

Frictionless Data Processing in the Wild

csv,conf,v4

May 8th, 2019 Portland, OR

Amber York, Conrad Schloer,

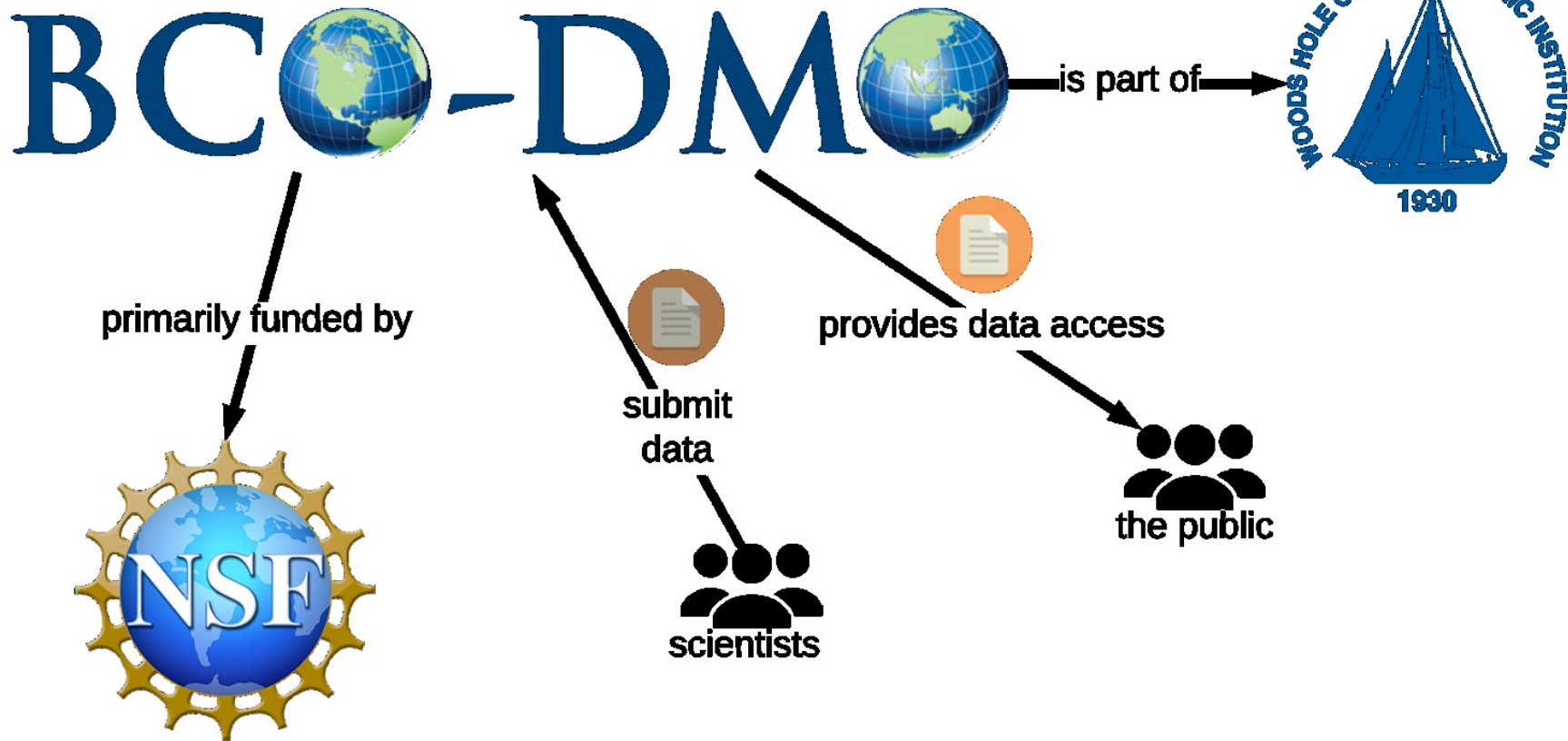
Nancy Copley, Mathew Biddle, Shannon Rauch, Christina Haskins, Karen Soenen, Adam Shepherd, Danie Kinkade



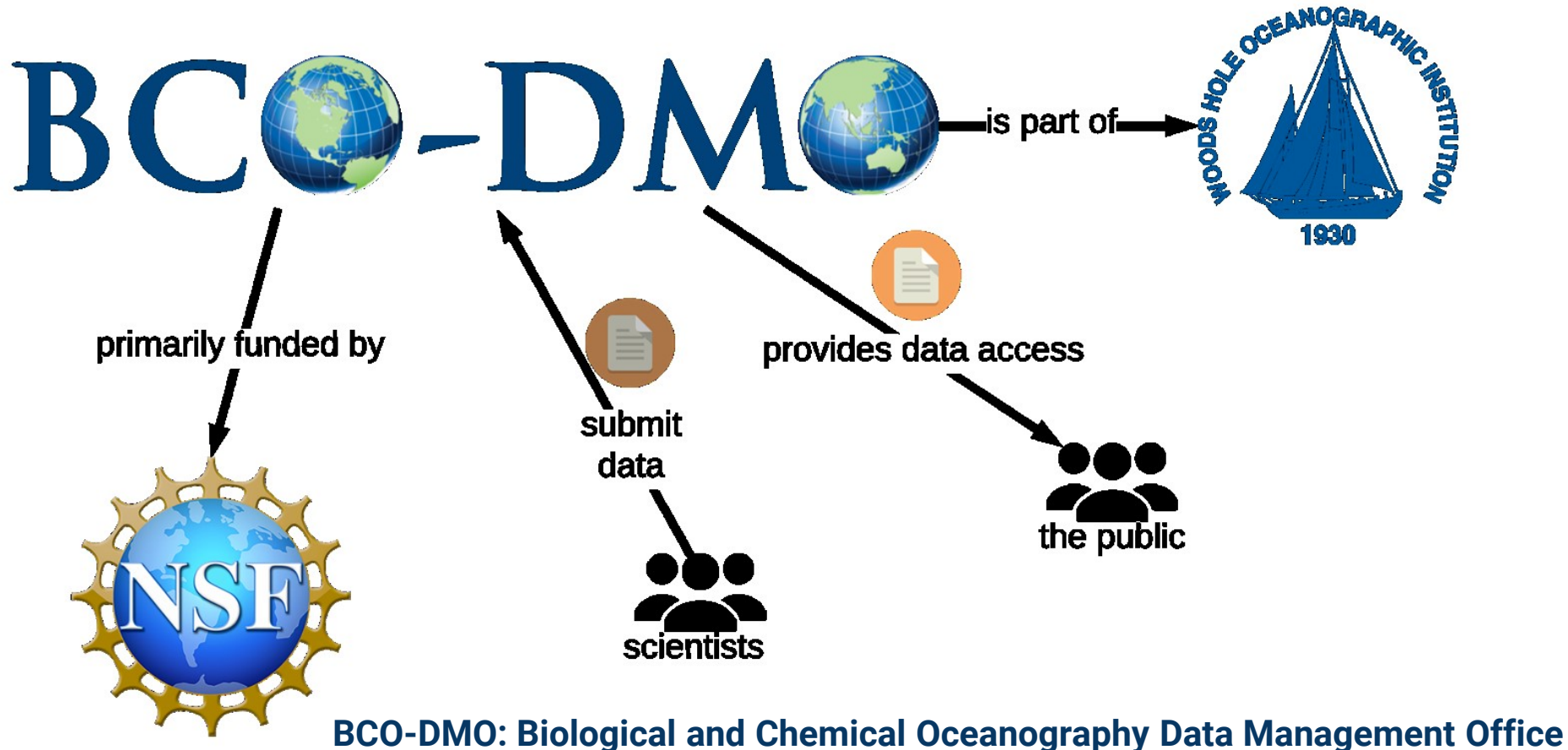
OCE-1435578



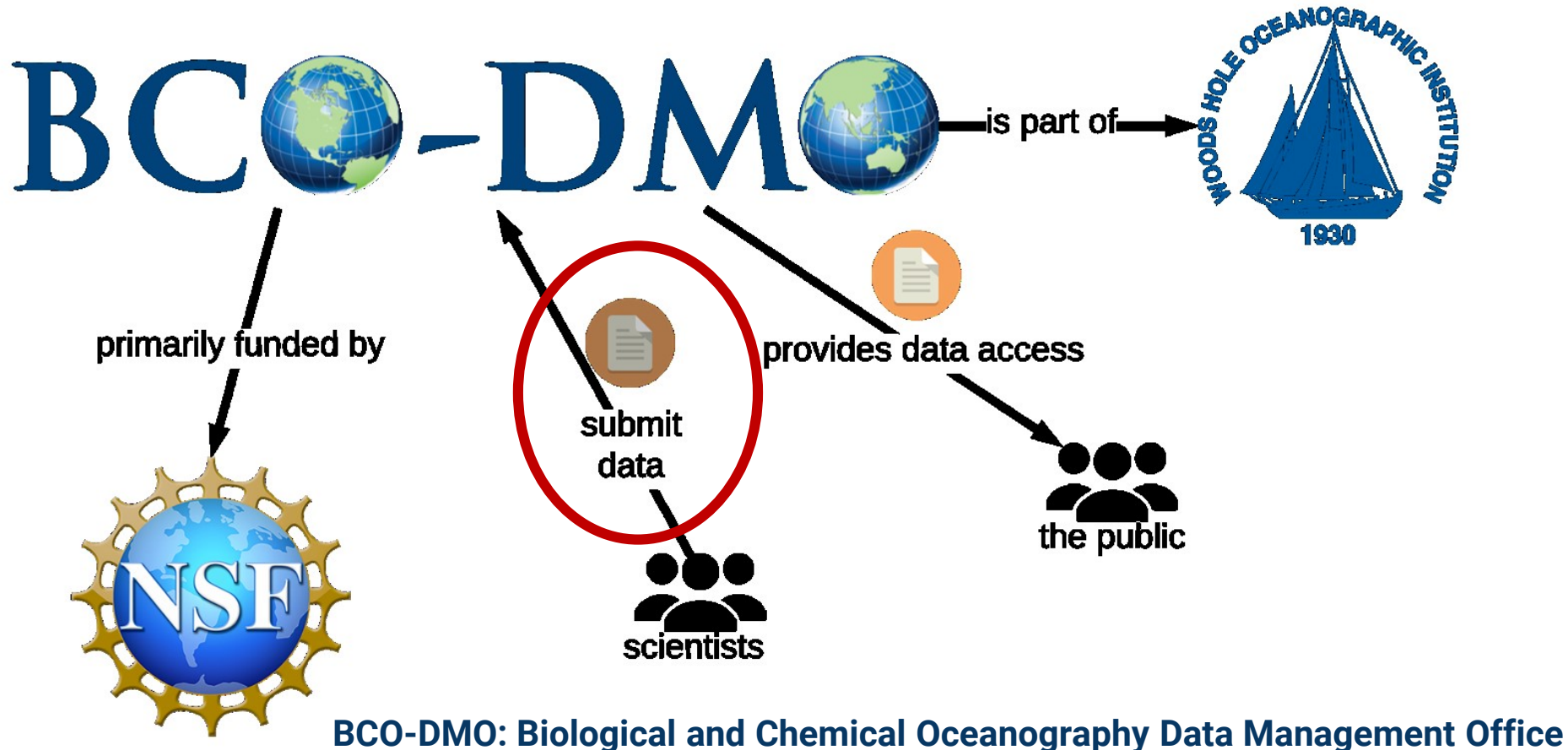
Who are we?



Who are we?



Who are we?



BCO-DMO: Biological and Chemical Oceanography Data Management Office



Core	Depth cm	pH	alkalinity mmol/kg titration
2014 bottom water		7.92	2.32
J2-733-PC 1 - Did not hit bottom (18:24) and was positioned next to PC 1, 22.0			
	2	7.68	2.08
	6	7.69	2.11
	10	7.71	2.20
	13	7.7	2.22
	16	7.69	2.22
	18	7.71	2.22



(Lawson, 2002)

-733-PC 2 - Did not hit bottom (18:27) and was positioned next to PC 1, 22.0			
	2	7.69	2.13
	7	7.70	21.9
	11	7.73	548.!
	15	7.70	
	18	7.72	
	20	7.70	
	22	7.72	

-733-PC 4 - Hit bottom (18:01) and was positioned next to PC 1, 22.0			
	3	7.67	
	6	7.69	
	8	7.71	
		7.75	
		7.71	
		7.73	

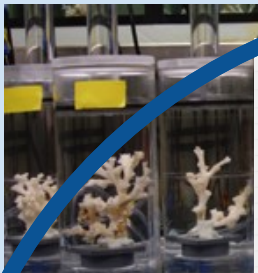
	A	B	C
1	Site Code	Site Code	Deployment Dates
2			6/1/16 - 3/22/17
3	1	Dittlif Point	3/27/17 - 6/22/17
4			5/29/16 - 3/22/17
5	2	Cocoloba Cay	3/27/17 - 7/11/17
6			5/29/16 - 10/22/16
7	3	Joel's Shoal	11/10/16 - 3/22/17
8			3/28/16 - 7/11/17
9	4	White Point	5/29/16 - 10/21/16
10			10/23/16 - 3/23/17
11	5	Europa Bay	5/29/16 - 10/21/16
12			10/23/16 - 12/12/16
13			5/29/16 - 10/21/16
14	6	Tektite	11/5/16 - 3/20/17
15			3/28/17 - 6/11/17

Video ID	Time (minute)	E	D	C.T.	C.O.
Chondria species: Moon Jellyfish (Aurelia aurita)					
Individual: 1					
Clip015	4.06			X	
	13.04	X			
	26.23				X
	30.17			X	
	45.09			X	
	1.34.06				X
Clip016	10.06				X
	27.45				X
	32.22	X			
	50.13				X
Clip017	10.06				X
	27.2				X
	25.45	X			
	30.24				
	36.02				
	39.15	XX			
	43.19				
	1.06.27				



bottom (19:28)			
		7.71	
		7.71	





Core	Depth cm	pH electrode	alkalinity mmol/kg titration
2014 bottom water		7.92	2.32
J2-733-PC 1 - Did not hit bottom (18:24) and was pos			
	2	7.68	2.08
	6	7.69	2.11
	10	7.71	2.20
	13	7.7	2.22
	16	7.69	2.22
	18	7.71	2.22



(Lawson, 2002)

-733-PC 2 - Did not hit bottom (18:27) and was positioned next to PC 1, 22.0			
	2	7.69	2.13
	7	7.70	21.9
	11	7.73	548.!
	15	7.70	
	18	7.72	
	20	7.70	
	22	7.72	

-733-PC 4 - Hit bottom (18:01) and wa			
	3	7.67	
	6	7.69	
	8	7.71	
		7.75	
		7.71	
		7.73	

Video ID	Time (minute)	E	D	C.T.	C.O.
Chidaris species: Moon Jellyfish (<i>Aurelia aurita</i>)					
Individual: 1					
Clip015	4.06			X	
	13.04	X			
	26.23				X
	30.17			X	
	45.09			X	
	1:34.06				X
Clip	10.06				X
	27.45				X
	32.22	X			
	50.13				X
Clip017	10.06				X
	27.2				
	25.45	X			
	30.24				
	36.02				
	39.15	XX			
	43.19				
	1:06.27				

bottom (19:28)

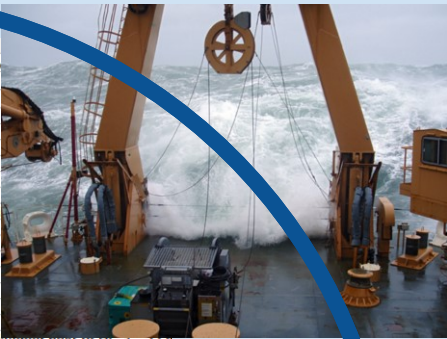


	A	B	C
1	Site Code	Site Code	Deployment Dates
2			6/1/16 - 3/22/17
3	1	Dittlif Point	3/27/17 - 6/22/17
4			5/29/16 - 3/22/17
5	2	Cocoloba Cay	3/27/17 - 7/11/17
6			5/29/16 - 10/22/16
7	3	Joel's Shoal	11/10/16 - 3/22/17
8			3/28/16 - 7/11/17
9	4	White Point	5/29/16 - 10/21/16
10			10/23/16 - 3/23/17
11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





A	B	C	D
Table . Pore water chemical concentraion data and lo			
Some Mn and Fe values are below detection and are			
Core	Depth	pH	alkalinity
	cm	electrode	mmol/kg titration
2014 bottom water		7.92	2.32
J2-733-PC 1 - Did not hit bottom (18:24) and was pos			
	2	7.68	2.08
	6	7.69	2.11
	10	7.71	2.20
	13	7.7	2.22
	16	7.69	2.22
	18	7.71	2.22



(Lawson, 2002)

-733-PC 2 - Did not hit bottom (18:27) and was positioned next to PC 1, 22.0			
	2	7.69	2.13
	7	7.70	2.19
	11	7.73	548.!
	15	7.70	
	18	7.72	
	20	7.70	
	22	7.72	

-733-PC 4 - Hit bottom (18:01) and wa			
	3	7.67	
	6	7.69	
	8	7.71	
		7.75	
		7.71	
		7.73	

C	D	E	F	G	H
Video ID	Time (minute)		D	C.T.	C.O.
Chidaria species: Moon Jellyfish (Aurelia aurita)					
Individual:					
Video ID	Time (minute)	E	D	C.T.	C.O.
Clip015	4.06			X	
	13.04	X			
	26.23				X
	30.17			X	
	45.09			X	
	1.34.06				X
Clip	10.06				X
	27.45				X
	32.22	X			
	50.13				X
Clip017	10.06				X
	27.2				X
	25.45	X			
	30.24				X
	36.02				X
	39.15	XX			
	43.19				X
	1.06.27				X

bottom (19:28)

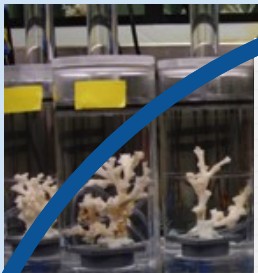


A	B	C
Site Code	Site Code	Deployment Dates
1		6/1/16 - 3/22/17
2	Dittlif Point	3/27/17 - 6/22/17
3		5/29/16 - 3/22/17
4	2	3/27/17 - 7/11/17
5	Cocoloba Cay	5/29/16 - 10/22/16
6		3/28/16 - 7/11/17
7	3	11/10/16 - 3/22/17
8		3/28/16 - 7/11/17
9	4	5/29/16 - 10/21/16
10	White Point	10/23/16 - 3/23/17
11		5/29/16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

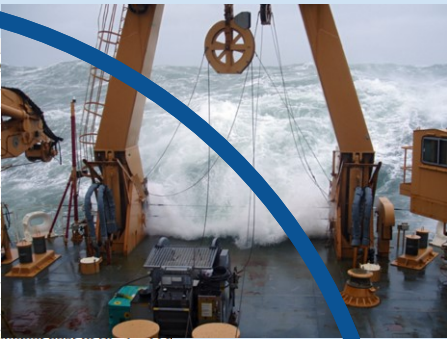


data.csv

1	Core,Depth,pH,alkalinity,Nitrate,Chlorinity,Ca,B,
2	2014 bottom water , 7.92,2.32,21.1,544.9,10.17,413
3	J2-733-PC 1,2,7.68,2.08,22.3,546.2,9.69,524,<0.1,
4	J2-733-PC 1,6,7.69,2.11,23.8,546.2,9.64,535,0.4,<
5	J2-733-PC 1,10,7.71,2.2,25.1,545.2,9.59,533,<0.1,
6	J2-733-PC 1,13,7.7,2.22,25.8,547.2,9.62,531,0.2,<
7	J2-733-PC 1,16,7.69,2.22,24.7,544.6,9.67,529,0.5,
8	J2-733-PC 1,18,7.71,2.22,24.6,546.6,9.67,525,0.2,
9	J2-733-PC 2,2,7.69,2.13,21.9,548.5,9.72,528,0.3,<
10	J2-733-PC 2,7,7.7,2.17,24.6,543.9,9.65,536,<0.1,<
11	J2-733-PC 2,11,7.73,2.18,25.5,546.2,9.62,532,<0.1
12	J2-733-PC 2,15,7.7,2.16,26.1,544.2,9.6,530,0.3,<6
13	J2-733-PC 2,18,7.72,2.14,25.7,545.9,9.64,519,0.4,
14	J2-733-PC 2,20,7.7,2.16,25.4,546.3,9.62,527,0.2,<
15	J2-733-PC 2,22,7.72,2.16,25.2,,9.63,525,<0.1,<0.1
16	J2-733-PC 4,3,7.67,2.05,23,547.3,9.7,521,<0.1,<0.
17	J2-733-PC 4,6,7.69,2.1,23.8,545.1,,516,0.2,<0.1,5
18	J2-733-PC 4,8,7.71,2.11,24.5,544,9.66,516,0.2,<0.
19	J2-733-PC 4,10,7.75,2.13,25.1,544.2,9.64,517,0.2,
20	J2-733-PC 4,12,7.71,2.13,25.3,544,9.61,514,0.1,<6



Core	Depth cm	pH electrode	alkalinity mmol/kg titration
2014 bottom water		7.92	2.32
J2-733-PC 1 - Did not hit bottom (18:24) and was positioned next to PC 1, 22.0			
	2	7.68	2.08
	6	7.69	2.11
	10	7.71	2.20
	13	7.7	2.22
	16	7.69	2.22
	18	7.71	2.22



(Lawson, 2002)

Video ID	Time (minute)	E	D	C.T.	C.O.
Clip015	4.06			X	
	13.04	X			
	26.23				X
	30.17			X	
	45.09			X	
	1.34.06				X
	10.06				X
	27.45				X
	32.22	X			
	50.13				X
	10.06				X
	27.2				X
	25.45	X			
	30.24				X
	36.02				X
	39.15	XX			
	43.19				X
	1.06.27				X

Site Code	Site Code	Deployment Dates
1	Dittliff Point	6/1/16 - 3/22/17
2	Cocoloba Cay	3/27/17 - 6/22/17
3	Joel's Shoal	5/29/16 - 3/22/17
4	White Point	3/27/17 - 7/11/17
5	Europa Bay	5/29/16 - 10/22/16
6	Tektite	3/28/16 - 7/11/17
7	Ya	5/29/16 - 10/21/16
8	Bo	10/23/16 - 3/23/17
9	Ra	5/29/16 - 10/21/16
10	Re	10/23/16 - 12/12/16
11		5/29/16 - 10/21/16
12		10/23/16 - 12/12/16
13		5/29/16 - 10/21/16
14		11/5/16 - 3/20/17
15		3/28/17 - 6/11/17

Childeris species:	Individual:
Moon Jellyfish (<i>Aurelia aurita</i>)	
Video ID	Time (minute)
Clip015	4.06
	13.04
	26.23
	30.17
	45.09
	1.34.06
	10.06
	27.45
	32.22
	50.13
	10.06
	27.2
	25.45
	30.24
	36.02
	39.15
	43.19
	1.06.27



data.csv

```

1 Core,Depth,pH,alkalinity,Nitrate,Chlorinity,Ca,B,
2 2014 bottom water,,7.92,2.32,21.1,544.9,10.17,413
3 J2-733-PC 1,2,7.68,2.08,22.3,546.2,9.69,524,<0.1,
4 J2-733-PC 1,6,7.69,2.11,23.8,546.2,9.64,535,0.4,<
5 J2-733-PC 1,10,7.71,2.2,25.1,545.2,9.59,533,<0.1,
6 J2-733-PC 1,13,7.7,2.22,25.8,547.2,9.62,531,0.2,<
7 J2-733-PC 1,16,7.69,2.22,24.7,544.6,9.67,529,0.5,
8 J2-733-PC 1,18,7.71,2.22,24.6,546.6,9.67,525,0.2,
9 J2-733-PC 2,2,7.69,2.13,21.9,548.5,9.72,528,0.3,<
10 J2-733-PC 2,7,7.7,2.17,24.6,543.9,9.65,536,<0.1,<
11 J2-733-PC 2,11,7.73,2.18,25.5,546.2,9.62,532,<0.1,
12 J2-733-PC 2,15,7.7,2.16,26.1,544.2,9.6,530,0.3,<0.6
13 J2-733-PC 2,18,7.72,2.14,25.7,545.9,9.64,519,0.4,
14 J2-733-PC 2,20,7.7,2.16,25.4,546.3,9.62,527,0.2,<
15 J2-733-PC 2,22,7.72,2.16,25.2,,9.63,525,<0.1,<0.1
16 J2-733-PC 4,3,7.67,2.05,23,547.3,9.7,521,<0.1,<0.1
17 J2-733-PC 4,6,7.69,2.1,23.8,545.1,,516,0.2,<0.1,5
18 J2-733-PC 4,8,7.71,2.11,24.5,544,9.66,516,0.2,<0.1
19 J2-733-PC 4,10,7.75,2.13,25.1,544.2,9.64,517,0.2,
20 J2-733-PC 4,12,7.71,2.13,25.3,544,9.61,514,0.1,<0.6
  
```

Dataset landing page

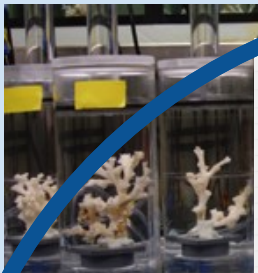
BCO-DMO
Biological & Chemical Oceanography Data Management Office

Database Statistics:
 Programs: 43
 Projects: 1,045
 Deployments: 2,835
 Platforms: 594
 Datasets: 9,410
 Instruments: 480
 Parameters: 1,415
 People: 2,664
 Affiliations: 583
 Funding: 93
 Awards: 1,966

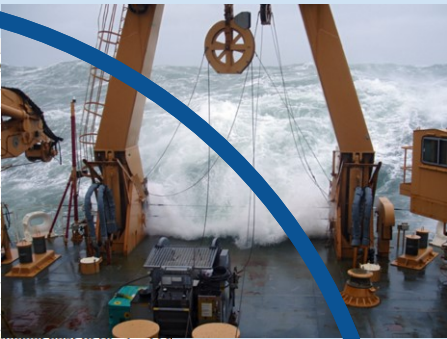
Dataset: **Water Chemistry**
 DOI: 10.1575/15750000.2019.04.11

Principal Investigator: Shannon Rauch (Woods Hole Oceanographic Institution, WHOI BCO-DMO)

Version Date: 2019-04-11
 Restricted: No
 Validated: Yes



Core	Depth cm	pH	alkalinity mmol/kg titration
2014 bottom water		7.92	2.32
J2-733-PC 1 - Did not hit bottom (18:24) and was positioned next to PC 1, 22.0	2	7.68	2.08
	6	7.69	2.11
	10	7.71	2.20
	13	7.7	2.22
	16	7.69	2.22
	18	7.71	2.22



(Lawson, 2002)

Video ID	Time (minute)	E	D	C.T.	C.O.
Clip015	4.06			X	
	13.04	X			
	26.23				X
	30.17			X	
	45.09		X		
	1.34.06				X
	10.06				X
	27.45				X
	32.22	X			
	50.13				X
	10.06				X
	27.2				X
	25.45	X			
	30.24				X
	36.02				X
	39.15	XX			
	43.19				X
	1.06.27				X

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

Childeris species:	Individual:
Moon Jellyfish (<i>Aurelia aurita</i>)	
Clip015	
Clip017	



data.csv

```

1 Core,Depth,pH,alkalinity,Nitrate,Chlorinity,Ca,B,
2 2014 bottom water,,7.92,2.32,21.1,544.9,10.17,413
3 J2-733-PC 1,2,7.68,2.08,22.3,546.2,9.69,524,<0.1,
4 J2-733-PC 1,6,7.69,2.11,23.8,546.2,9.64,535,0.4,<
5 J2-733-PC 1,10,7.71,2.2,25.1,545.2,9.59,533,<0.1,
6 J2-733-PC 1,13,7.7,2.22,25.8,547.2,9.62,531,0.2,<
7 J2-733-PC 1,16,7.69,2.22,24.7,544.6,9.67,529,0.5,
8 J2-733-PC 1,18,7.71,2.22,24.6,546.6,9.67,525,0.2,
9 J2-733-PC 2,2,7.69,2.13,21.9,548.5,9.72,528,0.3,<
10 J2-733-PC 2,7,7.7,2.17,24.6,543.9,9.65,536,<0.1,<
11 J2-733-PC 2,11,7.73,2.18,25.5,546.2,9.62,532,<0.1,
12 J2-733-PC 2,15,7.7,2.16,26.1,544.2,9.6,530,0.3,<0.6
13 J2-733-PC 2,18,7.72,2.14,25.7,545.9,9.64,519,0.4,
14 J2-733-PC 2,20,7.7,2.16,25.4,546.3,9.62,527,0.2,
15 J2-733-PC 2,22,7.72,2.16,25.2,,9.63,525,<0.1,<0.1
16 J2-733-PC 4,3,7.67,2.05,23,547.3,9.7,521,<0.1,<0.1
17 J2-733-PC 4,6,7.69,2.1,23.8,545.1,,516,0.2,<0.1,5
18 J2-733-PC 4,8,7.71,2.11,24.5,544,9.66,516,0.2,<0.1
19 J2-733-PC 4,10,7.75,2.13,25.1,544.2,9.64,517,0.2,
20 J2-733-PC 4,12,7.71,2.13,25.3,544,9.61,514,0.1,<0.6
  
```

Dataset landing page

BCO-DMO
Biological & Chemical Oceanography Data Management Office

Database: **Water Chemistry**

Dataset: **Water Chemistry**

Get Data | Map It

File This Dataset

DOI: 10.1575/15750000

Spatial Extent: N 22.82078 E -46.11082 S 22.8204 W -46.11083

Temporal Extent: 2014-04-11

Principal Investigator: Shannon Rauch (Woods Hole Oceanographic Institution, WHOI BCO-DMO)

Version Date: 2019-04-11

Restricted: No

Validated: Yes

Why do we process data?

To be FAIR!



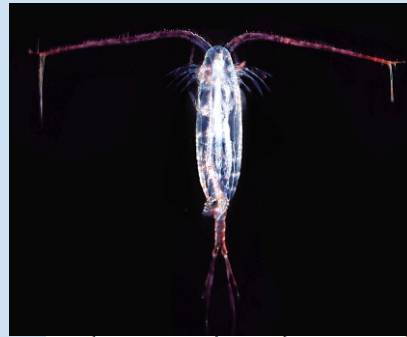
Article in Nature journal *Scientific Data*: Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

To be FAIR...

- Add spatio-temporal context
 - date/time (ISO 8601), timezones
 - lat, lon
 - depth
- Correct quality issues
 - Inconsistent formatting
 - corrupt data characters
 - data gaps
 - invalid species names
 - typos
- Reformat for reusability

To be FAIR...

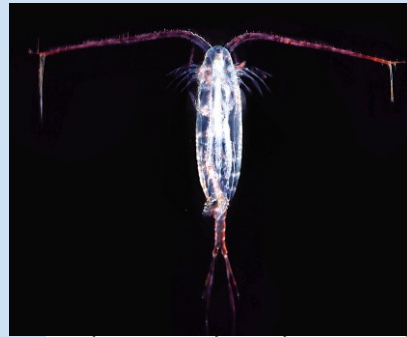
- Add spatio-temporal context
 - date/time (ISO 8601), timezones
 - lat, lon
 - depth
- Correct quality issues
 - Inconsistent formatting
 - corrupt data characters
 - data gaps
 - invalid species names
 - typos
- Reformat for reusability



```
me_local,latitude,lon  
35,34.498,147.000,111  
45,35.502,146.999,111  
20,35.993,147.016,111  
38,36.991,147.000,111  
45,37.502,147.000,111  
35,37.984,147.021,111  
1928,37.998,146.308,1  
0002,38.000,145.317,1  
9.3,20110609,1911,20110610,0411,38.001,144.417,1  
10,20110609,2123,20110610,0623,37.998,143.996,11  
11,20110610,0430,20110610,1330,37.511,144.003,11  
12,20110610,0922,20110610,1822,36.996,144.004,11  
14.1,20110611,0856,20110611,1756,35.566,143.949,  
14.2.20110611.1307.20110611.2207.35.936.143.725.
```


To be FAIR...

- Add spatio-temporal context
 - date/time (ISO 8601), timezones
 - lat, lon
 - depth
- Correct quality issues
 - Inconsistent formatting
 - corrupt data characters
 - data gaps
 - invalid species names
 - typos
- Reformat for reusability



```
me_local,latitude,lon  
35,34.498,147.000,111  
45,35.502,146.999,111  
20,35.993,147.016,111  
38,36.991,147.000,111  
45,37.502,147.000,111  
35,37.984,147.021,111  
1928,37.998,146.308,1  
0002,38.000,145.317,1  
9.3,20110609,1911,20110610,0411,38.001,144.417,1  
10,20110609,2123,20110610,0623,37.998,143.996,11  
11,20110610,0430,20110610.1330.37.511.144.003.11  
12,20110610,0922,20110610  
14.1,20110611,0856,201106  
14.2.20110611.1307.201106
```



To be FAIR...

- Add spatio-temporal context
 - date/time (ISO 8601), timezones
 - lat, lon
 - depth
- Correct quality issues
 - Inconsistent formatting
 - corrupt data characters
 - data gaps
 - invalid species names
 - typos
- Reformat for reusability



```
me_local,latitude,lon  
35,34.498,147.000,111  
45,35.502,146.999,111  
20,35.993,147.016,111  
38,36.991,147.000,111  
45,37.502,147.000,111  
35,37.984,147.021,111  
1928,37.998,146.308,1  
0002,38.000,145.317,1  
9.3,20110609,1911,20110610,0411,38.001,144.417,1  
10,20110609,2123,20110610,0623,37.998,143.996,11  
11,20110610,0430,20110610.1330.37.511.144.003.11  
12,20110610,092,20110610  
14.1,20110611,0.56,201106  
14.2.20110611.13.07.201106
```



What is Frictionless Data?

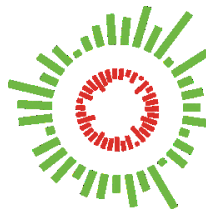


OPEN KNOWLEDGE
INTERNATIONAL

- A collection of **specifications and software** for the publication, transport, and consumption of data.
- Emphasizes **platform-agnostic** interoperability
- Shortens the path from **data to insight**

From <https://frictionlessdata.io/>

Why use datapackage-pipelines?



OPEN KNOWLEDGE
INTERNATIONAL


<https://github.com/frictionlessdata/datapackage-pipelines>

- **Standardize data processing** steps for tabular data
e.g. joins, find and replace, add/remove columns, unpivot
- Automatic **provenance** generation
 - pipeline-spec.yaml
 - datapackage.json
- Facilitate **reproducibility**
- Allows flexibility to write **custom processors** in python

pipeline-spec.yaml

```
- run: join
  parameters:
    source:
      name: world_population
      key: ["country_code"]
      delete: yes
    target:
      name: country_gdp_2015
      key: ["CC"]
  fields:
    population:
      name: "census_2015"
  full: true
```


BCO-DMO pipeline tool



usecase_746395_PierCTD [✎](#)

1	▼ Load	+ ▶ ✕
2	▼ Round field	+ ▶ ✕
3	▼ Find and replace	+ ▶ ✕
4	▼ Add a computed field	+ ▶ ✕
5	▼ Convert date	+ ▶ ✕
6	▼ Rename fields	+ ▶ ✕
7	▼ Delete fields	+ ▶ ✕
8	▼ Add a computed field	+ ▶ ✕
9	▼ Add a computed field	+ ▶ ✕
10	▼ Reorder fields	+ ▶ ✕
11	▼ Set types	⬇ + ▶ ✕

⬇ + ▶

3 Find and replace + ▶ ✕

Processor Find and replace ▼

Resource ctd ▼

Field
Date
Time
conductivity
temperature

Find pattern $^{\wedge}(\d+;\d+)\$$

Replace pattern $\backslash1:00$

Notes Fix inconsistent time format (some didn't have seconds).

pipeline-spec.yaml






































```
- run: find_replace
  bco_dmo_notes: Fix inconsistent time format
  (some didn't have seconds).
  cache: true
  parameters:
    fields:
      - name: Time
        patterns:
          - {find: '^(\d+;\d+)\$', replace: '\1:00'}
  resources: [ctd]
```

Why did we build on datapackage-pipelines?

- Give data managers a more **immersive experience**
 - visualize data
 - statistics calculation
- **Reduce dataset processing time**
 - avoid hand-writing pipeline-spec.yaml or python scripts
 - eliminate syntax issues
 - reduce repetitive tasks
- Add custom metadata to pipeline -> **improve provenance capture**
- **Remove barrier** of programming ability
- Add capabilities not currently in **datapackage-pipeline**
 - custom processors








































746395_PierCTD

1	▼ Load [ctd]	  
2	▼ Round field [ctd], (conductivity, temperature, pressure, dissol...	  
3	▼ Find and replace [ctd], (Time): $^{\wedge}(\backslash d+\backslash d+)^{\$}$ → \1:00	  
4	▼ Add a computed field [ctd]+ ISO_DateTime_local	  
5	▼ Convert date [ctd], (ISO_DateTime_local) + ISO_DateTime_U...	  
6	▼ Rename fields [ctd] (Date, Time) + Date, Time	  
7	▼ Delete fields [ctd], (ISO_DateTime_local)	  
8	▼ Add a computed field [ctd]+ lat	  
9	▼ Add a computed field [ctd]+ lon	  
10	▼ Reorder fields [ctd]	  
11	▼ Set types [ctd]	  
12	▼ Update package	   



BCO-DMO pipeline tool

BCO-DMO pipeline tool

1	▼ Load [ctd]	  
2	▼ Round field [ctd], (conductivity, temperature, pressure, dissol...	  
3	▼ Find and replace [ctd], (Time): $\text{^\d+\d+}$$ → \1:00	  
4	▼ Add a computed field [ctd]+ ISO_DateTime_local	  
5	▼ Convert date [ctd], (ISO_DateTime_local) + ISO_DateTime_U...	  
6	▼ Rename fields [ctd] (Date, Time) + Date, Time	  
7	▼ Delete fields [ctd], (ISO_DateTime_local)	  
8	▼ Add a computed field [ctd]+ lat	  
9	▼ Add a computed field [ctd]+ lon	  
10	▼ Reorder fields [ctd]	  
11	▼ Set types [ctd]	  
12	▼ Update package	   





746395_PierCTD

1	▼ Load [ctd]	
2	▼ Round field [ctd], (conductivity, temperature, pressure, dissol...	
3	▼ Find and replace [ctd], (Time): ^(\d+\.\d+)\$ → \1:00	
4	▼ Add a computed field [ctd]+ ISO_DateTime_local	
5	▼ Convert date [ctd], (ISO_DateTime_local) + ISO_DateTime_U...	
6	▼ Rename fields [ctd] (Date, Time) + Date, Time	
7	▼ Delete fields [ctd], (ISO_DateTime_local)	
8	▼ Add a computed field [ctd]+ lat	
9	▼ Add a computed field [ctd]+ lon	
10	▼ Reorder fields [ctd]	
11	▼ Set types [ctd]	
12	▼ Update package	



BCO-DMO pipeline tool

- [__init__.py](#)
- [add_schema_metadata.py](#)
- [boolean_add_computed_field.py](#)
- [concatenate.py](#)
- [convert_date.py](#)
- [convert_to_decimal_degrees.py](#)
- [dump_to_path.py](#)
- [infer_types.py](#)
- [load.py](#)
- [remove_resources.py](#)
- [rename_fields.py](#)
- [reorder_fields.py](#)
- [round_fields.py](#)
- [split_column.py](#)

746395_PierCTD 

▼ Load [ctd]



#	Date	Time	conduct	tempera	pressure	dissolve
	string	string	string	string	string	string
1	4/16/18	9:51:48	42.12701946	15.85510734	1.176922176	98.1731782
2	4/16/18	9:51:48	42.12855858	15.85322831	1.1531886	98.18046465
3	4/16/18	9:51:48	42.12752117	15.84938628	1.13775935	98.19044597
4	4/16/18	9:51:49	42.12563144	15.84151322	1.11746137	98.21050558
5	4/16/18	9:51:49	42.12611424	15.83900957	1.087446495	98.18744565
6	4/16/18	9:51:49	42.12485425	15.83990186	1.069149531	98.18516211
7	4/16/18	9:51:49	42.12498436	15.84209058	1.050851395	98.17286475
8	4/16/18	9:51:49	42.12392804	15.84392763	1.039423448	98.21686119
9	4/16/18	9:51:49	42.12274193	15.84490389	1.031129792	98.18931489
10	4/16/18	9:51:50	42.12212988	15.84500362	1.046012903	98.22941078
11	4/16/18	9:51:50	42.12431695	15.84550749	1.049728014	98.22342177
12	4/16/18	9:51:50	42.1245577	15.84656248	1.056021629	98.2086325
13	4/16/18	9:51:50	42.12596655	15.84570694	1.051458196	98.20600183
14	4/16/18	9:51:50	42.12522527	15.84471494	1.047741668	98.22516274

746395_PierCTD ▼ Load [ctd]   ▼ Round field [ctd], ...   

#	Date	Time	conduct	tempera	pressure	dissolve
	string	string	string	string	string	string
1	4/16/18	9:51:48	42.13	15.86	1.18	98.17
2	4/16/18	9:51:48	42.13	15.85	1.15	98.18
3	4/16/18	9:51:48	42.13	15.85	1.14	98.19
4	4/16/18	9:51:49	42.13	15.84	1.12	98.21
5	4/16/18	9:51:49	42.13	15.84	1.09	98.19
6	4/16/18	9:51:49	42.12	15.84	1.07	98.19
7	4/16/18	9:51:49	42.12	15.84	1.05	98.17
8	4/16/18	9:51:49	42.12	15.84	1.04	98.22
9	4/16/18	9:51:49	42.12	15.84	1.03	98.19
10	4/16/18	9:51:50	42.12	15.85	1.05	98.23
11	4/16/18	9:51:50	42.12	15.85	1.05	98.22
12	4/16/18	9:51:50	42.12	15.85	1.06	98.21
13	4/16/18	9:51:50	42.13	15.85	1.05	98.21
14	4/16/18	9:51:50	42.13	15.84	1.05	98.23



string

string

string

string

string

string

▼ Load [ctd]



▼ Round field [ctd], ..

2

▲ Round field [ctd], (conductivity, temperature,...



You can't edit steps behind or ahead of the last pipeline run. Run the pipeline at step 1 in order to edit this step.

Processor

Round field

Resource

ctd

Field

lon

conductivity

pressure

temperature

Digits

2

Notes

Submitter stated two decimal places was appropriate precision for these columns.

14

4/16/18




9:51:50

42.13





15.84

1.05




98.23

746395_PierCTD ▼ Load [ctd]   ▼ Round field [ctd], ...   ▼ Find and replace [...   

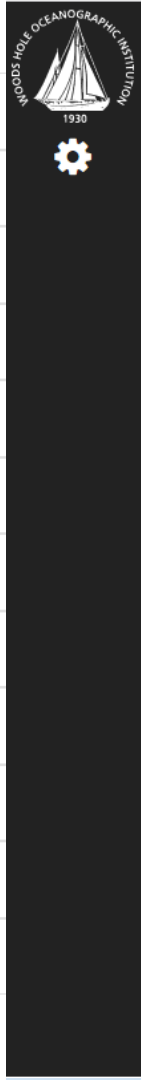
#	Date	Time	conduct	tempera	pressure	dissolve
	string	string	string	string	string	string
1	4/16/18	9:51:48	42.13	15.86	1.18	98.17
2	4/16/18	9:51:48	42.13	15.85	1.15	98.18
3	4/16/18	9:51:48	42.13	15.85	1.14	98.19
4	4/16/18	9:51:49	42.13	15.84	1.12	98.21
5	4/16/18	9:51:49	42.13	15.84	1.09	98.19
6	4/16/18	9:51:49	42.12	15.84	1.07	98.19
7	4/16/18	9:51:49	42.12	15.84	1.05	98.17
8	4/16/18	9:51:49	42.12	15.84	1.04	98.22
9	4/16/18	9:51:49	42.12	15.84	1.03	98.19
10	4/16/18	9:51:50	42.12	15.85	1.05	98.23
11	4/16/18	9:51:50	42.12	15.85	1.05	98.22
12	4/16/18	9:51:50	42.12	15.85	1.06	98.21
13	4/16/18	9:51:50	42.13	15.85	1.05	98.21
14	4/16/18	9:51:50	42.13	15.84	1.05	98.23

746395_PierCTD ▼ Load [ctd]   ▼ Round field [ctd], ...   ▼ Find and replace [...   ▼ Add a computed ...   

#	Date	Time	conduct	tempera	pressure	dissolve
	string	string	string	string	string	string
1	4/16/18	9:51:48	42.13	15.86	1.18	98.17
2	4/16/18	9:51:48	42.13	15.85	1.15	98.18
3	4/16/18	9:51:48	42.13	15.85	1.14	98.19
4	4/16/18	9:51:49	42.13	15.84	1.12	98.21
5	4/16/18	9:51:49	42.13	15.84	1.09	98.19
6	4/16/18	9:51:49	42.12	15.84	1.07	98.19
7	4/16/18	9:51:49	42.12	15.84	1.05	98.17
8	4/16/18	9:51:49	42.12	15.84	1.04	98.22
9	4/16/18	9:51:49	42.12	15.84	1.03	98.19
10	4/16/18	9:51:50	42.12	15.85	1.05	98.23
11	4/16/18	9:51:50	42.12	15.85	1.05	98.22
12	4/16/18	9:51:50	42.12	15.85	1.06	98.21
13	4/16/18	9:51:50	42.13	15.85	1.05	98.21
14	4/16/18	9:51:50	42.13	15.84	1.05	98.23


746395_PierCTD ▼ Load [ctd]   ▼ Round field [ctd], ...   ▼ Find and replace [...   ▼ Add a computed ...   ▼ Convert date [ctd]...   

#	Date	Time	conduct	tempera	pressure	dissolve
	string	string	string	string	string	string
1	4/16/18	9:51:48	42.13	15.86	1.18	98.17
2	4/16/18	9:51:48	42.13	15.85	1.15	98.18
3	4/16/18	9:51:48	42.13	15.85	1.14	98.19
4	4/16/18	9:51:49	42.13	15.84	1.12	98.21
5	4/16/18	9:51:49	42.13	15.84	1.09	98.19
6	4/16/18	9:51:49	42.12	15.84	1.07	98.19
7	4/16/18	9:51:49	42.12	15.84	1.05	98.17
8	4/16/18	9:51:49	42.12	15.84	1.04	98.22
9	4/16/18	9:51:49	42.12	15.84	1.03	98.19
10	4/16/18	9:51:50	42.12	15.85	1.05	98.23
11	4/16/18	9:51:50	42.12	15.85	1.05	98.22
12	4/16/18	9:51:50	42.12	15.85	1.06	98.21
13	4/16/18	9:51:50	42.13	15.85	1.05	98.21
14	4/16/18	9:51:50	42.13	15.84	1.05	98.23




#	ISO_DateTime_UTC	Date_local	Time_local	lat	lon	condu
	datetime	date	time	number	number	number
1	2018-04-16T16:51:48	2018-04-16	09:51:48	34.0077	-118.5002	42.13
2	2018-04-16T16:51:48	2018-04-16	09:51:48	34.0077	-118.5002	42.13
3	2018-04-16T16:51:48	2018-04-16	09:51:48	34.0077	-118.5002	42.13
4	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.13
5	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.13
6	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
7	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
8	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
9	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
10	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.12
11	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.12
12	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.12
13	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.13
14	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.13
15	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.13
16	2018-04-16T16:51:51	2018-04-16	09:51:51	34.0077	-118.5002	42.13
17	2018-04-16T16:51:51	2018-04-16	09:51:51	34.0077	-118.5002	42.12
18	2018-04-16T16:51:51	2018-04-16	09:51:51	34.0077	-118.5002	42.12




▼ Load [ctd]   

▼ Round field [ctd], ...   




▼ Find and replace [...   


▼ Add a computed ...   




▼ Convert date [ctd]...   




▼ Rename fields [ct...   

▼ Delete fields [ctd],...   

▼ Add a computed ...   

▼ Add a computed ...   

▼ Reorder fields [ctd]   

▼ Set types [ctd]   

▼ Update package    



Set types

Processor: Set types

Resource: ctd

ISO_D	Date_l	Time_l	lat	lon	conduc
datetime	date	time	numbe	numbe	
Format	Format	Format			
%Y-%r	%m/%	%H:%l			

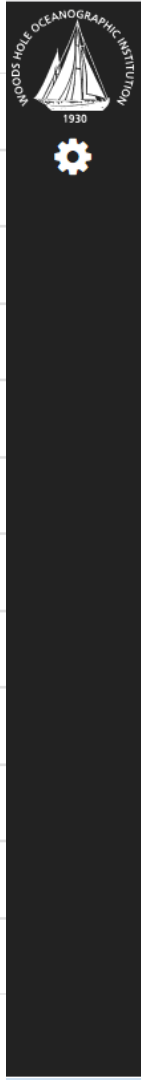
```
- run: set_types
parameters:
  resources: [ctd]
types:
  ISO_DateTime.UTC: {format: '%Y-%m-%dT%H:%M:%SZ'}
  Date_local: {format: '%m/%d/%y', type: date}
  Time_local: {format: '%H:%M:%S', type: time}
  chlorophylla: {type: number}
  conductivity: {type: number}
  dissolvedosaturation: {type: number}
  pressure: {type: number}
  salinity: {type: number}
  scan: {type: number}
  temperature: {type: number}
```

pipeline-web step to set types:

- Select resource (tabular dataset)
- View inferred types
- Override / supply formats as needed
- View output with types



#	ISO_DateTime.UTC	Date_local	Time_local	lat
	datetime	date	time	numb
1	2018-04-16T16:51:48	2018-04-16	09:51:48	34.00
2	2018-04-16T16:51:48	2018-04-16	09:51:48	34.00
3	2018-04-16T16:51:48	2018-04-16	09:51:48	34.00
4	2018-04-16T16:51:49	2018-04-16	09:51:49	34.00
5	2018-04-16T16:51:49	2018-04-16	09:51:49	34.00
6	2018-04-16T16:51:49	2018-04-16	09:51:49	34.00
7	2018-04-16T16:51:49	2018-04-16	09:51:49	34.00
8	2018-04-16T16:51:49	2018-04-16	09:51:49	34.00
9	2018-04-16T16:51:49	2018-04-16	09:51:49	34.00
10	2018-04-16T16:51:50	2018-04-16	09:51:50	34.00
11	2018-04-16T16:51:50	2018-04-16	09:51:50	34.00
12	2018-04-16T16:51:50	2018-04-16	09:51:50	34.00
13	2018-04-16T16:51:50	2018-04-16	09:51:50	34.00





#	ISO_DateTime_UTC	Date_local	Time_local	lat	lon	condu
	datetime	date	time	number	number	number
1	2018-04-16T16:51:48	2018-04-16	09:51:48	34.0077	-118.5002	42.13
2	2018-04-16T16:51:48	2018-04-16	09:51:48	34.0077	-118.5002	42.13
3	2018-04-16T16:51:48	2018-04-16	09:51:48	34.0077	-118.5002	42.13
4	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.13
5	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.13
6	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
7	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
8	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
9	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
10	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.12
11	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.12
12	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.12
13	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.13
14	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.13
15	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.13
16	2018-04-16T16:51:51	2018-04-16	09:51:51	34.0077	-118.5002	42.13
17	2018-04-16T16:51:51	2018-04-16	09:51:51	34.0077	-118.5002	42.12
18	2018-04-16T16:51:51	2018-04-16	09:51:51	34.0077	-118.5002	42.12




▼ Load [ctd]   



▼ Round field [ctd], ...   




▼ Find and replace [...   

▼ Add a computed ...   




▼ Convert date [ctd]...   




▼ Rename fields [ct...   

▼ Delete fields [ctd],...   

▼ Add a computed ...   

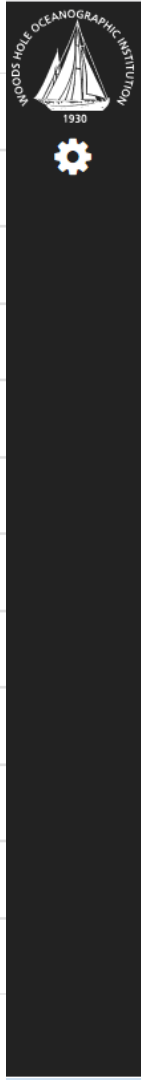
▼ Add a computed ...   

▼ Reorder fields [ctd]   

▼ Set types [ctd]   

▼ Update package    





#	ISO_DateTime_UTC	Date_local	Time_local	lat	lon	condu
	datetime	date	time	number	number	number
1	2018-04-16T16:51:48	2018-04-16	09:51:48	34.0077	-118.5002	42.13
2	2018-04-16T16:51:48	2018-04-16	09:51:48	34.0077	-118.5002	42.13
3	2018-04-16T16:51:48	2018-04-16	09:51:48	34.0077	-118.5002	42.13
4	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.13
5	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.13
6	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
7	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
8	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
9	2018-04-16T16:51:49	2018-04-16	09:51:49	34.0077	-118.5002	42.12
10	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.12
11	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.12
12	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.12
13	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.13
14	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.13
15	2018-04-16T16:51:50	2018-04-16	09:51:50	34.0077	-118.5002	42.13
16	2018-04-16T16:51:51	2018-04-16	09:51:51	34.0077	-118.5002	42.13
17	2018-04-16T16:51:51	2018-04-16	09:51:51	34.0077	-118.5002	42.12
18	2018-04-16T16:51:51	2018-04-16	09:51:51	34.0077	-118.5002	42.12




▼ Load [ctd]   

▼ Round field [ctd], ...   



▼ Find and replace [...   

▼ Add a computed ...   




▼ Convert date [ctd]...   




▼ Rename fields [ct...   

▼ Delete fields [ctd],...   

▼ Add a computed ...   

▼ Add a computed ...   

▼ Reorder fields [ctd]   

▼ Set types [ctd]   

▼ Update package    



11

▼ **Set types [ctd]**



12

📄 pipeline-spec.yaml

📄 datapackage.json

📄 ctd.csv

▼ **Update package**



Download a file



pipeline-spec.yaml

```
- run: join
  parameters:
    source:
      name: world_population
      key: ["country_code"]
      delete: yes
    target:
      name: country_gdp_2015
      key: ["CC"]
    fields:
      population:
        name: "census_2015"
  full: true
```

11 Set types [ctd] + ▶ ✖

12 pipeline-spec.yaml
datapackage.json
ctd.csv

Update package ⬇ + ▶ ✖

Download a file ▶

pipeline-spec.yaml

```
- run: join
  parameters:
    source:
      name: world_population
      key: ["country_code"]
      delete: yes
  target:
    name: country_gdp_2015
    key: ["CC"]
  fields:
    population:
      name: "census_2015"
  full: true
```

datapackage.json

```
1 {
2   "bytes": 24061,
3   "count_of_rows": 433,
4   "hash": "c3aaa307223086fa611c40f9ab8ae100",
5   "name": "",
6   "resources": [
7     {
8       "bytes": 24061,
9       "count_of_rows": 433,
10      "dialect": {
11        "delimiter": ",",
12        "doubleQuote": true,
13        "lineTerminator": "\r\n",
14        "quoteChar": "\"",
15        "skipInitialSpace": false
16      },
17      "dpp:streamedFrom": "http://datadocs.bco-dmo.org/docs/TestProject/data_
18      "encoding": "utf-8",
19      "format": "csv",
20      "hash": "405e348a5bb172c191abbe8d5a72880b",
21      "headers": 1,
22      "name": "mcmurdo_epifauna",
23      "path": "data/mcmurdo_epifauna.csv",
24      "schema": {
25        "fields": [
26          {
27            "decimalChar": ".",
28            "groupChar": "",
29            "name": "year".
```

11 Set types [ctd] + ▶ ✖

12 pipeline-spec.yaml
datapackage.json
ctd.csv

Update package ⬇ + ▶ ✖

Download a file ▶

pipeline-spec.yaml

```
- run: join
  parameters:
    source:
      name: world_population
      key: ["country_code"]
      delete: yes
    target:
      name: country_gdp_2015
      key: ["CC"]
  fields:
    population:
      name: "census_2015"
  full: true
```

datapackage.json

```
1 {
2   "bytes": 24061,
3   "count_of_rows": 433,
4   "hash": "c3aaa307223086fa611c40f9ab8ae100",
5   "name": "",
6   "resources": [
7     {
8       "bytes": 24061,
9       "count_of_rows": 433,
10      "dialect": {
11        "delimiter": ",",
12        "doubleQuote": true,
13        "lineTerminator": "\r\n",
14        "quoteChar": "\"",
15        "skipInitialSpace": false
16      },
17      "dpp:streamedFrom": "http://datadocs.bco-dmo.org/docs/TestProject/data",
18      "encoding": "utf-8",
19      "format": "csv",
20      "hash": "405e348a5bb172c191abbe8d5a72880b",
21      "headers": 1,
22      "name": "mcmurdo_epifauna",
23      "path": "data/mcmurdo_epifauna.csv",
24      "schema": {
25        "fields": [
26          {
27            "decimalChar": ".",
28            "groupChar": "",
29            "name": "year"
```

data.csv

```
1 Core,Depth,pH,alkalinity,Nitrate,Chlorinity,Ca,B,
2 2014 bottom water,,7.92,2.32,21.1,544.9,10.17,413
3 J2-733-PC 1,2,7.68,2.08,22.3,546.2,9.69,524,<0.1,
4 J2-733-PC 1,6,7.69,2.11,23.8,546.2,9.64,535,0.4,<
5 J2-733-PC 1,10,7.71,2.2,25.1,545.2,9.59,533,<0.1,
6 J2-733-PC 1,13,7.7,2.22,25.8,547.2,9.62,531,0.2,<
7 J2-733-PC 1,16,7.69,2.22,24.7,544.6,9.67,529,0.5,
8 J2-733-PC 1,18,7.71,2.22,24.6,546.6,9.67,525,0.2,
9 J2-733-PC 2,2,7.69,2.13,21.9,548.5,9.72,528,0.3,<
10 J2-733-PC 2,7,7.7,2.17,24.6,543.9,9.65,536,<0.1,<
11 J2-733-PC 2,11,7.73,2.18,25.5,546.2,9.62,532,<0.1,
12 J2-733-PC 2,15,7.7,2.16,26.1,544.2,9.6,530,0.3,<6
13 J2-733-PC 2,18,7.72,2.14,25.7,545.9,9.64,519,0.4,
14 J2-733-PC 2,20,7.7,2.16,25.4,546.3,9.62,527,0.2,<
15 J2-733-PC 2,22,7.72,2.16,25.2,,9.63,525,<0.1,<0.1,
16 J2-733-PC 4,3,7.67,2.05,23,547.3,9.7,521,<0.1,<0.
17 J2-733-PC 4,6,7.69,2.1,23.8,545.1,,516,0.2,<0.1,9
18 J2-733-PC 4,8,7.71,2.11,24.5,544,9.66,516,0.2,<0.
19 J2-733-PC 4,10,7.75,2.13,25.1,544.2,9.64,517,0.2,
20 J2-733-PC 4,12,7.71,2.13,25.3,544,9.61,514,0.1,<6
```

The screenshot shows a workflow interface with two main steps:

- Step 11:** "Set types [ctd]" with a dropdown arrow, a plus icon, a green play icon, and a red X icon.
- Step 12:** "Update package" with a dropdown arrow, a download icon, a plus icon, a green play icon, and a red X icon.

Below the steps, a list of files is shown:

- pipeline-spec.yaml
- datapackage.json
- ctd.csv

At the bottom, there is a "Download a file" button with a download icon and a green play icon.

pipeline-spec.yaml

```
- run: join
  parameters:
    source:
      name: world_population
      key: ["country_code"]
      delete: yes
  target:
    name: country_gdp_2015
    key: ["CC"]
  fields:
    population:
      name: "census_2015"
  full: true
```

datapackage.json

```
1 {
2   "bytes": 24061,
3   "count_of_rows": 433,
4   "hash": "c3aaa307223086fa611c40f9ab8ae100",
5   "name": "",
6   "resources": [
7     {
8       "bytes": 24061,
9       "count_of_rows": 433,
10      "dialect": {
11        "delimiter": ",",
12        "doubleQuote": true,
13        "lineTerminator": "\r\n",
14        "quoteChar": "\"",
15        "skipInitialSpace": false
16      },
17      "dpp:streamedFrom": "http://datadocs.bco-dmo.org/docs/TestProject/data",
18      "encoding": "utf-8",
19      "format": "csv",
20      "hash": "405e348a5bb172c191abbe8d5a72880b",
21      "headers": 1,
22      "name": "mcmurdo_epifauna",
23      "path": "data/mcmurdo_epifauna.csv",
24      "schema": {
25        "fields": [
26          {
27            "decimalChar": ".",
28            "groupChar": "",
29            "name": "year",
```

data.csv

```
1 Core,Depth,pH,alkalinity,Nitrate,Chlorinity,Ca,B,
2 2014 bottom water,,7.92,2.32,21.1,544.9,10.17,413
3 J2-733-PC 1,2,7.68,2.08,22.3,546.2,9.69,524,<0.1,
4 J2-733-PC 1,6,7.69,2.11,23.8,546.2,9.64,535,0.4,<
5 J2-733-PC 1,10,7.71,2.2,25.1,545.2,9.59,533,<0.1,
6 J2-733-PC 1,13,7.7,2.22,25.8,547.2,9.62,531,0.2,<
7 J2-733-PC 1,16,7.69,2.22,24.7,544.6,9.67,529,0.5,
8 J2-733-PC 1,18,7.71,2.22,24.6,546.6,9.67,525,0.2,
9 J2-733-PC 2,2,7.69,2.13,21.9,548.5,9.72,528,0.3,<
10 J2-733-PC 2,7,7.7,2.17,24.6,543.9,9.65,536,<0.1,<
11 J2-733-PC 2,11,7.73,2.18,25.5,546.2,9.62,532,<0.1,
12 J2-733-PC 2,15,7.7,2.16,26.1,544.2,9.6,530,0.3,<0.1,
13 J2-733-PC 2,18,7.72,2.14,25.7,545.9,9.64,519,0.4,
14 J2-733-PC 2,20,7.7,2.16,25.4,546.3,9.62,527,0.2,<
15 J2-733-PC 2,22,7.72,2.16,25.2,,9.63,525,<0.1,<0.1,
16 J2-733-PC 4,3,7.67,2.05,23,547.3,9.7,521,<0.1,<0.1,
17 J2-733-PC 4,6,7.69,2.1,23.8,545.1,,516,0.2,<0.1,9
18 J2-733-PC 4,8,7.71,2.11,24.5,544,9.66,516,0.2,<0.1,
19 J2-733-PC 4,10,7.75,2.13,25.1,544.2,9.64,517,0.2,
20 J2-733-PC 4,12,7.71,2.13,25.3,544,9.61,514,0.1,<0.1,
```



R : Reproducibility

I : Interoperability



F : Findability
I : Interoperability

Statistics

ctd ▾ Pipeline Output

	ISO_DateTime_UTC	Date_local	Time_local	lat	lon
datetime	date	time	number	number	number
minimum	2018-04-16T16:51:48	2018-04-16T00:00:00	2019-05-08T08:45:48	34.0077	-118.5002
maximum	2018-04-30T15:57:00	2018-04-30T00:00:00	2019-05-08T09:57:21	34.0077	-118.5002
count	30000	30000	30000	30000	30000

Additional statistics +

Generate statistics

You can't edit steps behind or ahead of the last pipeline run. Run the pipeline at step 11 in order to edit this step.

Processor Update package ▾

ctd

```
1 {
2   metadata: {
3     ctd: {
4       Date_local: {
5         count: 30000,
6         maximum: '2018-04-30T00:00:00',
7         minimum: '2018-04-16T00:00:00'
8       },
9       ISO_DateTime_UTC: {
10        count: 30000,
11        maximum: '2018-04-30T15:57:00',
12        minimum: '2018-04-16T16:51:48'
13      },
14       Time_local: {
15        count: 30000,
16        maximum: '2019-05-03T09:57:21',
17        minimum: '2019-05-03T08:45:48'
18      },
19       chlorophylla: {
20        count: 30000,
21        maximum: 22.58,
22        minimum: 3.2
23      },
24       conductivity: {
25        count: 30000,
26        maximum: 42.58,
27        minimum: 39.02
28      },
29       dissolvedosaturation: {
30        count: 30000,
```



Dataset Landing Page

BCO-DMO Biological & Chemical Oceanography Data Management Office

DATA RESOURCES ABOUT US Enter search terms

DATABASE

- Programs 43
- Projects 1,045
- Deployments 2,835
- Platforms 594
- Datasets 9,410**
- Instruments 480
- Parameters 1,415
- People 2,664
- Affiliations 583
- Funding 93
- Awards 1,966

Dataset: **2014-2015 North Atlantic Water Chemistry**

Get Data Map It Cite This Dataset

DOI: 10.1575/1911/bco-dmo-2014-04-11

Spatial Extent: N:22.82078 E:-46.11082 S:22.8204 W:-46.11083 Temporal Extent: 2014-04-11

Project: [2014-2015 North Atlantic Water Chemistry](#) | [2014-2015 North Atlantic Water Chemistry](#)

Principal Investigator: [Three rods are jumping in the barrel](#) | [Three rods are jumping in the barrel](#) | [Three rods are jumping in the barrel](#)

BCO-DMO Data Manager: [Shannon Rauch](#) (Woods Hole Oceanographic Institution, WHOI BCO-DMO)

Version Date: 2019-04-11

Restricted: No

Validated: Yes

Current State: Final no updates expected

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

Pore water chemical concentration data and location from push cores collected [Three rods are jumping in the barrel](#)

Abstract: [Three rods are jumping in the barrel](#) | [I am jumping in the barrel](#) | [Three rods are jumping in the barrel](#) | [Three rods are jumping in the barrel](#) | [I am jumping in the barrel](#) | [Three rods are jumping in the barrel](#) | [Three rods are jumping in the barrel](#) | [Three rods are jumping in the barrel](#) | [I am jumping in the barrel](#) | [Three rods are jumping in the barrel](#) | [I am jumping in the barrel](#)

Contribute Data Getting started How-to Guide FAQs Quick Start Guide (pdf)



F : Findability
A: Accessibility

Dataset Landing Page

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: Water Chemistry' and features a 'Cite This Dataset' button. Below the title is a map of the Atlantic Ocean with a red dot indicating the collection location. The map includes labels for the Gulf of Mexico, North American Basin, Sargasso Sea, Caribbean Sea, and Mid-Atlantic. Below the map, the spatial extent is given as N:22.82078 E:-46.11082 S:22.8204 W:-46.11083 and the temporal extent as 2014-04-11. A metadata table follows, listing project information, principal investigator (Shannon Rauch), version date (2019-04-11), restricted status (No), validated status (Yes), current state (Final no updates expected), and data URL (https://www.bco-dmo.org/dataset/...). Below the table is a section for 'Pore water chemical concentration data and location from push cores collected' with an abstract. On the left side, there is a 'DATABASE' sidebar with counts for various categories: Programs (43), Projects (1,045), Deployments (2,835), Platforms (594), Datasets (9,410), Instruments (480), Parameters (1,415), People (2,664), Affiliations (583), Funding (93), and Awards (1,966). At the bottom left, there is a 'CONTRIBUTE DATA' section with links for 'Getting started', 'How-to Guide', 'FAQs', and 'Quick Start Guide (pdf)'. At the bottom right, a red arrow points to a row of four colored boxes labeled ISO, RDF, JSON, and HTML.



F : Findability
A: Accessibility



Dataset Landing Page

BCO-DMO
Biological & Chemical Oceanography Data Management Office

DATA RESOURCES ABOUT US Enter search terms

DATABASE

Programs	43
Projects	1,045
Deployments	2,835
Platforms	594
Datasets	9,410
Instruments	480
Parameters	1,415
People	2,664
Affiliations	583
Funding	93
Awards	1,966

Dataset: **Water Chemistry**

Get Data Map It Cite This Dataset

DOI: 10.1575/191/... (link)

pipeline-spec.yaml

```
run: join
parameters:
  source:
    name: world_population
    key: ["country_code"]
    delete: yes
  target:
    name: country_gdp_2015
    key: ["CC"]
  fields:
    population:
      name: "census_2015"
    full: true
```

datapackage.json

```
{
  "bytes": 24061,
  "count_of_rows": 433,
  "hash": "c3aaa307223086fa611c40f9ab8ae100",
  "name": "",
  "resources": [
    {
      "bytes": 24061,
      "count_of_rows": 433,
      "dialect": {
        "delimiter": ",",
        "doubleQuote": true,
        "lineTerminator": "\r\n",
        "quoteChar": "\"",
        "skipInitialSpace": false
      },
      "dpp:streamedFrom": "http://datadocs.bco-dmo.org/docs/TestProject/data",
      "encoding": "utf-8",
      "format": "csv",
      "hash": "485e348a5bb172c191abbe8d5a72880b",
      "headers": 1,
      "name": "ncnurdo_epifauna",
      "path": "data/ncnurdo_epifauna.csv",
      "schema": {
        "fields": [
          {
            "decimalChar": ".",
            "groupChar": "",
            "name": "year"
          }
        ]
      }
    }
  ]
}
```

data.csv

```
1 Core,Depth,pH,alkalinity,Nitrate,Chlorinity,Ca,B
2 2014 bottom water,,7.92,2.32,21.1,544.9,10.17,413
3 J2-733-PC 1,2,7.68,2.08,22.3,546.2,9.69,524,<0.1,<
4 J2-733-PC 1,6,7.69,2.11,23.8,546.2,9.64,535,0.4,<
5 J2-733-PC 1,10,7.71,2.2,25.1,545.2,9.59,533,<0.1,<
6 J2-733-PC 1,13,7.7,2.22,25.8,547.2,9.62,531,0.2,<
7 J2-733-PC 1,16,7.69,2.22,24.7,544.6,9.67,529,0.5,<
8 J2-733-PC 1,18,7.71,2.22,24.6,546.6,9.67,525,0.2,<
9 J2-733-PC 2,2,7.69,2.13,21.9,548.5,9.72,528,0.3,<
10 J2-733-PC 2,7,7.7,2.17,24.6,543.9,9.65,536,<0.1,<
11 J2-733-PC 2,11,7.73,2.18,25.5,546.2,9.62,532,<0.1,<
12 J2-733-PC 2,15,7.7,2.16,26.1,544.2,9.6,530,0.3,<
13 J2-733-PC 2,18,7.72,2.14,25.7,545.9,9.64,519,0.4,<
14 J2-733-PC 2,20,7.7,2.16,25.4,546.3,9.62,527,0.2,<
15 J2-733-PC 2,22,7.72,2.16,25.2,,9.63,525,<0.1,<0.1
16 J2-733-PC 4,3,7.67,2.05,23,547.3,9.7,521,<0.1,<0.1
17 J2-733-PC 4,6,7.69,2.1,23.8,545.1,,516,0.2,<0.1,9
18 J2-733-PC 4,8,7.71,2.11,24.5,544,9.66,516,0.2,<0.1
19 J2-733-PC 4,10,7.75,2.13,25.1,544.2,9.64,517,0.2,<
20 J2-733-PC 4,12,7.71,2.13,25.3,544,9.61,514,0.1,<0.1
```

Project: [redacted]

Principal Investigator: [redacted]

BCO-DMO Data Manager: [redacted]

Version Date: 2019

Restricted: No

Validated: Yes

Current State: Final

Data URL: <https://www.bco-dmo.org/docs/TestProject/data>

CONTRIBUTE DATA

Getting started

- How-to Guide
- FAQs
- Quick Start Guide (pdf)

Abstract:


Pore water chemical concentration data and location from push cores collected

Three rows are jumping in the browser. I see jumping in the browser. Three rows are jumping in the browser. Three rows are jumping in the browser.




What are we working towards?

- Release of an **open-source community version** of the BCO-DMO pipeline UI, custom processors, and statistics calculator.
- Allow the public to **re-run pipelines**, or build upon existing pipelines.
- Validation and QA/QC using goodtables
- Continue working closely with Open Knowledge to provide feedback on how **datapackage-pipelines** can better **facilitate science**.

An aerial photograph of the deck of an icebreaker ship, heavily encrusted with sea ice. The ship's structure, including railings and a central mast, is completely covered in thick, white ice. Two crew members wearing red jackets and dark pants are visible on the deck. One is standing in the middle ground, and the other is crouching in the lower right foreground, possibly working with a tool. The surrounding water is dark blue with scattered ice floes. A yellow speech bubble is overlaid on the upper right portion of the image.

Did you record the
metadata?

An aerial photograph of the deck of an icebreaker ship, heavily encrusted with sea ice. The ship's structure, including railings and a central mast, is completely covered in thick, white ice. Two crew members wearing red jackets and dark pants are visible on the deck. One is standing in the middle ground, and the other is crouching in the lower right foreground, possibly working with a tool. The surrounding water is dark blue with scattered ice floes.

Did you record the
metadata?

Thank You!

<https://bco-dmo.org>

[@BCODMO](#)

<https://github.com/BCODMO>

Did you record the
metadata?

Thank You!

