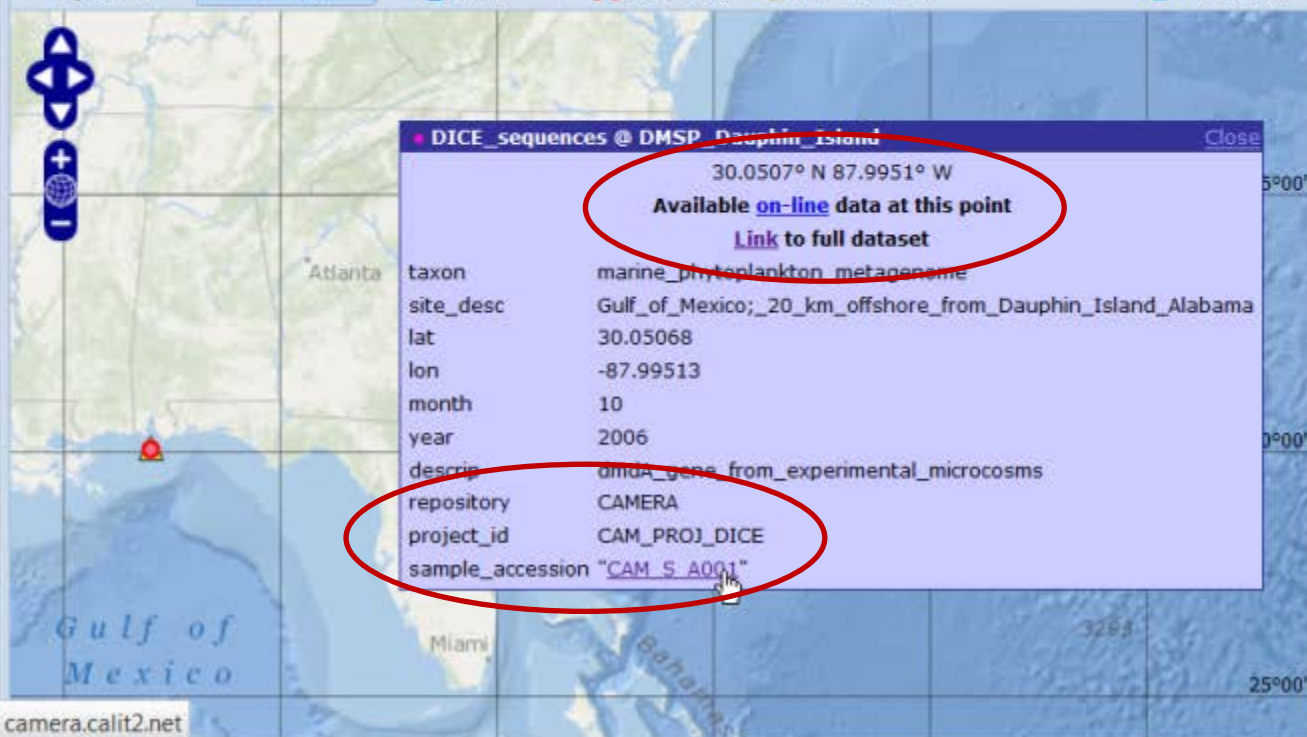
Highlight selected deployments on map? ☒

Select deployment(s) to view dataset(s). Right-click a

- DMSP\_Dauphin\_Island
- DMSP\_Sapelo\_Island
- DMSP\_Sargasso\_Sea

## Map

Zoom in Pan Query Clear query Clear highlights Map options



## Datasets

[Available datasets](#) [Mapped datasets](#)

Group by: deployment Remove all

Select dataset(s) to add to your map. Right-click a row

Dataset	Deployment
Deployment: DMSP_Dauphin_Island	
bacterial_production	DMSP_DauphinIsl...
chlorophyll_a	DMSP_DauphinIsl...
DICE_sequences	DMSP_DauphinIsl...
fluor_PAM	DMSP_DauphinIsl...
microcosm_summary	DMSP_DauphinIsl...

# DATASET WITH LINKS TO ...

## Links to accession numbers at CAMERA & GenBank

Screenshot of a web browser showing the BCO-DMO/En-Gen\_DMSP\_Cycling/DICE\_sequences page, Level 1.

Browser tabs: Links to published ... x, Biological and Ch... x, /BCO-DMO/En-Ge... x, /BCO-DMO/En-Ge... x, CAMERA - Comm... x, /BCO-DMO/En-Ge... x

Address bar: dmoserv3.whoi.edu/jg/serv/BCO-DMO/En-Gen\_DMSP\_Cycling/DICE\_sequences.html9{dir=dmos

Navigation buttons: Directory, Documentation, Download & Other Operations

Level 0, Next Level, Flat Listing

# Links to published DICE (Dauphin Island Cubitainer Experiment) dmdA, cDNA, and RNA sequences  
# PI: Mary Ann Moran (UGA); Co-PIs: Ronald Kiene (DISL) & William Whitman (UGA)  
# Notes: sample\_accession links open in new window/tab.  
# Version: 23 Nov 2012

taxon	site_desc	lat	lon	month	year
marine_phytoplankton_metagenome	Gulf_of_Mexico;_20_km_offshore_from_Dauphin_Island_Alabama	30.05068	-87.99513	10	2006

descrip	repository	project_id	sample_accession
dmdA_gene_from_experimental_microcosms	CAMERA	CAM_PROJ_DICE	<a href="#">CAM S A001</a>
dmdA_gene_from_control_microcosms	CAMERA	CAM_PROJ_DICE	<a href="#">CAM S A002</a>
cDNA_from_control_microcosms	CAMERA	CAM_PROJ_DICE	<a href="#">DICE SMPL DICE C1</a>
cDNA_from_control_microcosms	CAMERA	CAM_PROJ_DICE	<a href="#">DICE SMPL DICE C2</a>
cDNA_from_experimental_microcosms	CAMERA	CAM_PROJ_DICE	<a href="#">DICE SMPL DICE E1</a>
cDNA_from_experimental_microcosms	CAMERA	CAM_PROJ_DICE	<a href="#">DICE SMPL DICE E2</a>
RNA_probes_from_microarray	NCBI_Gene_Expression_Omnibus	nd	<a href="#">GSE21325</a>

# HETEROGENEOUS DATA

- Data types: biological, chemical, geological, physical, biogeochemistry, (\*omics and environmental)
- Study types: field studies, mesocosm, laboratory experiments, model results
- Research themes: OA, biodiversity, ecosystem status/health

**CHALLENGE**



# EXAMPLE: MOCNESS DATA

CTD sensor data



MOCNESS Sampling

Date: 4/28/04  
 Filename (raw): MOC04.raw File name (processed): MOC04.raw

**Net Information**  
 Net Size: 1 m Net Condition: good  
 Net Mesh: 5.53

**Start** **End**  
 Location (Lat/Lon): 47°55'26.64" N 123°03'00" W Location (Lat/Lon): 47°55'04.56" N 123°03'00" W  
 Time (Local): 0035 Time (GMT): 0135 Time (Local): 0210 Time (GMT): 0642  
 MOCNESS Battery Voltage: 19.2 MOCNESS Battery Voltage:

**Environmental Parameters**  
 Wind (Speed/Direction): 8.7, 30° Sea State: very low, 1.0 m  
 Water Depth (m): 6.5 Light and Clouds: 60% of sky, 100%  
 Air Temperature: -1.6°C SST: -1.15°C

Net #	Depth Interval	Depth Open	Time Open	Depth Close	Time Close	Volume Filtered	Comments
0	0-200	0	0135	200	0521	846.45	NR NR
1	200-400	200	0501	595	0531	401	NR NR 10 m/min
2	400-600	400	0531	595	0546	755	NR NR 10 m/min
3	600-800	600	0546	195	0603	435	NR NR 10 m/min
4	800-1000	800	0603	99	0626	1238	NR NR 10 m/min
5	1000-1200	1000	0626	74	0630	293	NR NR 5 m/min
6	1200-1400	1200	0630	47	0636	32	NR NR 5 m/min
7	1400-1600	1400	0636	266	0642	978	NR NR 5 m/min
8	1600-1800	1600	0642	0	0642	574	NR NR 5 m/min

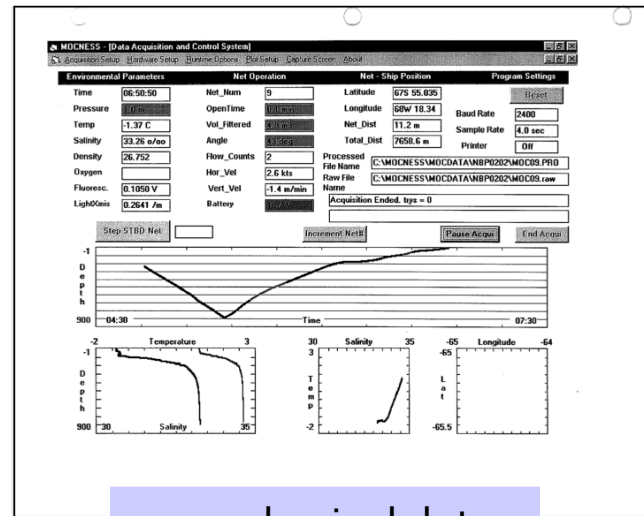
Reset system in mid 600 m net of web to release file box volume followed  
 State m. 600 m net

MOCNESS TOW #: 1  
 Date: 4/28/04  
 Region: LaBoca  
 People Processing Sample: CDR, AK

**NET DEPTH** **Taxonomic Composition**

0	0-200	Copepods, krill, large fish, jellyfish, large krill
1	200-400	Euphausiids (present), copepods, 1 red medusa
2	400-600	Krill, euphausiids, copepods, 1 medusa
3	600-800	large euphausiids, euphausiids, copepods, large krill
4	800-1000	large euphausiids, large krill, 1 large euphausiid, 1 large jellyfish, 1 small krill
5	1000-1200	Krill, euphausiids, copepods, euphausiids, large krill
6	1200-1400	3rd Yell (Am), white euphausiids, 1 euphausiid, 1 small krill
7	1400-1600	large euphausiids, 1 euphausiid, 1 small krill
8	1600-1800	large euphausiids, 1 euphausiid, 1 small krill

raw biology data



raw physical data

... the samples are processed by people, observations recorded by people, and digital data sets created by people

/globec/soglobec/zooabund\_nbp --cruiseid eq NBP0202, station eq 34, net eq 4-- Level 3

Directory... Documentation... Plotting and Other Operations...

Level 0 Next level Flat listing

Net # 0 1 2 3 4 5 6 7 8

Depth Interval 0-200 200-400 400-600 600-800 800-1000 1000-1200 1200-1400 1400-1600 1600-1800

Depth Open 0 200 400 600 800 1000 1200 1400 1600

Time Open 0135 0501 0531 0546 0603 0626 0630 0636 0642

Depth Close 200 595 595 195 99 47 266 0

Time Close 0521 0531 0546 0603 0626 0630 0636 0642

Volume Filtered 846.45 401 755 435 1238 293 32 978 574

Comments NR NR NR NR NR NR NR NR NR

Net # 0 1 2 3 4 5 6 7 8

Depth Interval 0-200 200-400 400-600 600-800 800-1000 1000-1200 1200-1400 1400-1600 1600-1800

Depth Open 0 200 400 600 800 1000 1200 1400 1600

Time Open 0135 0501 0531 0546 0603 0626 0630 0636 0642

Depth Close 200 595 595 195 99 47 266 0

Time Close 0521 0531 0546 0603 0626 0630 0636 0642

Volume Filtered 846.45 401 755 435 1238 293 32 978 574

Comments NR NR NR NR NR NR NR NR NR

digital biology data

Directory... Documentation... Plotting and Other Operations...

Level 0 Next level Flat listing

Net # 0 1 2 3 4 5 6 7 8

Depth Interval 0-200 200-400 400-600 600-800 800-1000 1000-1200 1200-1400 1400-1600 1600-1800

Depth Open 0 200 400 600 800 1000 1200 1400 1600

Time Open 0135 0501 0531 0546 0603 0626 0630 0636 0642

Depth Close 200 595 595 195 99 47 266 0

Time Close 0521 0531 0546 0603 0626 0630 0636 0642

Volume Filtered 846.45 401 755 435 1238 293 32 978 574

Comments NR NR NR NR NR NR NR NR NR

Net # 0 1 2 3 4 5 6 7 8

Depth Interval 0-200 200-400 400-600 600-800 800-1000 1000-1200 1200-1400 1400-1600 1600-1800

Depth Open 0 200 400 600 800 1000 1200 1400 1600

Time Open 0135 0501 0531 0546 0603 0626 0630 0636 0642

Depth Close 200 595 595 195 99 47 266 0

Time Close 0521 0531 0546 0603 0626 0630 0636 0642

Volume Filtered 846.45 401 755 435 1238 293 32 978 574

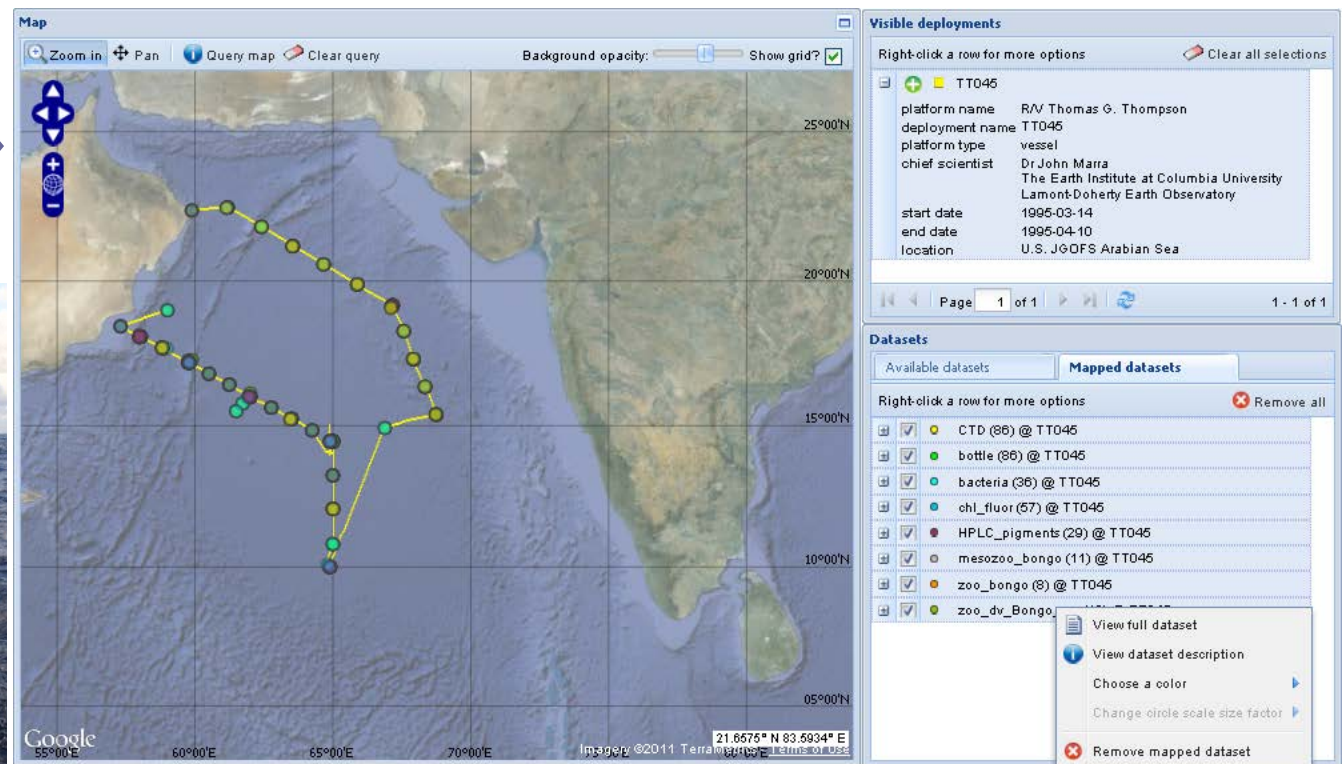
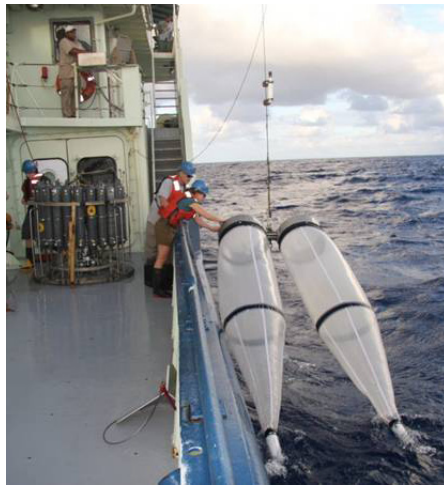
Comments NR NR NR NR NR NR NR NR NR

digital physical data

# FIELD DATA TO DATABASE

*in situ* data from research cruises are documented and contributed to the online data system and discoverable through a variety of user interfaces

Original data from  
Bongo net tows  
and CTD/Niskin  
Rosette





# DATA REPOSITORY CONNECTIONS

BCO-DMO is flexible enough to accommodate the wide variety of ocean data from NSF OCE funded research.

BCO-DMO catalog of data has been extended through connections with other data repositories (e.g. OBIS, GenBank, CAMERA, LTER).

**SOLUTION**

# DATA ACCESS SEMANTICS

Faceted search >>>



BC-DM MapServer Geospatial Interface (Beta)

Search

Return to classic map Start over

Programs

- Arctic Summer Cloud Ocean Study (2)
- Census of Marine Life (467)
- Comparative Analysis of Marine Ecosystem Organization (1)
- Emerging Topics in Biogeochemical Cycles (14)
- Gulf of Mexico - Deepwater Horizon Oil Spill (56)

Projects

- Algal Communities in Distress: Impacts and Consequences (3)
- Amazon influence on the Atlantic: CarbOn export from Nitrogen fixation by DiAtom Symbioses (12)
- Anammox, denitrification and nitrogen fixation in the Black Sea (3)
- Aplysina Red Band Syndrome: Investigating the etiology,

People

- Abbott, Mark (4)
- Abraham, Edward (12)

Datasets

Available datasets Mapped datasets

Group by: dataset

Select dataset(s) to add to your project. Click on the plus icon for more options.

Dataset	Deployment
Dataset: 13C samples	
13C samples	KM0715
Dataset: 15N2 samples	
15N2 samples	KM0715
Dataset: 16S rRNA clone library	
16S rRNA clone library	KM102.05

Page 1 of 685

Visible deployments

- AB\_63\_1
- AB\_63\_2
- AB\_63\_3
- AB\_63\_4A
- AB\_63\_A
- AB\_64\_5
- AB\_64\_6
- AB\_64\_7
- AB\_64\_8
- ACIDIC-CM...
- ACIDIC-FC...
- ACIDIC-SH...
- AE0811

Map

Clear selections

Clear query Clear highlights

Map options

# THANK YOU

## Questions . . .



*Woods Hole, Massachusetts, USA*