

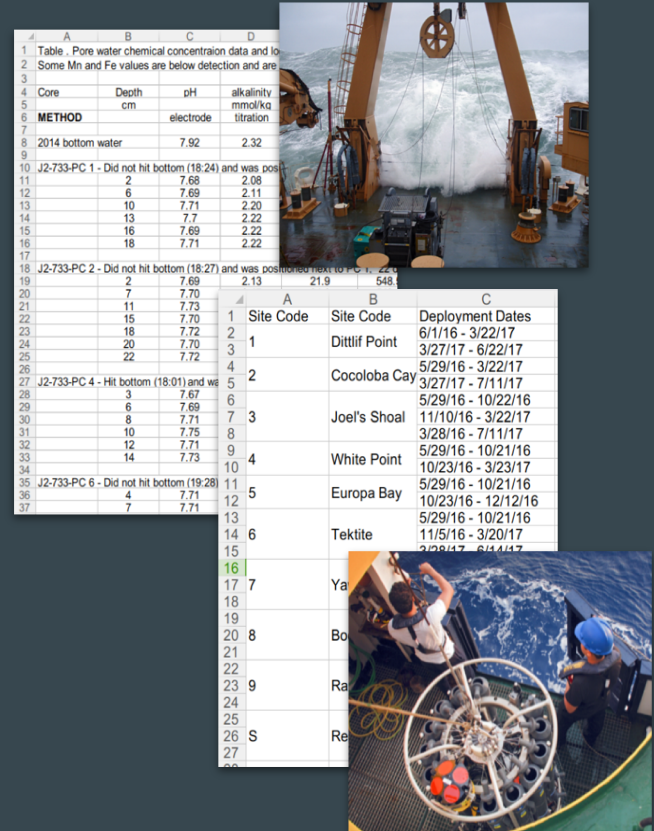
Data Management and Reporting: BCO-DMO Data Management Services & Best Practices

Shannon Rauch



Outline

1. Intro/overview of BCO-DMO
 - a. History/mission, who we are
 - b. Types of data we handle
 - c. Data lifecycle
 - i. Proposal:
 1. OCE Data Policy requirements
 2. DMPTool template
 - ii. Contribution
 1. How to submit data files & metadata
 - iii. Discovery & Access
 - iv. Data Publication, Reuse
 - v. Preservation
2. Tips for Successful Data Management
 - a. Cruise planning & reporting
 - b. Event logger (R2R) & sample logs
 - c. IGSN & WoRMs identifiers
 - d. Data formatting tips/best practices



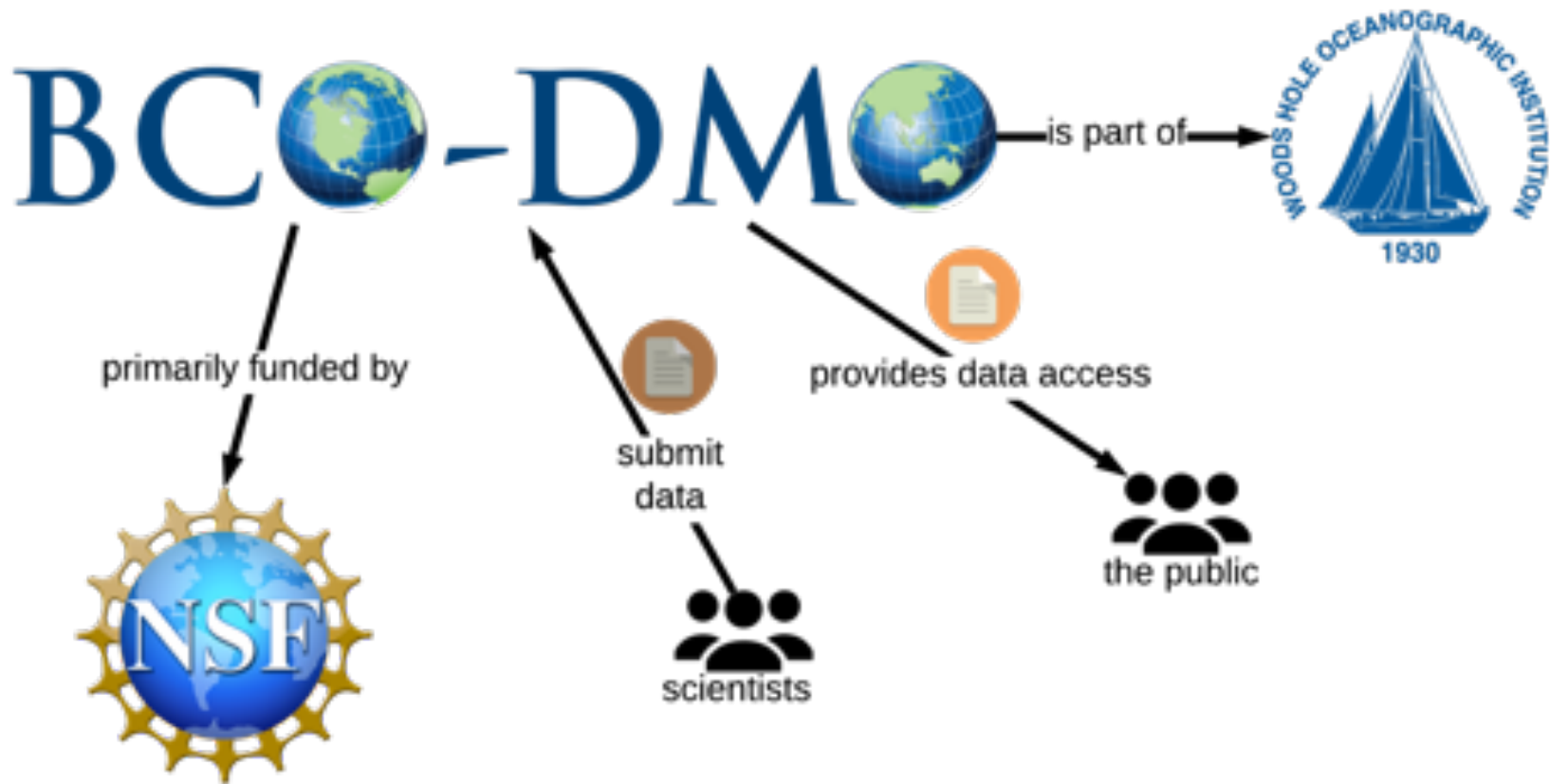
	A	B	C	D
1	Table. Pore water chemical concentration data and lo			
2	Some Mn and Fe values are below detection and are			
3				
4	Core	Depth	pH	alkalinity
5		cm		mmol/kg
6	METHOD		electrode	titration.
7				
8	2014 bottom water		7.92	2.32
9				
10	J2-733-PC 1 - Did not hit bottom (18.24) and was pos			
11		2	7.69	2.08
12		6	7.69	2.11
13		10	7.71	2.20
14		13	7.7	2.22
15		16	7.69	2.22
16		18	7.71	2.22
17				
18	J2-733-PC 2 - Did not hit bottom (18.27) and was positioned next to PC 1, 22			
19		2	7.69	2.13
20		7	7.70	2.19
21		11	7.73	548
22		15	7.70	
23		18	7.72	
24		20	7.70	
25		22	7.72	
26				
27	J2-733-PC 4 - Hit bottom (18.01) and was			
28		3	7.67	
29		6	7.69	
30		8	7.71	
31		10	7.75	
32		12	7.71	
33		14	7.73	
34				
35	J2-733-PC 6 - Did not hit bottom (19.28)			
36		4	7.71	
37		7	7.71	

	A	B	C
1	Site Code	Site Code	Deployment Dates
2			
3	1	Dittlif Point	6/1/16 - 3/22/17
4			3/27/17 - 6/22/17
5	2	Cocoloba Cay	5/29/16 - 3/22/17
6			3/27/17 - 7/11/17
7	3	Joel's Shoal	5/29/16 - 10/22/16
8			11/10/16 - 3/22/17
9			3/28/16 - 7/11/17
10	4	White Point	5/29/16 - 10/21/16
11			10/23/16 - 3/23/17
12	5	Europa Bay	5/29/16 - 10/21/16
13			10/23/16 - 12/12/16
14	6	Tektitle	5/29/16 - 10/21/16
15			11/5/16 - 3/20/17
16			2/28/17 - 6/11/17
17	7	Ya	
18			
19			
20	8	Bo	
21			
22			
23	9	Ra	
24			
25			
26	S	Re	
27			

BCO-DMO: The Biological & Chemical Oceanography Data Management Office

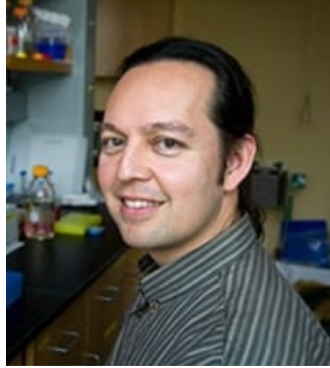
- Established in 2006, through the merging of the two data management offices for the U.S. GLOBEC and JGOFS projects.
- BCO-DMO is funded by the U.S. National Science Foundation (NSF) to provide data management services to NSF-funded projects* at no cost to the investigators.
- We now manage data from thousands of projects from researchers (PIs) across the U.S. studying a wide variety of scientific domains.

Who are we?



BCO-DMO: Biological and Chemical Oceanography Data Management Office

Who are we?



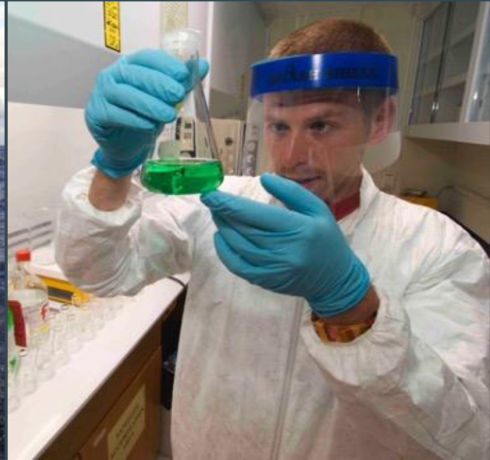
Peter Wiebe, Mak Saito, Amber York, Karen Soenen, Nancy Copley



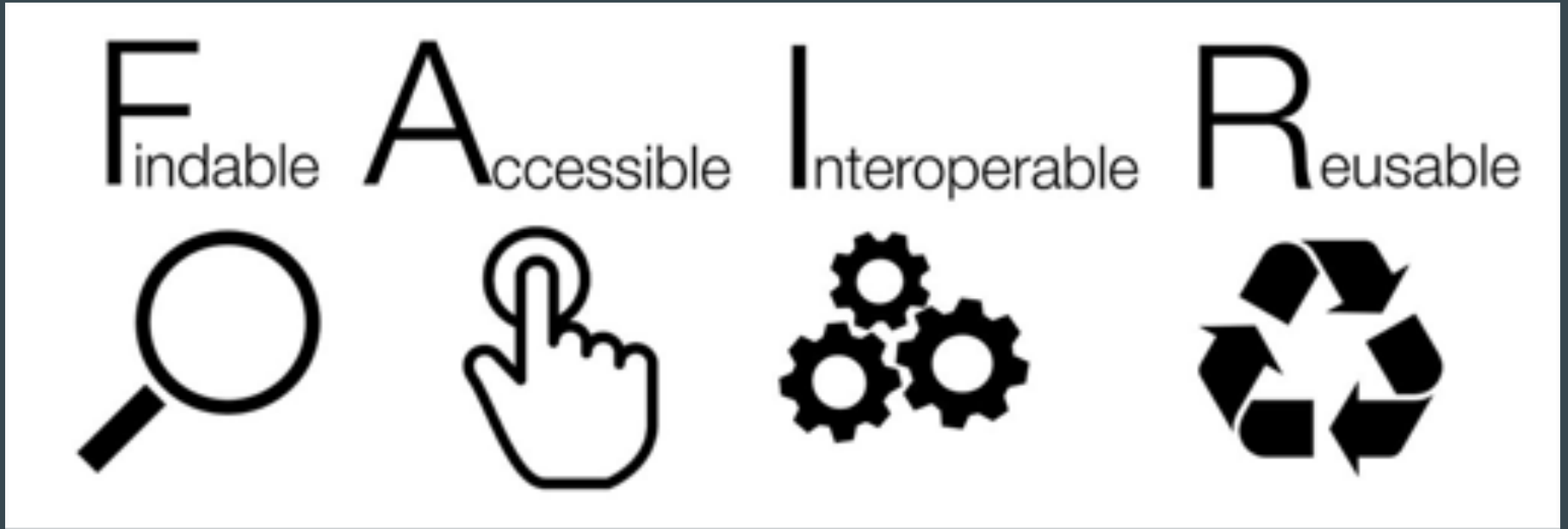
Danie Kinkade, Adam Shepherd, Shannon Rauch, Mathew Biddle, Tina Haskins

Mission

To work with principal investigators to curate and serve data and information from research projects funded by the National Science Foundation's Biological and Chemical Oceanography Sections and the Division of Polar Programs Antarctic Organisms and Ecosystems program.



Making Data FAIR



Original article describing FAIR data principles: Wilkinson, M.D., et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018, <https://doi.org/10.1038/sdata.2016.18>

Article more specific to FAIR in our domain: Stall, S., et al. (2018), Advancing FAIR data in Earth, space, and environmental science, *Eos*, 99, <https://doi.org/10.1029/2018EO109301>

At a Glance...

DATASETS = >9000

AWARDS TRACKED = >1900

CONTRIBUTORS = 2500

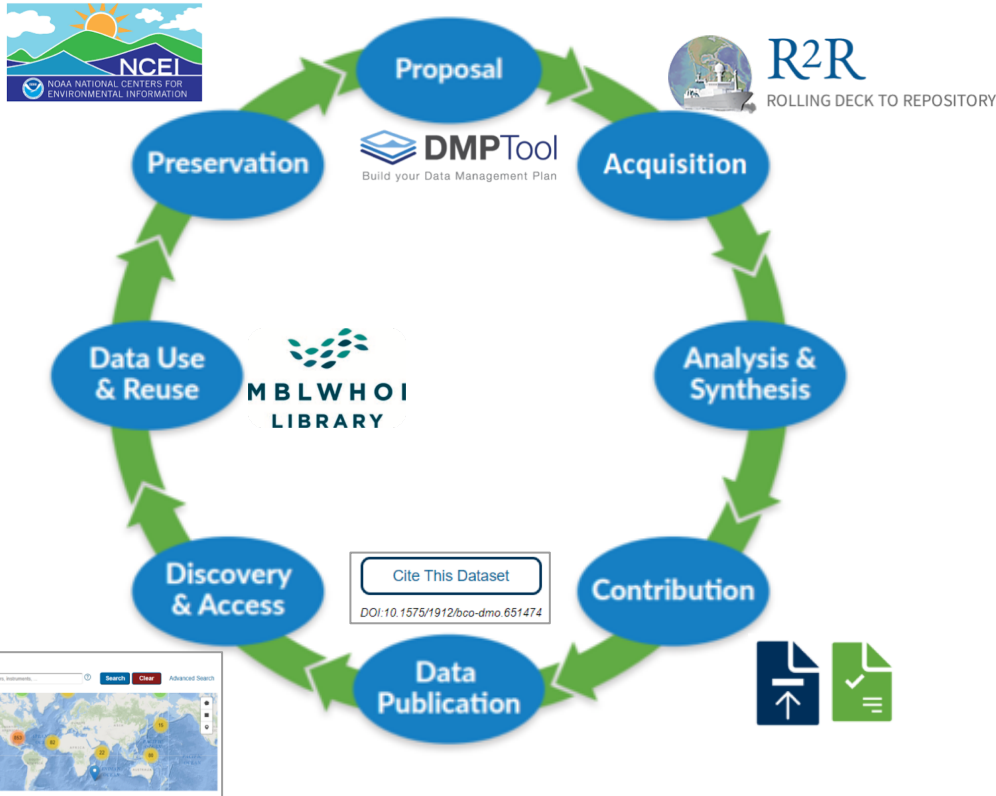
PROJECTS = > 1000

Species Abundances & Biomass
Acoustic Backscatter
Carbonate System Parameters
Chlorophyll & Accessory Pigments
Dissolved & Particulate Nutrients
Ice Cover and Thickness
'Omics-related
Optical Properties
Primary Production
Trace Element Isotopes
Temperature, Salinity, Oxygen

From...the water column, sediment cores,
porewater, controlled experiments, etc.



BCO-DMO supports data stewardship throughout the data lifecycle



- Provide guidance on data formatting and standards;
- Apply gross QC (e.g lat/lon error checking);
- Capture and record metadata;
- Make data and metadata available online (restricted or public access as appropriate);
- Ensure final archive of data in appropriate national data center;
- Provide Data Management Planning template;
- Assign dataset digital object identifiers (DOIs).

Data Management Planning: NSF Data Policies

Proposals submitted to NSF must include a supplementary document of no more than two pages labeled "Data Management Plan" (often referred to as 'the two page plan' or 'DMP'). This document must describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.

Investigators working under awards granted by the NSF Division of Ocean Sciences (OCE) have additional conditions to which they must adhere, as described in the Division of Ocean Sciences Sample and Data Policy.

NSF 17-037
Division of Ocean Sciences (OCE) Sample and Data Policy

This document replaces **NSF 11-060, Division of Ocean Sciences Sample and Data Policy, May 24, 2011.**

Per the NSF policy on the **Dissemination and Sharing of Research Results**, Principal Investigators (PIs) are expected to share with other researchers and the public, at no more than incremental cost and within a reasonable time, the data, samples, physical collections, and other supporting materials created or gathered in the course of work under NSF grants.

All NSF proposals must include a Data Management Plan that describes what data/samples will be collected, what analyses will be done, and how the project will provide open and rapid access to data, samples, derived data products (e.g., models and model output), and other information on the project during and after the project's completion. The Data Management Plan also must specifically discuss how the investigators will achieve the specific OCE data archiving and reporting requirements described below in this document. If the project is not expected to generate new data, samples or derived data products, the Data Management Plan can include a statement that no detailed plan is needed, accompanied by a clear justification. See the **NSF Proposal & Award Policies & Procedures Guide (PAPPG)**, Chapter II.C.2.j for additional information.

DATA AND SAMPLE ARCHIVING REQUIREMENTS

The Division of Ocean Sciences requires that metadata files, full data sets, derived data products and physical collections must be made publicly accessible within two (2) years of collection. This includes software and derived data products (e.g., model results, output, and workflows). A brief description of **preferred data and physical collection archives and centers** and their criteria for submission can be found on the **OCE website** or through contact with the Cognizant Program Officer of the given award. Any limit on access to data, samples, or other information beyond the two-year moratorium period must be based on compelling justification, documented in the Data Management Plan of the proposal, or approved by the cognizant Program Officer.

Where no data or sample repository or archive exists for collected data and samples, the PI is required to identify a preservation plan in the Data Management Plan that complies with the general philosophy of sharing research products and data within two years of collection as described above.

See:

<https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp>



In short, NSF OCE requirements...

- Full data sets, derived data products (e.g. models and model output), and physical collections must be publicly accessible within 2 years of collection.
- The 2-page DMP should describe:
 - Types of samples, data, software, other materials expected to be produced by the project
 - Standards used to format the data and metadata
 - Policies and provisions for sharing
 - Plans for archiving



BCO-DMO
Biological & Chemical Oceanography Data Management Office

Makes it easier for you to meet the
data policy requirements!



DMP Template

BCO-DMO-created a template designed to meet the NSF OCE Data Policy Requirements

Available through DMPTool.org

DMPTool is a free, open-source, online application. It provides detailed guidance and walks researchers through the process of generating comprehensive plans.

Plans can be worked on collaboratively, saved, shared, and exported.



The screenshot shows the BCO-DMO website interface. At the top, there is a navigation bar with the BCO-DMO logo and links for DATA, RESOURCES, and ABOUT US. Below the navigation bar, there is a search bar. The main content area is titled "NSF Two Page Data Management Plan" and includes a table of statistics, a description of the template, and a link to the "Data Management Plan Template" on DMPTool.org.

DATABASE	
Programs	44
Projects	1,055
Deployments	2,864
Platforms	594
Datasets	9,414
Instruments	484
Parameters	1,415
People	2,692
Affiliations	565
Funding	93
Awards	1,963

NSF Two Page Data Management Plan
[Revised June 2018]

Proposals submitted to NSF must include a supplementary document of no more than two pages labeled "Data Management Plan" (often referred to as "the two page plan" or "DMP"). This supplementary document (a supplement to the 15 page proposal) should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See the full description in the Grant Proposal Guide (GPG) Chapter II.C.2 for full policy implementation.

Investigators working under awards granted by the NSF Division of Ocean Sciences (OCE) have additional conditions to which they must adhere, as described in the Division of Ocean Sciences Sample and Data Policy.

If the proposal is being submitted to NSF Geosciences Directorate (GEO) OCE Biological or Chemical Oceanography Sections or Division of Polar Programs (PLR) Antarctic Sciences (ANT) Organisms & Ecosystems Program then the two page plan can state that BCO-DMO staff will work with you to manage the data, and that data or model results generated during the proposed research project will be contributed to the BCO-DMO system.

GEOSPATIAL ACCESS

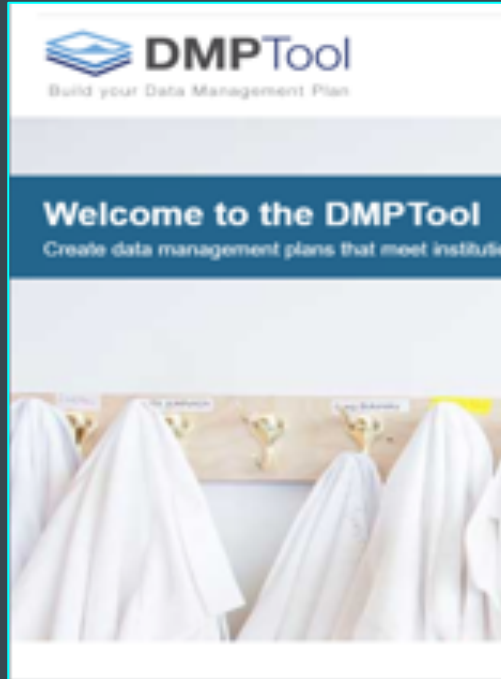
Data Management Plan Template

 **DMPTool**
Build your Data Management Plan

BCO-DMO has developed a Data Management Plan template to assist investigators in submission of plans that meet the NSF OCE Sample and Data Policy requirements. The template can be found and completed on the DMPTool.



DMPTool



DMPTool
Build your Data Management Plan

Welcome to the DMPTool
Create data management plans that meet institutional requirements

Find the BCO-DMO Template:

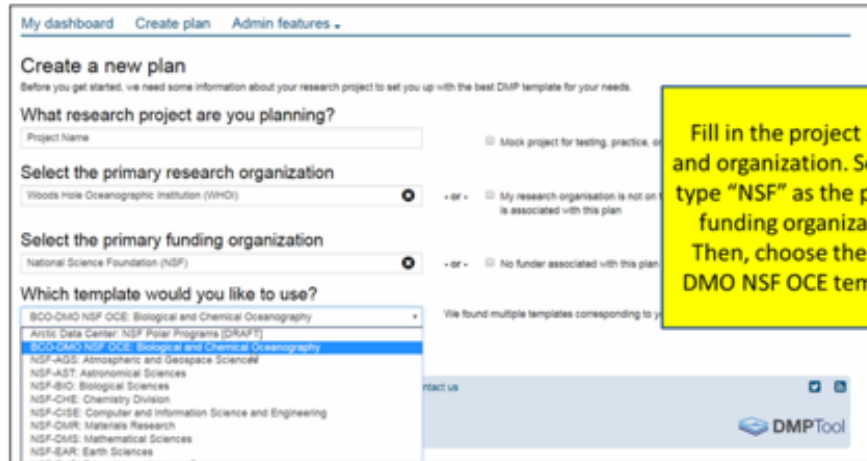


First, click on "Create plan"

My dashboard Create plan Admin features ▾

My dashboard Create plan

The table below lists the plans that you have created, and that have been shared with you by others. You can edit, share, download, make a copy, or remove these plans at any time.



My dashboard Create plan Admin features ▾

Create a new plan
Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

What research project are you planning?
Project Name Mock project for testing, practice, or training

Select the primary research organization
Woods Hole Oceanographic Institution (WHOI) + or My research organization is not on this list

Select the primary funding organization
National Science Foundation (NSF) + or No funder associated with this plan

Which template would you like to use?
BCO-DMO NSF OCE: Biological and Chemical Oceanography We found multiple templates corresponding to your selection

- ARIZ: Data Center: NSF Polar Programs (DRAFT)
- BCO-DMO NSF OCE: Biological and Chemical Oceanography**
- NSF-AGS: Atmospheric and Geospace Sciences
- NSF-AST: Astronomical Sciences
- NSF-BIO: Biological Sciences
- NSF-CHE: Chemistry Division
- NSF-CISE: Computer and Information Science and Engineering
- NSF-CMR: Materials Research
- NSF-CMS: Mathematical Sciences
- NSF-EAR: Earth Sciences

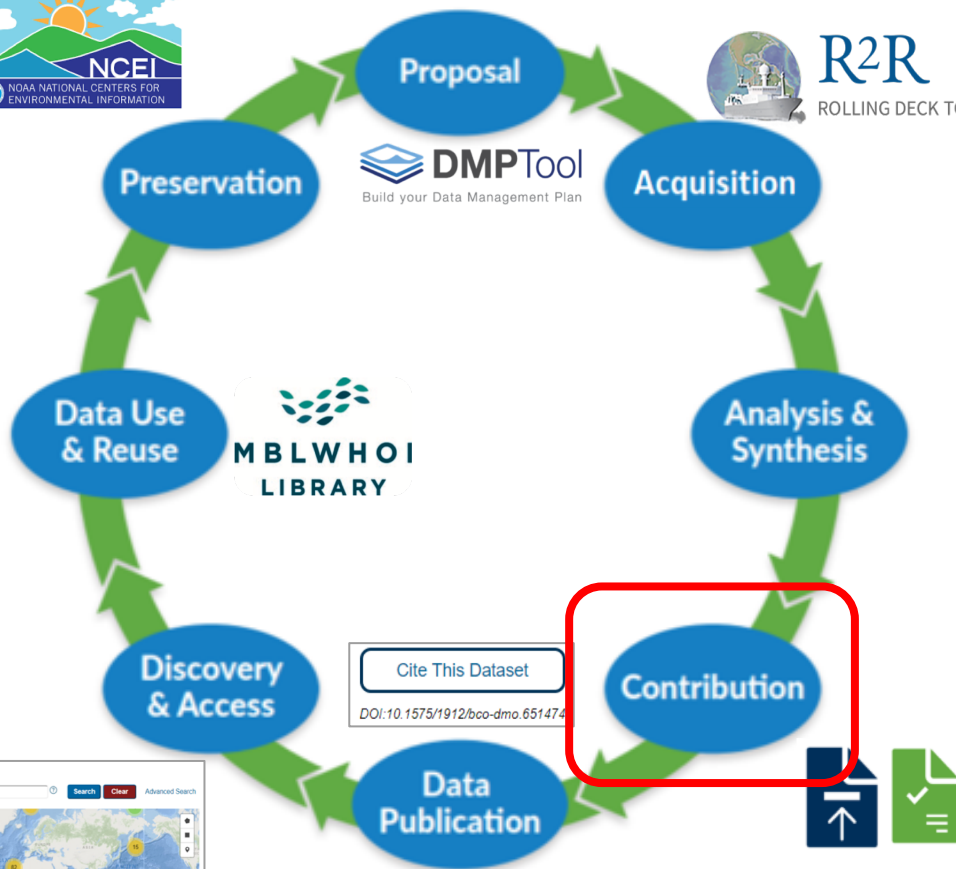
Contact us Facebook Twitter

DMPTool

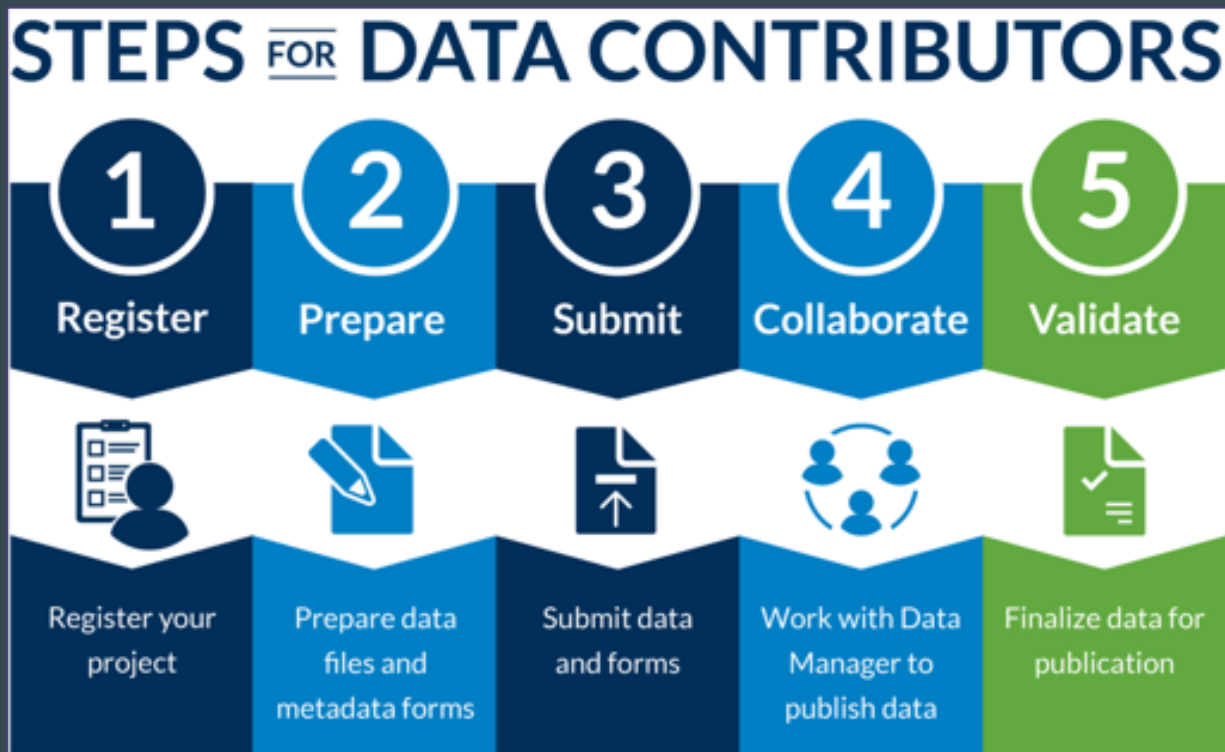
Fill in the project name and organization. Select or type "NSF" as the primary funding organization. Then, choose the BCO-DMO NSF OCE template.



R²R
ROLLING DECK TO REPOSITORY



Data Submission Process



Preparing Data

Submit data in the format most appropriate for your community.

BCO-DMO transforms data into a tabular version (csv, tsv).

For videos, images, and other formats not suitable to tsv/csv, we will work with you to arrive at the best data representation possible.

```

1 unix_timestamp,lat,lon,chl,CDOM,phycoer
2 UTC,degrees_north,degrees_east,RFU,RFU,
3 2015-05-20T21:32:40Z,20.0382,-155.83077
4 2015-05-20T21:42:40Z,20.03822,-155.8307
5 2015-05-20T21:52:40Z,20.03821,-155.8307
6 2015-05-20T21:52:40Z,20.0382,-155.83077
7 2015-05-20T21:32:40Z,20.0382,-155.83077
8 2015-05-20T21:42:40Z,20.03822,-155.8307
9 2015-05-20T21:52:40Z,20.03821,-155.8307
10 2015-05-20T21:52:40Z,20.0382,-155.83077
11 2015-05-20T21:32:40Z,20.0382,-155.83077
12 1970-01-01 2015-05-4 2015-05-20T21:42:40Z,20.03822,-155.8307
13 2015-05-2015-05-5 2015-05-20T21:52:40Z,20.03821,-155.8307
14 2015-05-2015-05-6 2015-05-20T23:41:22Z,20.03819,-155.8307
15 2015-05-2015-05-7 2015-05-20T23:51:22Z,20.03821,-155.8308
16 2015-05-2015-05-8 2015-05-21T00:01:22Z,20.03819,-155.8308
17 2015-05-2015-05-9 2015-05-21T00:11:22Z,20.0382,-155.83077
18 2015-05-2015-05-10 2015-05-21T00:21:22Z,20.03821,-155.8307
19 2015-05-2015-05-11 2015-05-21T00:31:22Z,20.03822,-155.8307
20 2015-05-2015-05-12 1970-01-01T00:00:00Z,,0,0,0,0
21 2015-05-2015-05-13 2015-05-21T01:20:18Z,20.03823,-155.8307
22 2015-05-2015-05-14 2015-05-21T01:30:18Z,20.03825,-155.8307
23 2015-05-2015-05-15 2015-05-21T01:40:18Z,20.03824,-155.8307
    
```

Same data type;
same headers;
are combined if
appropriate

```

/BCO-DMO/MAGI/c3 ---- Level 0
Directory Documentation Download & Other Operations
Level 0 Next Level Flat Listing
# Fluorescence (C3) data from the Honey Badger (G3) Wave Glider
# P.I. Tracy Villareal
# version 7 Jul 2017
=====
ISO_DateTime_UTC lat lon chl CDOM phycoerythrin
2015-05-20T21:32:40Z 20.03820 -155.83077 40.51 49.11 334.0
2015-05-20T21:42:40Z 20.03822 -155.83077 36.89 1186.51 224.5
2015-05-20T21:52:40Z 20.03821 -155.83076 29.58 1190.79 223.83
2015-05-20T23:41:22Z 20.03819 -155.83073 74.4 70.0 95.2
2015-05-20T23:51:22Z 20.03821 -155.83080 10.32 29.96 50.32
2015-05-21T00:01:22Z 20.03819 -155.83088 9.48 32.36 53.56
2015-05-21T00:11:22Z 20.03820 -155.83077 10.36 35.64 53.52
2015-05-21T00:21:22Z 20.03821 -155.83079 10.8 34.4 58.08
2015-05-21T00:31:22Z 20.03822 -155.83077 11.16 35.04 57.6
2015-05-21T01:20:18Z 20.03823 -155.83078 77.6 74.4 95.2
2015-05-21T01:30:18Z 20.03825 -155.83078 11.04 35.72 52.0
2015-05-21T01:40:18Z 20.03824 -155.83076 10.28 31.88 52.0
2015-05-21T01:50:18Z 20.03822 -155.83077 11.2 33.04 53.64
2015-05-21T02:00:18Z 20.03831 -155.83079 10.2 34.72 53.6
2015-05-23T02:10:10Z 20.03820 -155.83073 0.04 34.36 53.64
    
```

One Excel file w/ separate data types: these are split into different datasets

trial	date_local	time_local	site	lat	lon	survivors
1	2007-05-18	11:55	Rocas_Gordon	-0.56596	-90.14065	3
2	2007-05-18	11:56	Rocas_Gordon	-0.56596	-90.14065	3
3	2007-05-18	11:57	Rocas_Gordon	-0.56596	-90.14065	3
4	2007-05-18	11:58	Rocas_Gordon	-0.56596	-90.14065	3
5	2007-05-18	11:58	Rocas_Gordon	-0.56596	-90.14065	3
6	2007-05-18	11:59	Rocas_Gordon	-0.56596	-90.14065	3
7	2007-05-18	12:00	Rocas_Gordon	-0.56596	-90.14065	3
8	2007-05-18	12:01	Rocas_Gordon	-0.56596	-90.14065	3
9	2007-05-18	12:02	Rocas_Gordon	-0.56596	-90.14065	3
10	2007-05-18	12:03	Rocas_Gordon	-0.56596	-90.14065	3
11	2007-05-18	12:04	Rocas_Gordon	-0.56596	-90.14065	3
12	2007-05-18	12:05	Rocas_Gordon	-0.56596	-90.14065	3
13	2007-05-18	12:06	Rocas_Gordon	-0.56596	-90.14065	3
14	2007-05-18	12:07	Rocas_Gordon	-0.56596	-90.14065	3

```

/BCO/Trophic_Cascades/urchin_survivorship
Level 0 Directory Documentation Download & Other Operations
Level 0 Next Level Flat Listing
# survivors from sea urchin tethering experiments
# J. Mitman, F. Seitz (Brown U)
# version: 2016-01-15
=====
trial_date_local year_mon_local day_local site lat lon
1 2007-05-18 2007 05 18 Rocas_Gordon -0.56596 -90.14065
time_local_yrday_local ISO_DateTime_Local survivors
1155 138 1157 138 1157 138 1157 138
1156 138 1156 138 1156 138 1156 138
1200 138 1200 138 1200 138 1200 138
1201 138 1201 138 1201 138 1201 138
1202 138 1202 138 1202 138 1202 138
1203 138 1203 138 1203 138 1203 138
1204 138 1204 138 1204 138 1204 138
1205 138 1205 138 1205 138 1205 138
1206 138 1206 138 1206 138 1206 138
1207 138 1207 138 1207 138 1207 138
1208 138 1208 138 1208 138 1208 138
1209 138 1209 138 1209 138 1209 138
138-4965 2007-05-18T11:55:00.000 3
138-4972 2007-05-18T11:56:00.000 3
138-4979 2007-05-18T11:57:00.000 3
138-4986 2007-05-18T11:58:00.000 3
138-4993 2007-05-18T11:59:00.000 3
138-5000 2007-05-18T12:00:00.000 3
138-5007 2007-05-18T12:01:00.000 3
138-5014 2007-05-18T12:02:00.000 3
138-5021 2007-05-18T12:03:00.000 3
138-5028 2007-05-18T12:04:00.000 3
138-5035 2007-05-18T12:05:00.000 3
138-5042 2007-05-18T12:06:00.000 3
138-5049 2007-05-18T12:07:00.000 3
138-5056 2007-05-18T12:08:00.000 3
138-5063 2007-05-18T12:09:00.000 3
    
```



Preparing Metadata

Metadata: describes your data (where and how it was collected, by whom, analysis methods, funding sources, etc.)

The contents of your metadata form are directly used to populate the public Dataset Landing Page.

Allows your data to be understood and re-used by others.

BCO-DMO Dataset Landing Page: Wave Glider - Fluorescence - C3

Description: This dataset includes chlorophyll, phycoerythrin, and CDOM data from the AUV Honey Badger (Wave Glider model V2) during a 2015 deployment in the North Pacific Ocean.

Acquisition Description: Data were collected at the surface by the AUV Honey Badger. This deployment in the North Pacific Ocean is Badger and project name: <http://hawaii.jpl.nasa.gov/MNCO>

Processing Description: No calibration was deemed useful due to the duration of the mission and nature of questions asked. Sensors returned only RFU.

BCO-DMO dataset landing page

Dataset Metadata Form

BCO-DMO Dataset Metadata Submission Form

All data should be reported in table form. For full instructions, see our "See How to Get Started" page: <http://www.bco-dmo.org/how-get-started>. Please send your completed form or questions to info@bco-dmo.org

Dataset Name: [Preferred short name for the dataset (preferably 30 characters or less)]

Dataset Description: [Brief sentence describing these data (preferably 60 characters or less)]

Originating PI name and contact information:
Name: _____
Email: _____
Phone: _____
Mailing Address: _____

ORCID: [If you don't currently have an ORCID number, you can register for one at <https://orcid.org/register>. ORCID's are unique persistent identifiers assigned to individuals to enable easy disambiguation and ensure proper credit for work. It is free to register.]

Affiliation/Institution during data acquisition: _____

Dataset Name

Dataset Description

Methodology

Sampling and analytical procedures



Metadata Components

- How were the data generated?
 - How were samples collected?
 - How were they processed and analyzed?
 - Methods - include references and citations
 - Instruments - include manufacturer and model; as well as calibration information! (for both lab instruments and shipboard instruments)
 - Be specific; e.g. describe all the sensors on the CTD (not just "CTD")
- Describe the quality control
 - Replicate samples, inter-comparisons, reference materials
- What software or scripts did you use?
 - Are they publicly available?
 - Modeling project? Describe the inputs, parameters, conditions, etc.
- Consider the longevity of your data... they have value beyond your own research!



Submitting Data

Send applicable metadata forms and data files as attachments to info@bco-dmo.org

Talk to us if data are too large for email...

In the works: an auto-submit system. Stay tuned!

YOU WANT YOUR COUSIN TO SEND YOU A FILE? EASY.
HE CAN EMAIL IT TO— ... OH, IT'S 25 MB? HMM...

DO EITHER OF YOU HAVE AN FTP SERVER? NO, RIGHT.
IF YOU HAD WEB HOSTING, YOU COULD UPLOAD IT...

HMM. WE COULD TRY ONE OF THOSE MEGASHARE UPLOAD SITES,
BUT THEY'RE FLAKY AND FULL OF DELAYS AND PORN POPUPS.

HOW ABOUT AIM DIRECT CONNECT? ANYONE STILL USE THAT?

OH, WAIT, DROPBOX! IT'S THIS RECENT STARTUP FROM A FEW
YEARS BACK THAT SYNC'S FOLDERS BETWEEN COMPUTERS.
YOU JUST NEED TO MAKE AN ACCOUNT, INSTALL THE—



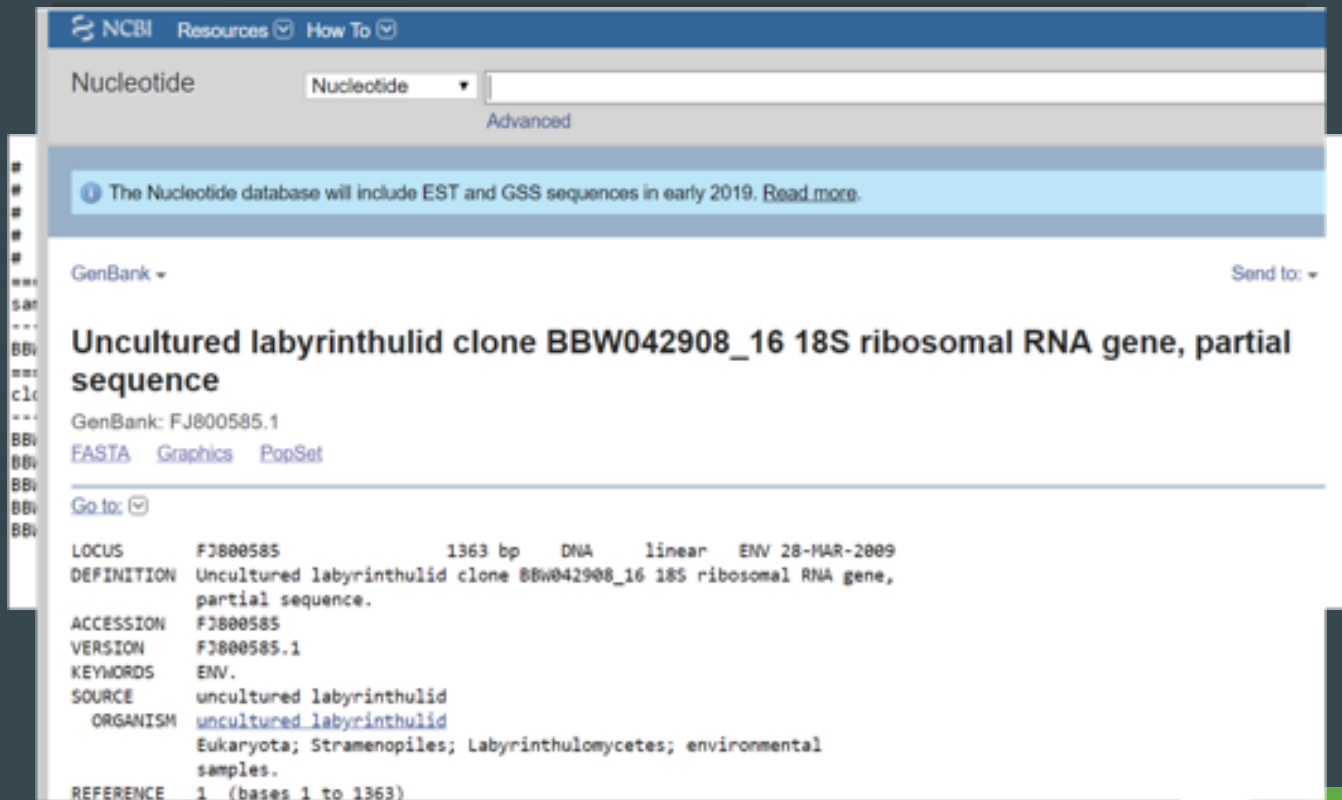
I LIKE HOW WE'VE HAD THE INTERNET FOR DECADES,
YET "SENDING FILES" IS SOMETHING EARLY
ADOPTERS ARE STILL FIGURING OUT HOW TO DO.

Contributing Sequence Data

Sequences themselves are best served by specialized repositories, like NCBI's GenBank.

BCO-DMO serves the related environmental data and links out to NCBI, etc.

This allows all your project data to be discoverable from one place, but served by the domain repositories best suited to do so.



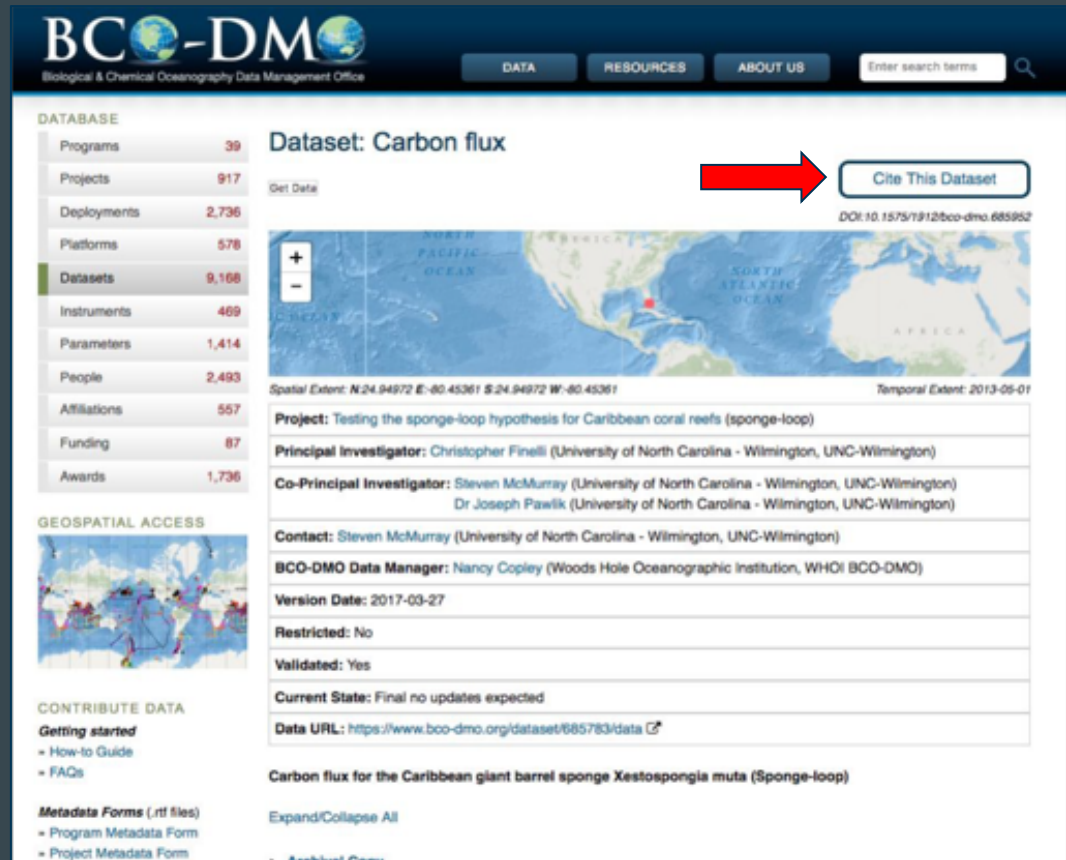
The screenshot shows the NCBI Nucleotide database interface. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' links. Below this is a search bar with 'Nucleotide' selected and a search box. A blue banner below the search bar contains the message: 'The Nucleotide database will include EST and GSS sequences in early 2019. Read more.' The main content area displays the entry for 'Uncultured labyrinthulid clone BBW042908_16 18S ribosomal RNA gene, partial sequence'. The entry includes the GenBank ID 'FJ800585.1', links for 'FASTA', 'Graphics', and 'PopSet', and a 'Go to' button. The entry details are as follows:

LOCUS	FJ800585	1363 bp	DNA	linear	ENV 28-MAR-2009
DEFINITION	Uncultured labyrinthulid clone BBW042908_16 18S ribosomal RNA gene, partial sequence.				
ACCESSION	FJ800585				
VERSION	FJ800585.1				
KEYWORDS	ENV.				
SOURCE	uncultured labyrinthulid				
ORGANISM	uncultured_labyrinthulid				
	Eukaryota; Stramenopiles; Labyrinthulomycetes; environmental samples.				
REFERENCE	1 (bases 1 to 1363)				

Data Publication

Once data & metadata are validated by the submitter, a Digital Object Identifier (DOI) is assigned to the dataset.

This makes it easier to cite the dataset and discover it (e.g. from a publication to the repository).



The screenshot shows the BCO-DMO (Biological & Chemical Oceanography Data Management Office) website. The header includes the BCO-DMO logo and navigation buttons for DATA, RESOURCES, and ABOUT US, along with a search bar. The main content area is titled 'Dataset: Carbon flux'. On the left, a 'DATABASE' sidebar lists various categories with counts: Programs (39), Projects (917), Deployments (2,736), Platforms (578), Datasets (9,168), Instruments (469), Parameters (1,414), People (2,493), Affiliations (557), Funding (87), and Awards (1,736). Below this is a 'GEOSPATIAL ACCESS' section with a world map. The 'CONTRIBUTE DATA' section includes links for 'Getting started', 'Metadata Forms', and 'How-to Guide'. The main dataset page features a 'Cite This Dataset' button (highlighted with a red arrow), a map of the Caribbean region, and a table of metadata including project details, investigators, contact information, version date, and data URL.

BCO-DMO
Biological & Chemical Oceanography Data Management Office

DATA RESOURCES ABOUT US Enter search terms

DATABASE

Programs	39
Projects	917
Deployments	2,736
Platforms	578
Datasets	9,168
Instruments	469
Parameters	1,414
People	2,493
Affiliations	557
Funding	87
Awards	1,736

Dataset: Carbon flux

Get Data **Cite This Dataset**
DOI: 10.1575/1512/bco-dmo.685952

Project: Testing the sponge-loop hypothesis for Caribbean coral reefs (sponge-loop)

Principal Investigator: Christopher Finelli (University of North Carolina - Wilmington, UNC-Wilmington)

Co-Principal Investigator: Steven McMurray (University of North Carolina - Wilmington, UNC-Wilmington)
Dr Joseph Pawlik (University of North Carolina - Wilmington, UNC-Wilmington)

Contact: Steven McMurray (University of North Carolina - Wilmington, UNC-Wilmington)

BCO-DMO Data Manager: Nancy Copley (Woods Hole Oceanographic Institution, WHOI BCO-DMO)

Version Date: 2017-03-27

Restricted: No

Validated: Yes

Current State: Final no updates expected

Data URL: <https://www.bco-dmo.org/dataset/685783/data>

Carbon flux for the Caribbean giant barrel sponge *Xestospongia muta* (Sponge-loop)

Expand/Collapse All



Data Publication

A suggested citation format is provided, supporting data re-use.



The screenshot shows the BCO-DMO website interface. At the top, there is a navigation bar with "DATA", "RESOURCES", and "ABOUT US" buttons, and a search bar. Below the navigation bar, a sidebar on the left lists database categories: Programs (39), Projects (917), Deployments (2,736), Platforms (578), Datasets (9,168), and Instruments (469). The main content area displays "Dataset: Carbon flux" with a "Get Data" button and a "Cite This Dataset" button. A map of the Caribbean region is visible, with a red dot indicating the location of the data. A white pop-up box is overlaid on the map, containing the following text:

Data Citation:

Finelli, C., Pawlik, J., McMurray, S. (2017) Carbon flux for the Caribbean giant barrel sponge *Xestospongia muta* (Sponge-loop). Biological and Chemical Oceanography Data Management Office (BCO-DMO). Dataset version 2017-03-27 [if applicable, indicate subset used]. doi:10.1575/1912/bco-dmo.685952 [access date]

Terms of Use

All data sets are licensed under a [Creative Commons Attribution 4.0 International License \(CC BY 4\)](#). Per the CC BY 4 license it is understood that any use of the data set will properly acknowledge the individual(s) listed above using the suggested data citation. If you wish to use this data set, it is highly recommended that you contact the original principal investigator(s) (PI). Should the relevant PI be unavailable, please contact BCO-DMO (info@bco-dmo.org) for additional guidance. For general guidance please see the BCO-DMO [Terms of Use](#) document.

Version Date: 2017-03-27

Restricted: No

Validated: Yes

Current State: Final no updates expected

Data URL: <https://www.bco-dmo.org/dataset/685783/data>

Carbon flux for the Caribbean giant barrel sponge *Xestospongia muta* (Sponge-loop)

Expand/Collapse All

Archival Copy



Data Discovery & Access

- BCO-DMO data holdings are freely accessible to the public
- No login or account creation needed
- Discoverable via text and geospatial search interfaces

BCO-DMO
Biological & Chemical Oceanography Data Management Office

DATA RESOURCES ABOUT US Enter search terms

DATABASE

Programs	39
Projects	916
Deployments	2,737
Platforms	578
Datasets	9,168
Instruments	469
Parameters	1,414
People	2,497
Affiliations	558
Funding	87
Awards	1,737

GEOSPATIAL ACCESS

Dataset Search

carbon flux + Pawlik Hide Advanced Search

Collection Date Start: 1995 End: 2018

Status Select a status

Validated Yes

Embargoed No 1

Search Clear

Total: 137 (1 of 10)

[+] Dataset	People	Award
[+] Carbon flux for the Caribbean giant barrel sponge <i>Xestospongia muta</i>	Principal Investigator	OCE-155850

Preservation

Once a project's data and metadata are published online at BCO-DMO, they are then submitted to an appropriate national data center for long-term preservation, e.g., the National Centers for Environmental Information (NCEI).




NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Home Access Data Submit Data Public Outreach About Search NCEI Data

Home > Data > Metadata > gov.noaa.ncei:0112166

BIOMASS, ANIMALS - INDIVIDUAL - COUNTS, SPECIES IDENTIFICATION, TAXONOMIC CODE and species abundance trawl data collected in the South Atlantic Ocean and South Pacific Ocean on the LAURENCE M. GOULD and NATHANIEL B. PALMER cruises LMG0104, LMG0203 and others as part of the Southern Ocean GLOBEC project from 2001-04-30 to 2002-09-08 (NCEI Accession 0112166)


Preview graphic

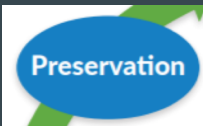
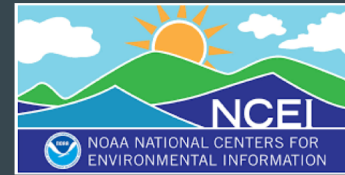
NOOC Accession 0112166 includes biological and trawl data collected aboard the LAURENCE M. GOULD and NATHANIEL B. PALMER during cruises LMG0104, LMG0203, NBP0104 and NBP0204 in the South Atlantic Ocean and South Pacific Ocean from 2001-04-30 to 2002-09-08. These data include BIOMASS, ANIMALS - INDIVIDUAL - COUNTS, SPECIES IDENTIFICATION, TAXONOMIC CODE and species abundance. The instruments used to collect these data include Multiple Opening/Closing Net and Environmental Sensing System (MOCNESS). These data were collected by Jose Torres of University of South Florida as part of Southern Ocean GLOBEC. The Biological and Chemical Oceanography Data Management Office (BCO-DMO) submitted these data to NOOC on 2013-07-24.

The following is the text of the abstract provided by BCO-DMO:

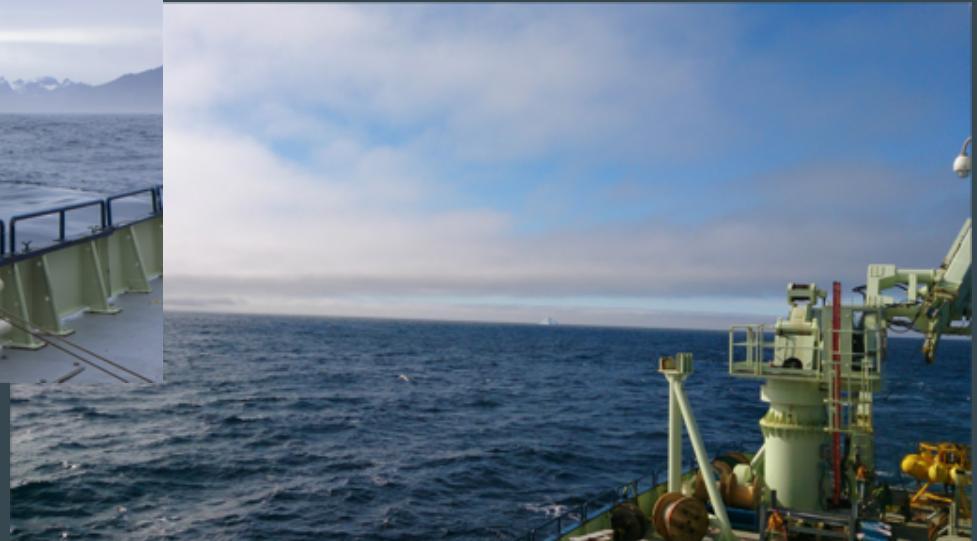
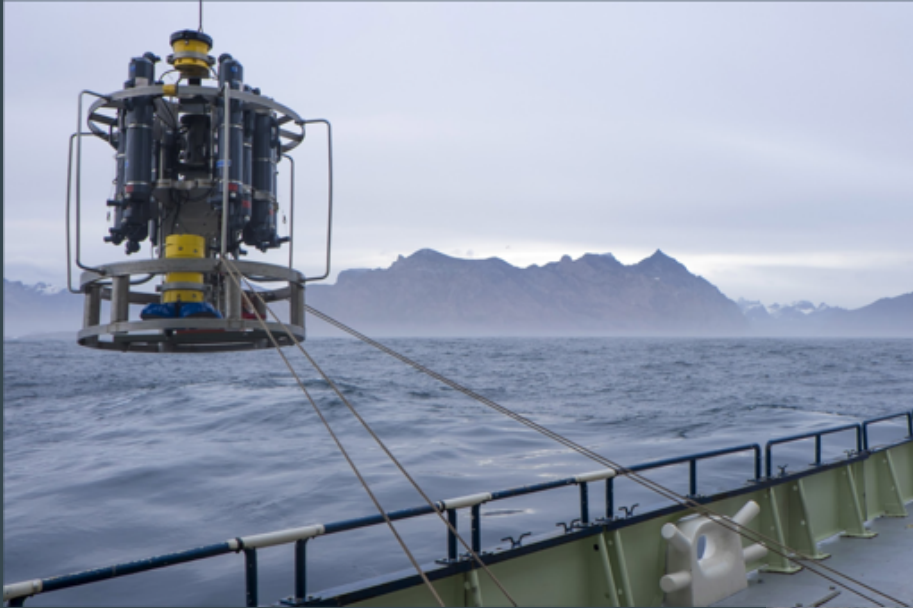
Fish abundance data from MOC-10 trawls.
[Show less](#)

[Dataset Citation](#)
[Dataset Identifiers](#)
[ISO 19115-2 Metadata](#)

Access Time & Location Documentation Description Credit Keywords Constraints Lineage



Successful Cruise Data Management...



Cruise Planning & Reporting

Cruise Plan (pre-cruise)

Document describing:

- Ship name, cruise ID
- Participants & their roles
- Dates & Ports of call
- Scientific objectives & planned operations
- Proposed cruise track & station locations
- List of instrumentation on board
- Station sampling plan & allocation of water/samples

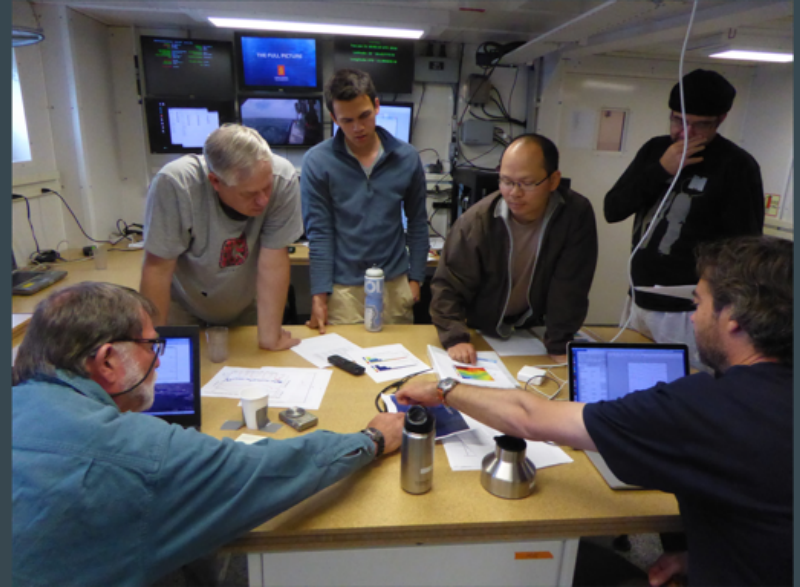
Cruise Report (post-cruise)

Document describing:

- Ship name, cruise ID
- Participants & their roles
- Actual dates & Ports of call
- Scientific objectives & operations accomplished
- Actual cruise track & station locations
- List of instrumentation used, description of problems encountered
- Preliminary results

Chief Scientist data responsibilities

- main point(s) of contact for BCO-DMO,
- submits event log, sample logs, cruise report, station list, hydrography, etc. to BCO-DMO,
- a list of expected datasets & the responsible PI -- very helpful!
- **** Make sure cruise participants know where to find final version of logs & hydrography ****



Event Logs

A chronological record of all scientific sampling events that happened during a cruise, wherein each sampling event is assigned a unique identifier. Recommended fields:

- Event #: unique to the cruise; perhaps unique across cruises for a multi-cruise project
- Instrument type/name/model
- Station #
- Cast # (if applicable)
- Date and time (specify UTC or local; if local, include time zone)

Suggested format ISO: yyyy-mm-ddTHH:MM:SS

- Latitude and longitude
- Sampling depth (for ranges use depth_min and depth_max)
- Depth of the water
- Investigator
- Notes/comments



The event log allows investigators to integrate data from different sampling devices used during a cruise.

Event Logs

The R2R Event Logger is installed on many UNOLS vessels.

The R2R Eventlogger is a program that creates a record of the scientific sampling events conducted during a cruise.

At the end of the cruise, the final event log can be exported as a plain text csv file.



Scientific Sampling Event Log

Event	Instrument	Action	Transect	Station	Cast	Time Local	Latitude	Longitude	Depth	12Khz	Author	Comment	Revisions
20100922 0900	Ship	crew/start				9:01 AM	43.492217	-71.4287			glawson		glawson & 22 Sep 2010 16:51
20100922 1040	GreenBomber	deploy			1	10:46 AM	43.42265	-71.405033			glawson	test deployment	glawson & 23 Sep 2010 19:46
20100922 1120	Hammerhead	deploy			1	11:22 AM	43.411083	-71.41975			glawson	test deployment	glawson & 23 Sep 2010 19:46
20100922 1240	Ship	OnStation				12:49 PM	43.32265	-71.4336			glawson	station #0	glawson & 22 Sep 2010 16:56
20100922 1300	VM	deploy			1	1:08 PM	43.315547	-71.430433			glawson	test deployment	glawson & 23 Sep 2010 18:55
20100922 1310	VM	recover			1	1:16 PM	43.312783	-71.430417			glawson	test recovery	glawson & 23 Sep 2010 18:57
20100922 1330	Hammerhead	recover			1	1:30 PM	43.308247	-71.430583			glawson	test recovery	glawson & 23 Sep 2010 19:47
20100922 1350	GreenBomber	recover			1	1:58 PM	43.30295	-71.433433			glawson	test recovery	glawson & 23 Sep 2010 19:48
20100922 1415	Ship	SafetyOffStart				2:15 PM	43.3008	-71.4231	25.35		glawson	meeting in the galley	glawson & 22 Sep 2010 18:15
20100922 1520	Ship	SafetyOffEnd					43.533283	-71.1237	28.54		glawson		glawson & 22 Sep 2010 19:25
20100922 1534	ObserverMannak	start				2:34 PM	43.517613	-71.087	28.11		zvanderhoop	Test Observer Protocol	
20100922 1613	ObserverMannak	end					43.383433	-70.930833	33.55		zvanderhoop	Test Observer Protocol	
20100923 0640	GreenBomber	deploy			2	6:40 AM	41.94695	-67.630183			glawson	latlon feed not working according to ship, latlon is 42 58 07.7 N and 67 37 01.1 W	glawson & 25 Sep 2010 22:01
20100923 0835	Hammerhead	deploy			2	6:55 AM	42.00027	-67.63275			glawson	latlon feed not working according to ship, latlon is 42 00 04.4 N and 67 37 06.5 W	glawson & 25 Sep 2010 22:15
20100923 0700	ADCP75	start				7:03 AM	42.0074	-67.6374			glawson	starting it with external trigger (lat lon is 42 00 44.4 N and 67 38 24.4 W)	glawson & 25 Sep 2010 22:18
20100923 0727	GreenBomber	start			1	7:27 AM	42.073617	-67.64635			glawson		

See:

<https://www.rvdata.us/about/event-log>

Log Sheets

- Per sampling device
- Physical or digital
 - If hard copies, scan into PDFs and, ideally, transcribe into a spreadsheet

VPR DATA SHEET Day / Night

Cruise TI 715 Location Sta 2 Cast# 2

Date 10/22/13 Wind Speed 5.6 m/s Direction 183.5°

Year/Day _____ Sea State Calm

Local Time 912 to 945 Start: Lat 42°20.226' Long 69°46.964'

GMT Time 1312 to 1345 End: Lat _____ Long 69°

Start Battery 27.7 End Battery 26.2 Magnification S1

Raw Filename: 138247489 Processed Directory _____

Start frame: 3990 End frame: 11300 Cast Depth: 251m

INTERESTING OBSERVATIONS MADE DURING EXTRACTION

Depth Frame#

25 m 566 5976 8380

9100

Event #:	Station:	Date:	Oxy Box:	Salt Box:	Nut Box:						
2065	3	10/20/10		9	Light Blue						
Cast:	7	Time up: 0630									
Bottle	Nominal Depth	GEOTRACES Number	Unfiltered Oxygen	OxyT	Nutrients	Salinity	Pigments	Th-234	U-238	Radium	
1	2100	5793			1			100		30	
2	2100	5794			2	2		100		29	
3	1500	5795			3	3				28	
4	900	5796			4	4				27	
5	375	5797			5	5				26	
6											
7											
8											
9											
10											
11											
12											
"13" Surface Pump											
Approximate Volume:			1		50	50		50		50	
Initials of Collector:					0.3	0.5	4-4.5	4.5	0.1	Remainder	
Comments:											

Log Sheets

Pump Cast: Stn 3

Pump ID	Flow S/N	Prev. End	Start (l)	End (l)	Calc. Diff
1	19008635	38608.4	38608.4	39691.0	1082.6
	19008636	29922.6	29922.6	30412.1	489.5
	19008637	70050.8	70050.9	71632.1	1581.2
2	19008629	60241.0	60241.0	61249.5	1008.5
	19008630	25809.6	25809.6	26356.7	547.1
	19008631	85585.7	85585.7	87137.8	1552.1
3	19008656	78742.9	78743.0	79145.3	402.3
	19008657	41886.1	41886.0	42047.4	161.4
	19008658	127823.1	127823.0	128395.1	572.1
4	19008638	27398.4	27398.4	28205.0	806.6
	19008639	26762.3	26762.3	27023.5	261.2
	19008640	55191.5	55191.5	56266.3	1074.8
5	20007741	45536.5	45536.5	45602.8	66.3
	20007742	37393.0	37393.0	37667.0	274.0
	20007743	83951.7	83951.7	84324.5	372.8
6	20008125	18408.5	Not Depl	18408.5	0.0
	20008126	25363.8	Not Depl	25363.8	0.0
	20008127	44858.0	Not Depl	44858.0	0.0
7	20007747	24458.0	Not Depl	24458.0	0.0
	20007748	12025.0	Not Depl	12025.0	0.0
	20007749	44450.7	Not Depl	44450.7	0.0

Dates Y/M/D and times in local time (and GMT)

GMT=LCL+ 5

Date 2013/11/01 (2013-11-01)

Event 4047

Station 3

Cast 4

Water depth 180 m

Delay duration 00:45:00

Pump duration 04:00:00

Trigger time 13:10 (18:10)

Pumps on 13:55 (18:55)

Pumps off 17:55 (22:55)

DESCRIPTION SCANNED_PUMP_LOG

- [#882_Mcl_81ns2_2c.pdf](#)
- [#813_Mcl_3_3c.pdf](#)
- [#819_8e7_3_7s.pdf](#)
- [#821_Mcl_3_9c.pdf](#)
- [#821_Mcl_3_9s.pdf](#)
- [#824_Mcl_3_12c.pdf](#)
- [#824_Mcl_3_12s.pdf](#)
- [#837_Mcl_2_3c.pdf](#)
- [#847_Mcl_3_4c.pdf](#)
- [#856_Mcl_4_3c.pdf](#)

GEOTR
PUMPS
PIs: I
Vents:

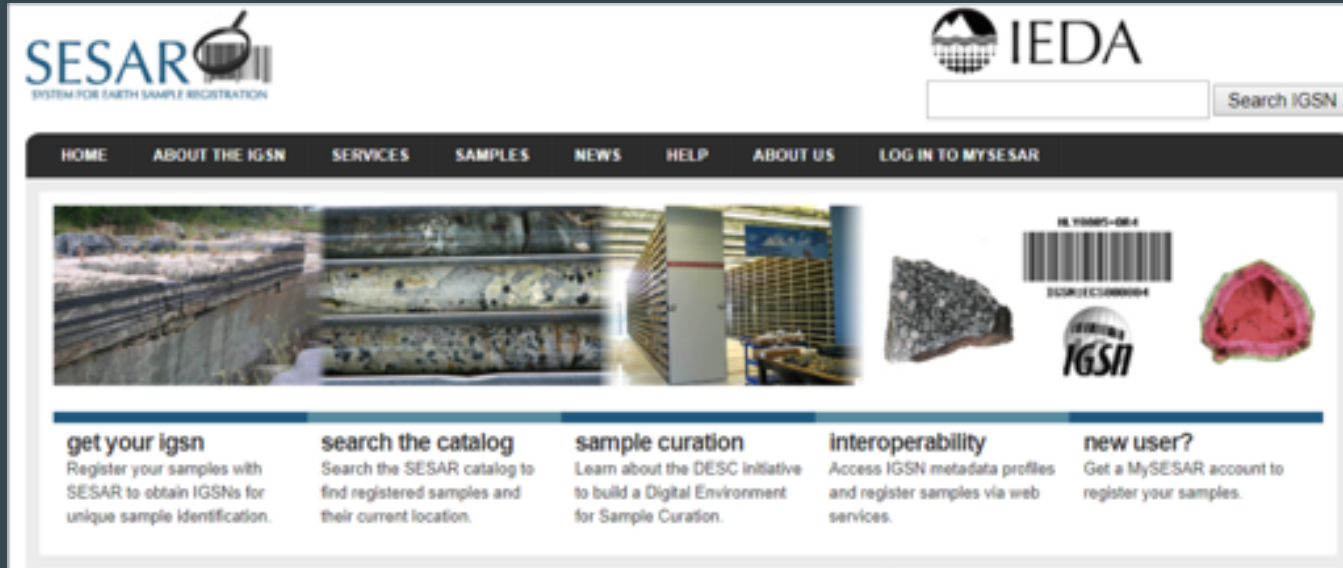
PUMP_LOG

#882_Mcl
#813_Mcl
#819_8e7
#821_Mcl
#821_Mcl
#824_Mcl
#824_Mcl
#837_Mcl
#847_Mcl
#856_Mcl

International Geo Sample Numbers (IGSNs)

The IGSN is a persistent unique identifier for physical samples and specimens that eliminates the problems associated with the ambiguous naming of samples. In the U.S., you can obtain IGSNs using the System for Earth Sample Registration (SESAR) at IEDA Data Facility.

See <https://igsn.github.io/overview/> and <http://www.geosamples.org/>



The screenshot shows the SESAR website interface. At the top left is the SESAR logo with the tagline 'SYSTEM FOR EARTH SAMPLE REGISTRATION'. At the top right is the IEDA logo and a search bar labeled 'Search IGSN'. Below the header is a navigation menu with links: HOME, ABOUT THE IGSN, SERVICES, SAMPLES, NEWS, HELP, ABOUT US, and LOG IN TO MYSESAR. The main content area features a large image of a sample storage facility with rows of sample trays. To the right of the image is a sample specimen, a barcode labeled 'IGSN: 100001-00-1', and a red sample. Below the image and barcode are five columns of text:

- get your igsn**
Register your samples with SESAR to obtain IGSNs for unique sample identification.
- search the catalog**
Search the SESAR catalog to find registered samples and their current location.
- sample curation**
Learn about the DESC initiative to build a Digital Environment for Sample Curation.
- interoperability**
Access IGSN metadata profiles and register samples via web services.
- new user?**
Get a MySESAR account to register your samples.

Taxonomy - WoRMs

The World Register of Marine Species (WoRMs) provides an authoritative and comprehensive list of names of marine organisms, including information on synonymy.

/BCO-DMO/Southern_Ocean_Pinnipeds/fossil_seal_aa_isotopes ---- Level 0

[Directory](#) [Documentation](#) [Download & Other Operations](#)

[Level 0](#) [Next Level](#) [Flat Listing](#)

Fossil Seal Amino Acid Isotopes
PIs: Paul L. Koch (UC Santa Cruz) & Brenda Hall (UMaine)
Co-PIs: Daniel P. Costa (UC Santa Cruz) & A. Rus Hoelzel (Durham University)
Version: 26 March 2018

Common_name	Scientific_name	WoRMS_LSID	AphiaID
Crabeater_seal	Lobodon_carcinophaga	urn:lsid:marinespecies.org:taxname:231416	231416
Southern_elephant_seal	Mirounga_leonina	urn:lsid:marinespecies.org:taxname:231413	231413
Weddell_seal	Leptonychotes_weddellii	urn:lsid:marinespecies.org:taxname:195932	195932

We recommend checking species names in WoRMs and including identifiers in your data when possible.

Data Preparation Tips

- Use consistent formatting!
 - For site/station names/numbers, dates, times, position, column names (e.g. "EVENT" vs "event" vs "event_number" vs "Event Number")

Example: Poor Data Entry

Site	Date	Plot	Species	Weight	Adult	Site	Plot	Adult	RodentSp	Weight
DeepWell	2/13/2010	1	DIPU	12.1	j	DW		1 y	PERO	12
Deep Well	Feb-10	2	Pero	13.22	j	RS		2 j	PERO	escaped <15
nioSalado	2/13/2010	1a	pero	16	N	RS		3 n	Clegap	91
nioStadu			ClvGap	18.92	gut away					
			Mean1	15.06						

Site	Plot	Adult	Species	grams	Comments
deep well	1 y		woodrat	15	
niosalado	2 y		PERO	24.5	
niosalado	3 y		Clegap	91	

Inconsistency between data collection events

- Location of Date information
- Inconsistent Date format
- Inconsistent Column names
- Different site spellings
- Codes used for site names for some data, but spelled out for others
- Text and numbers in same column – what is the mean of "escaped < 15?"

Data Preparation Tips

Columns of data are consistent: only numbers, dates, or text

Consistent names, codes, formats (date) used in each column.

Data are all in one table, which is much easier for a statistical program to work with than multiple small tables which each require human intervention.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	Site	Date	Plot	Species	Weight	Adult		Rodent Trapping 3/15/2010							
2	DeepWell	2/13/2010		1 DIPO	12.1	j		Site	Plot	Adult	RodentSp	Weight			
3	Deep Well	Feb-10		2 Pero	13.22	j		DW		1 y	Pero		12		
4	rioSalado	2/13/2010	1a	pero	16	N		RS		2 j	PERO	escaped <15			
5	riuStadu	*	1*	CleGap	18.92	gut away		RS		3 n	CleGap	91			
6				Mean1	15.06										
7															
8															
9															
10															
11															

	A	B	C	D	E	F	G	H
1	Date	Site	Plot	Species	Weight	Adult	Comments	
2	2/5/2010	Deep Well		1 DIPO	13.2	y		
3	2/4/2010	Deep Well		1 CLEGAP	11.6	j		
4	2/5/2010	Rio Salado		1 DIPO	14.2	y		
5	2/5/2010	Rio Salado		2 PERO	10.1	y		
6	3/15/2010	Deep Well		1 DIPO	15.2	y	plot burned	
7	3/15/2010	Deep Well		2 DIPO	21.7	y	pregnant	
8	3/15/2010	Rio Salado		1 CLEGAP	16.2	j		
9								
10								
11								
12								

Example:
This is better!

Data Formatting Best Practices

- Create descriptive column names without spaces or special characters. Use underscores instead of symbols.
 - Temp 30 meters → Temp_30_m
 - Species Code → species_code
- Avoid using numbers at the beginning of a column name (some programs have trouble with this)
- Use descriptive file names.
 - A file named `PIV_E_gracilis_20180524.csv` provides useful information about the data (type = Particle Image Velocimetry, species = *E. gracilis*, date= May 24, 2018).
 - Consistent formatting of file names allows for sorting and organizing of the files → keep in mind for images and video.
 - CTD files for example: `cruiseID_station_cast.csv` (e.g. `KM1104_12_1.csv`)



Data Formatting Best Practices

- Missing Data:
 - Again, be consistent!
 - Blank cells have no meaning...use "nd" or "NaN" (-999)
 - Bear in mind that "0" has meaning (0 = measured and not found as opposed to "not measured")
- Round data to appropriate number of decimal places
- Document all codes and quality flag definitions in the metadata
- Don't rely on Excel formatting to convey meaning (e.g. colored cells)



Any/all data: include **lat/lon**, **date**, and **time** whenever possible.

Quality Flags

You can incorporate data quality and uncertainty into your data and/or metadata.

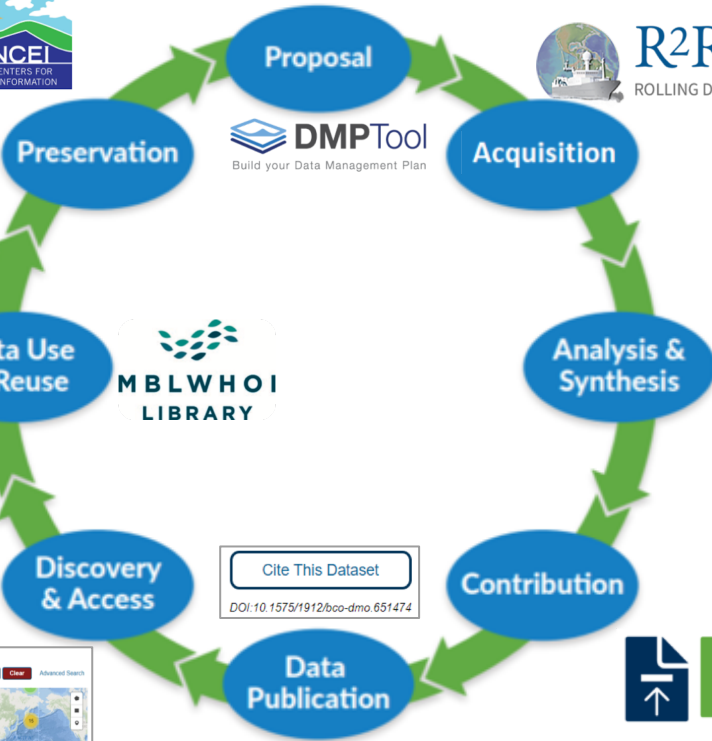
e.g., IODE quality flags. See: https://www.iode.org/mg54_3

Value	Primary-level flag short name	Definition
1	Good	Passed documented required QC tests
2	Not evaluated, not available or unknown	Used for data when no QC test performed or the information on quality is not available
3	Questionable/suspect	Failed non-critical documented metric or subjective test(s)
4	Bad	Failed critical documented QC test(s) or as assigned by the data provider
9	Missing data	Used as place holder when data are missing

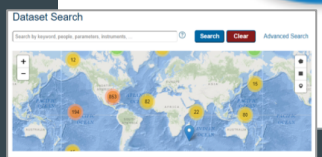
ODV, WOCE, other quality flag systems → use whatever is appropriate for your data/your community.



R2R
ROLLING DECK TO REPOSITORY



Cite This Dataset
DOI: 10.1575/1912/bco-dmo.651474



Resources

- BCO-DMO: <https://www.bco-dmo.org/resources> (DMP Template, BCO-DMO Quick Guide)
- OCE Data Policy: <https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp>
- DMPTool: <https://dmptool.org/>
- R2R: <https://www.rvdata.us/>
- NCBI - how to submit: <https://www.ncbi.nlm.nih.gov/guide/howto/submit-sequence-data/>
- World Register of Marine Species: <http://www.marinespecies.org/index.php>
- IGSN: <http://www.geosamples.org/>
- DataONE Data Management Hub: <https://dataoneorg.github.io/Education/>
- Data Management Short Course: <http://commons.esipfed.org/datamanagementshortcourse>



Wait, did you record the **metadata?**

Questions?

srauch@whoi.edu -- info@bco-dmo.org -- www.bcodmo.org