BCO-DMO

Biological & Chemical Oceanography Data Management Office

http://bco-dmo.org

DATA ACCESS TUTORIAL
2015 OCB PI Summer Workshop

Table of Contents

BCO-DMO

Biological & Chemical Oceanography Data Management Office

http://bco-dmo.org

DATA ACCESS TUTORIAL
2015 OCB PI SUMMER WORKSHOP

How to Submit Data

BCO-DMO works with investigators to publish data from research projects funded by the NSF Geosciences Directorate (GEO) Division of Ocean Sciences (OCE) Biological and Chemical Oceanography Sections and the Division of Polar Programs (PLR) Antarctic Sciences (ANT) Organisms and Ecosystems Program. Once published, BCO-DMO also fulfills NSF requirements for long-term preservation by submitting data to the appropriate national data center for archive (e.g., NOAA's National Oceanographic Data Center).

To start the process, please visit our "How to Get Started" page:
http://www.bco-dmo.org/how-get-started

Here you will find the necessary forms to contribute metadata related to programs, projects, deployments, and datasets. These forms are available in Rich Text Format (.rtf), making them accessible using any word processing application:

Program: A large coordinated research effort, usually comprising multiple projects generating several datasets. (e.g., OCB, U.S. GEOTRACES, SEES-OA)

Project: A project usually encompasses one or more datasets, and may be part of a larger program (e.g. CoFeMUG is a project affiliated with U.S. GEOTRACES), or could be (and most are) unaffiliated with any program.

Deployment: A specific cruise, mooring, laboratory, or some other unique event of data collection and/or analysis.

Dataset: A collection of values representing scientific observations, measurements, or derivations.

Step 1: Complete Metadata Forms

Use the forms described above to submit new information to the Office.

Step 2: Prepare Data Files

We accept data in any format, however comma or tab-delimited (preferred), or Excel spreadsheets are most common. If the dataset is too large to contribute as an email attachment, contact our office via info@bco-dmo.org for instructions on the best way to contribute your data.

Step 3: Email BCO-DMO

Once completed, metadata forms and data files can be emailed as attachments to the office along with a brief communication explaining the nature of the submission, by addressing the email to: info@bco-dmo.org

You will receive a response from the Office confirming your submission and informing you of the staff assigned to your dataset. BCO-DMO staff members will work with you to publish your data and metadata.

Additional Information:

Also available on the “How To Get Started” page are useful links to additional information on the NSF OCE data policy, submitting data in spreadsheets, data archives, and more. Feel free to contact the Office with any questions or concerns you may have regarding your data submission.

Data access: TEXT-BASED SEARCH

scenario 1: You have a general idea of what you are looking for.

Go to: http://bco-dmo.org

The DATABASE links in the left navigation area provide an idea of how the data are organized. Data sets are grouped by Program (largest collection), Project (smaller in scope than a Program), Deployment (cruise, mooring, and many other examples), Dataset (a logical collection of data).

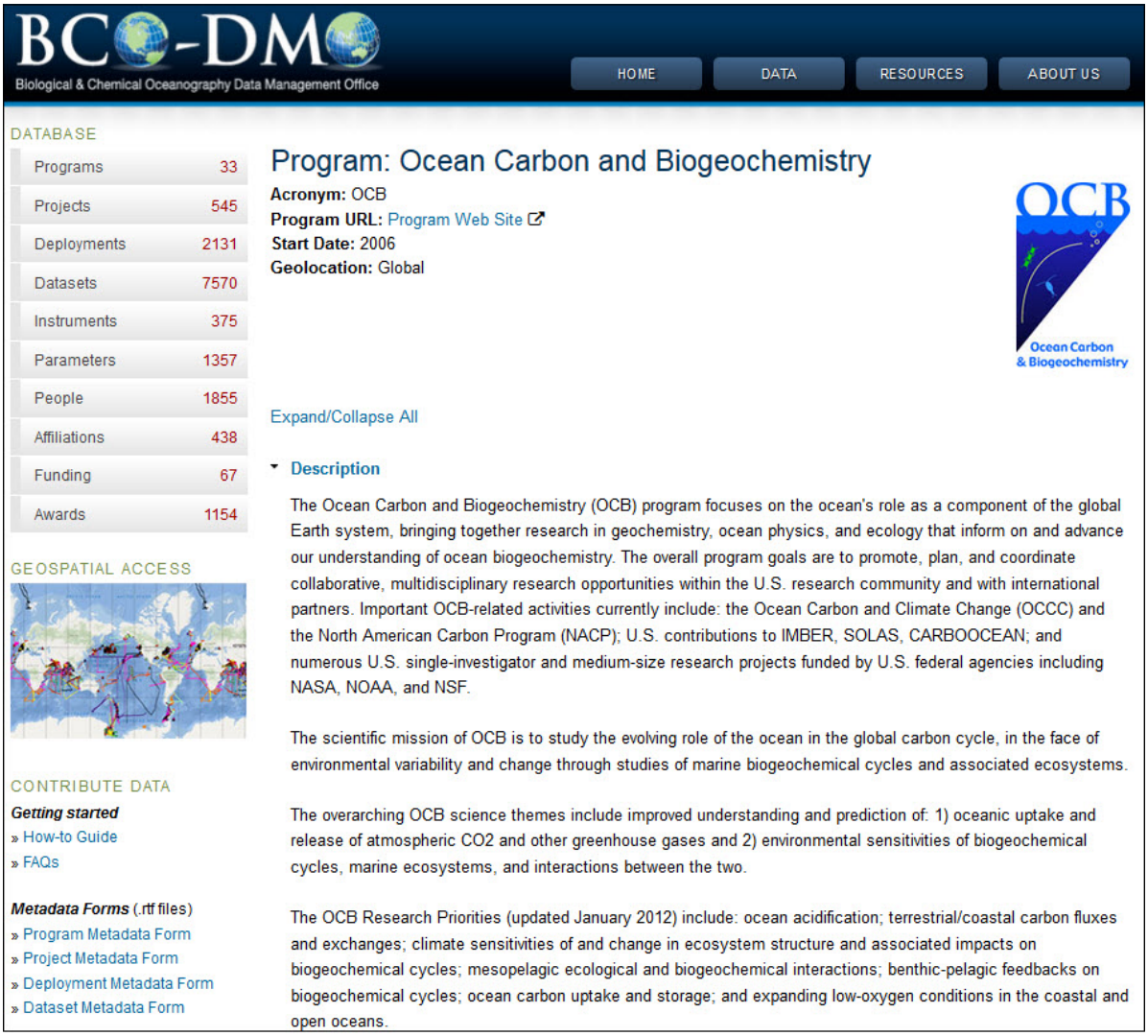
One could start with any of those navigation links and search to find the data of interest. If you know you are interested in data from an OCB project (for example, "CoFeMUG") you may start at the top level "Programs".

Select Programs from the left navigation area. Enter OCB in the search box. Click the Search button.

Select "Ocean Carbon and Biogeochemistry" from the Program list.

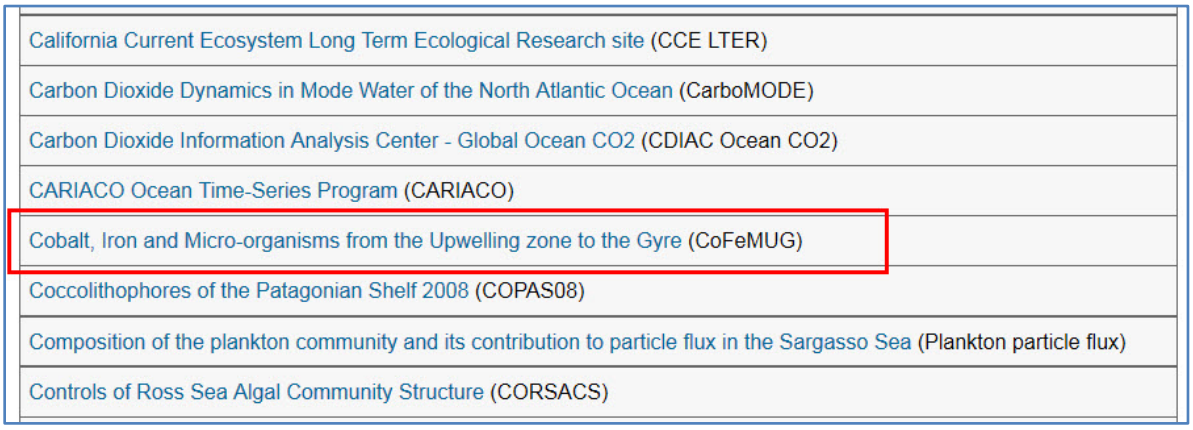
URL: http://www.bco-dmo.org/program/2015

Your browser display should look like the image on the next page, showing a brief description of the OCB program.



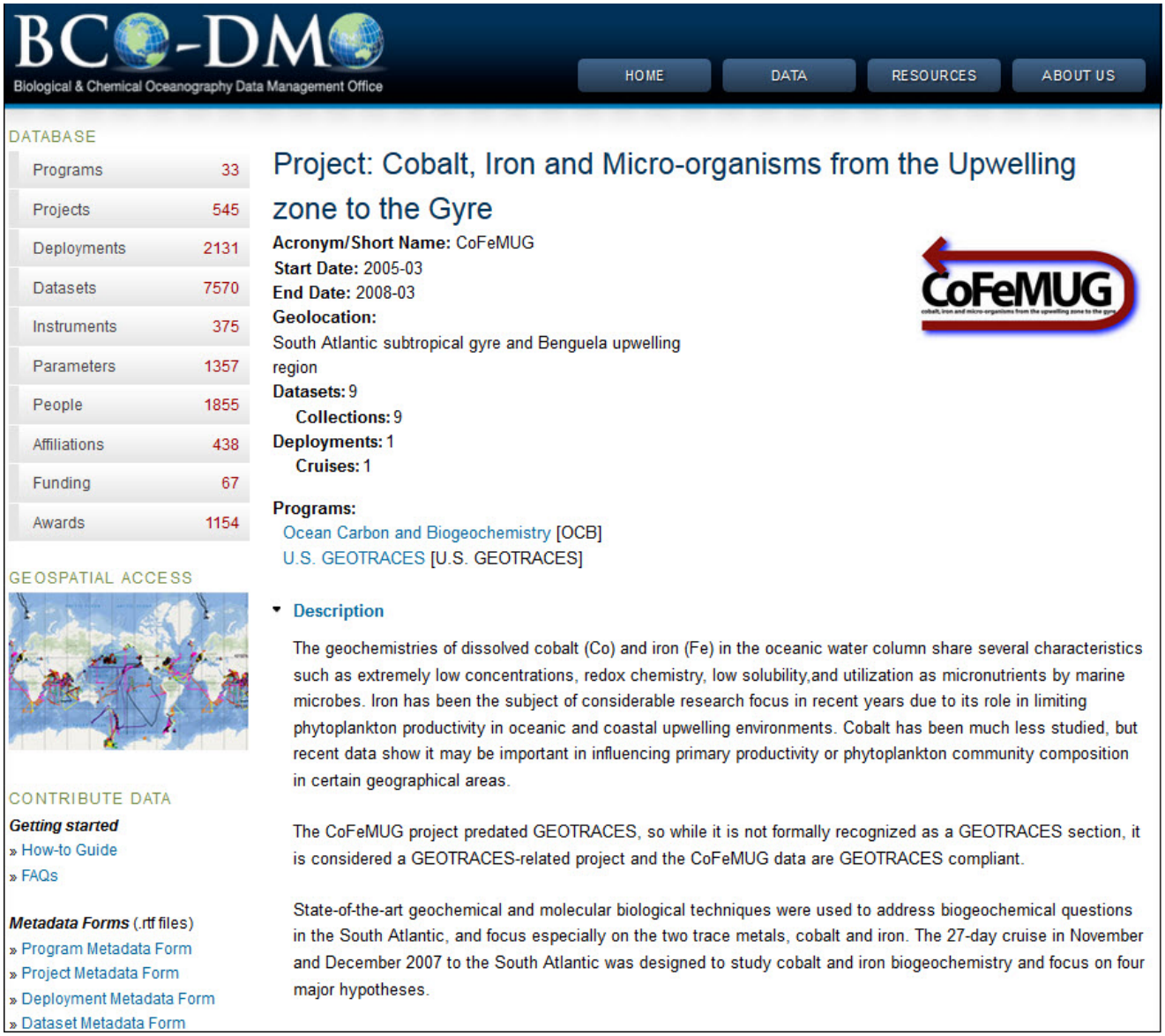
Scroll down and expand the Projects section below the “More Information” heading.

Select “Cobalt, Iron and Micro-organisms from the Upwelling zone to the Gyre” (CoFeMUG) from the list of project names.



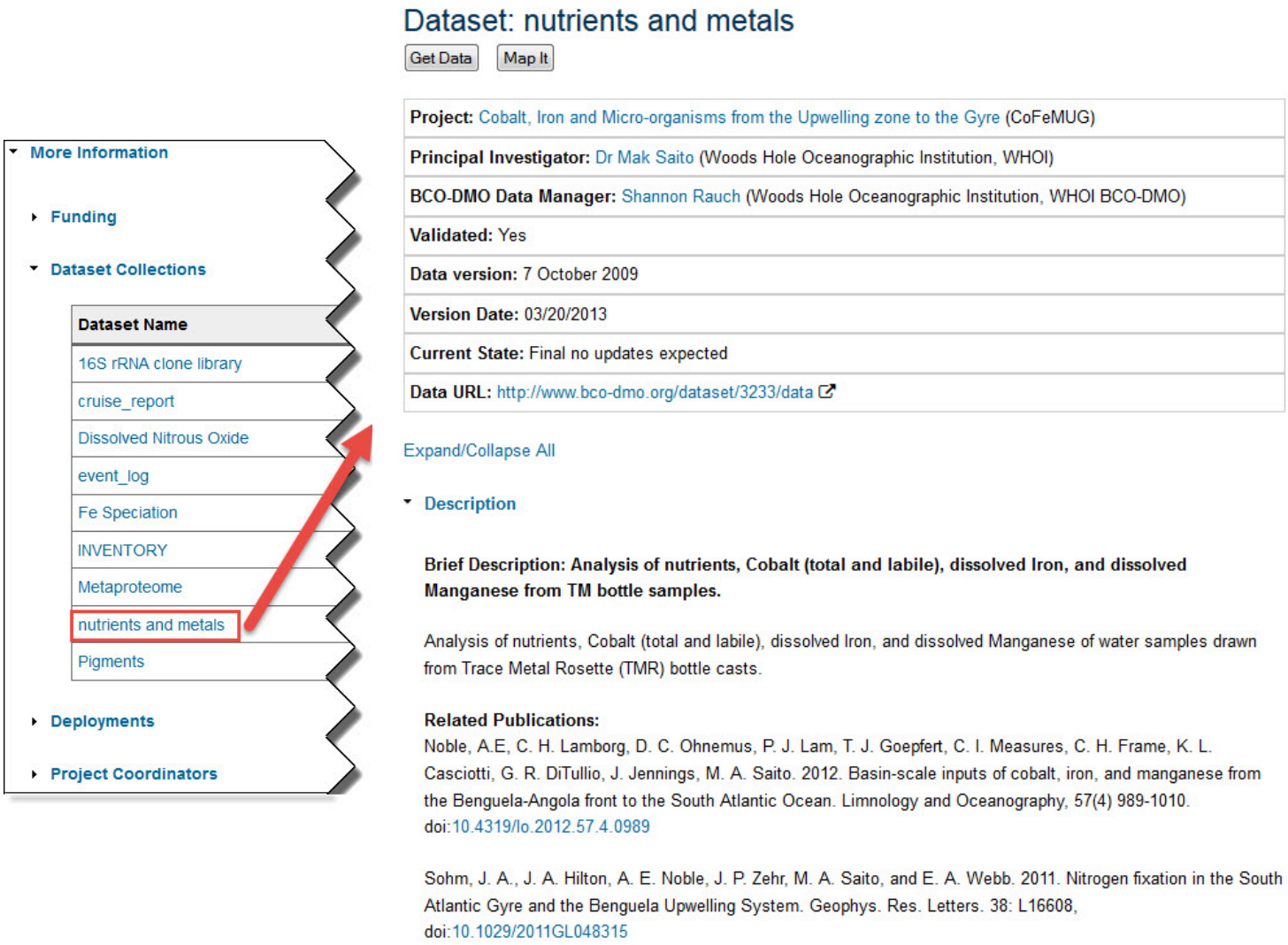
The browser displays the project description page.

URL: http://www.bco-dmo.org/project/2067

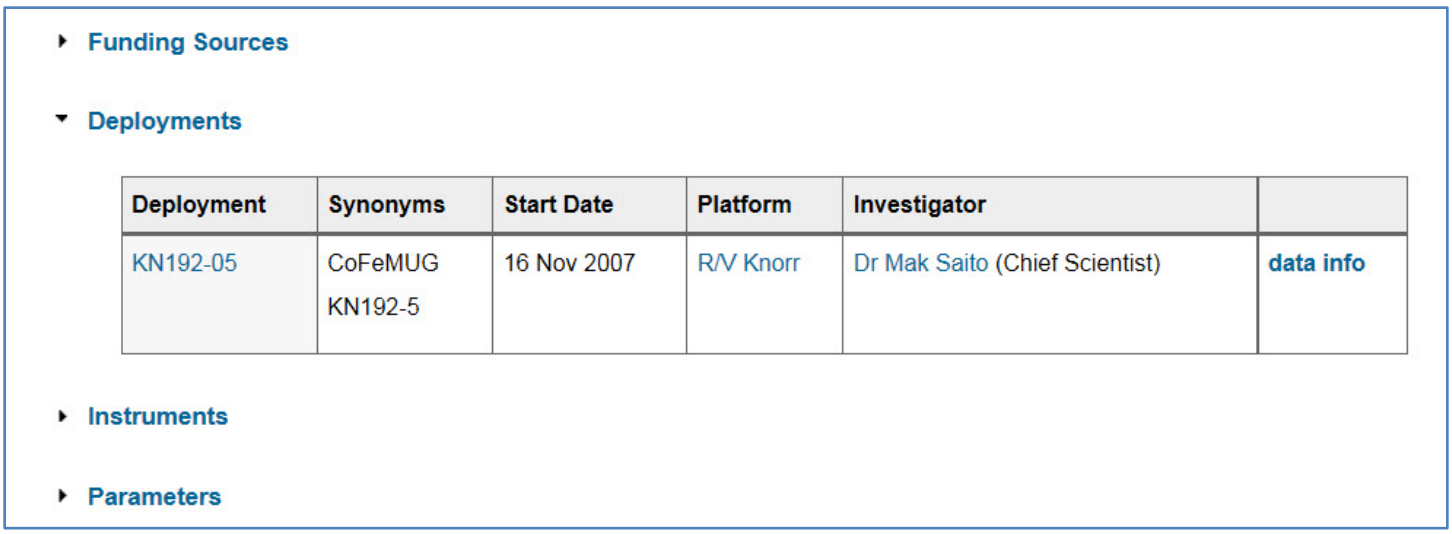


Scroll down that page and expand the Dataset Collections section.

Select the data set of interest from the list. For this example, choose “nutrients and metals”. URL: http://www.bco-dmo.org/dataset/3233



From the dataset display page, expand the Deployments section to see the cruises during which the nutrients and metal data were collected.

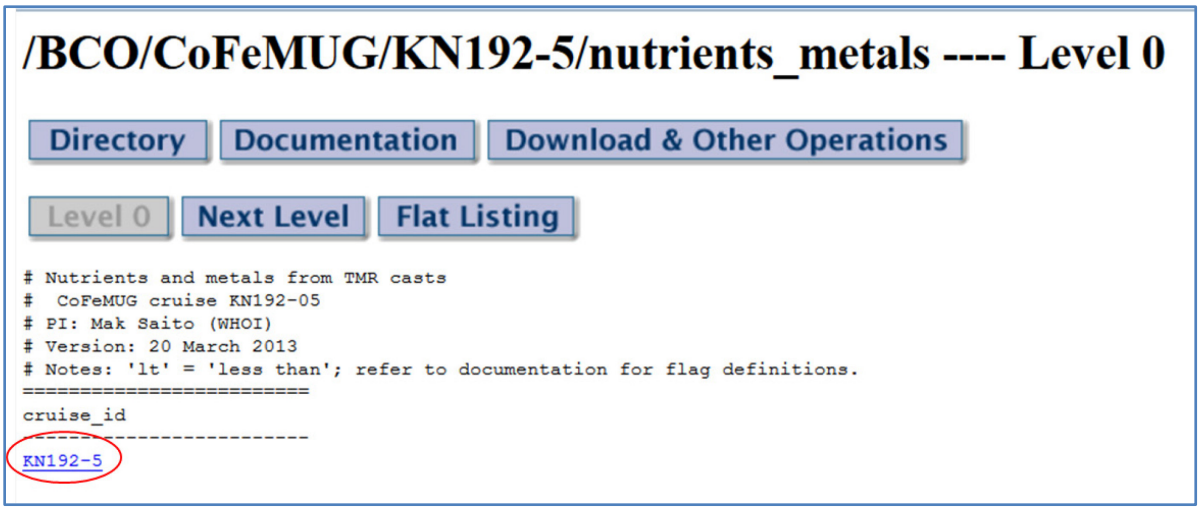


These data were collected on one cruise, KN192-05. If data were collected on multiple deployments, all deployment names would be listed here (e.g. cruises, moorings, floats, gliders, etc.). To see more information about the cruise, including a description and other available data, one could click on the deployment name.

However, we want the data …

For this dataset there are two choices: Get Data and Map It displayed near the top of the page. At this point, please click the Get Data button to access these data online.

Browser displays URL: http://data.bco-dmo.org/jg/serv/BCO/CoFeMUG/KN192-5/nutrients_metals.html0 in another window or browser tab depending on the browser configuration.



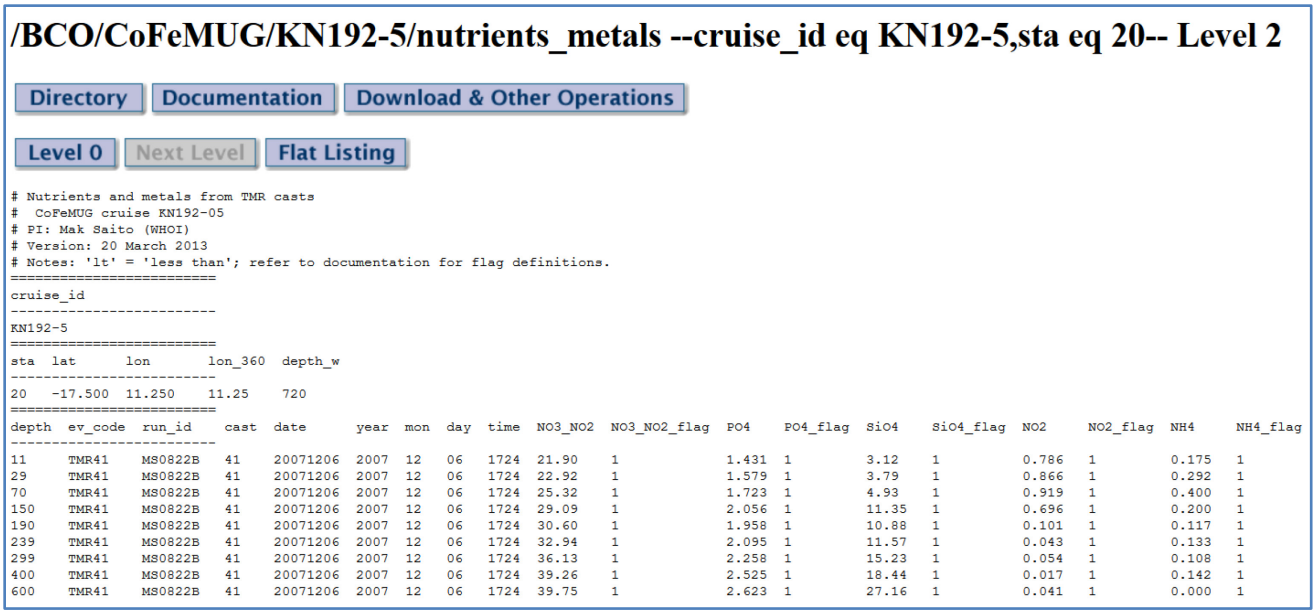
/BCO/CoFeMUG/KN192-5/nutrients_metals ---- Level 0

Directory | Documentation | Download & Other Operations

Level 0 | Next Level | Flat Listing

```
# Nutrients and metals from TMR casts
#  CoFeMUG cruise KN192-05
# PI: Mak Saito (WHOI)
# Version: 20 March 2013
# Notes: 'lt' = 'less than'; refer to documentation for flag definitions.
==========================
cruise_id
--------------------------
KN192-5
```

Click on the blue cruise ID at level 0 to expand the data display. Then, click on a station number to see the data for that station.



/BCO/CoFeMUG/KN192-5/nutrients_metals --cruise_id eq KN192-5,sta eq 20-- Level 2

Directory | Documentation | Download & Other Operations

Level 0 | Next Level | Flat Listing

```
# Nutrients and metals from TMR casts
#  CoFeMUG cruise KN192-05
# PI: Mak Saito (WHOI)
# Version: 20 March 2013
# Notes: 'lt' = 'less than'; refer to documentation for flag definitions.
==========================
cruise_id
--------------------------
KN192-5
==========================
sta  lat      lon       lon_360  depth_w
--------------------------
20   -17.500  11.250    11.25    720
==========================
```

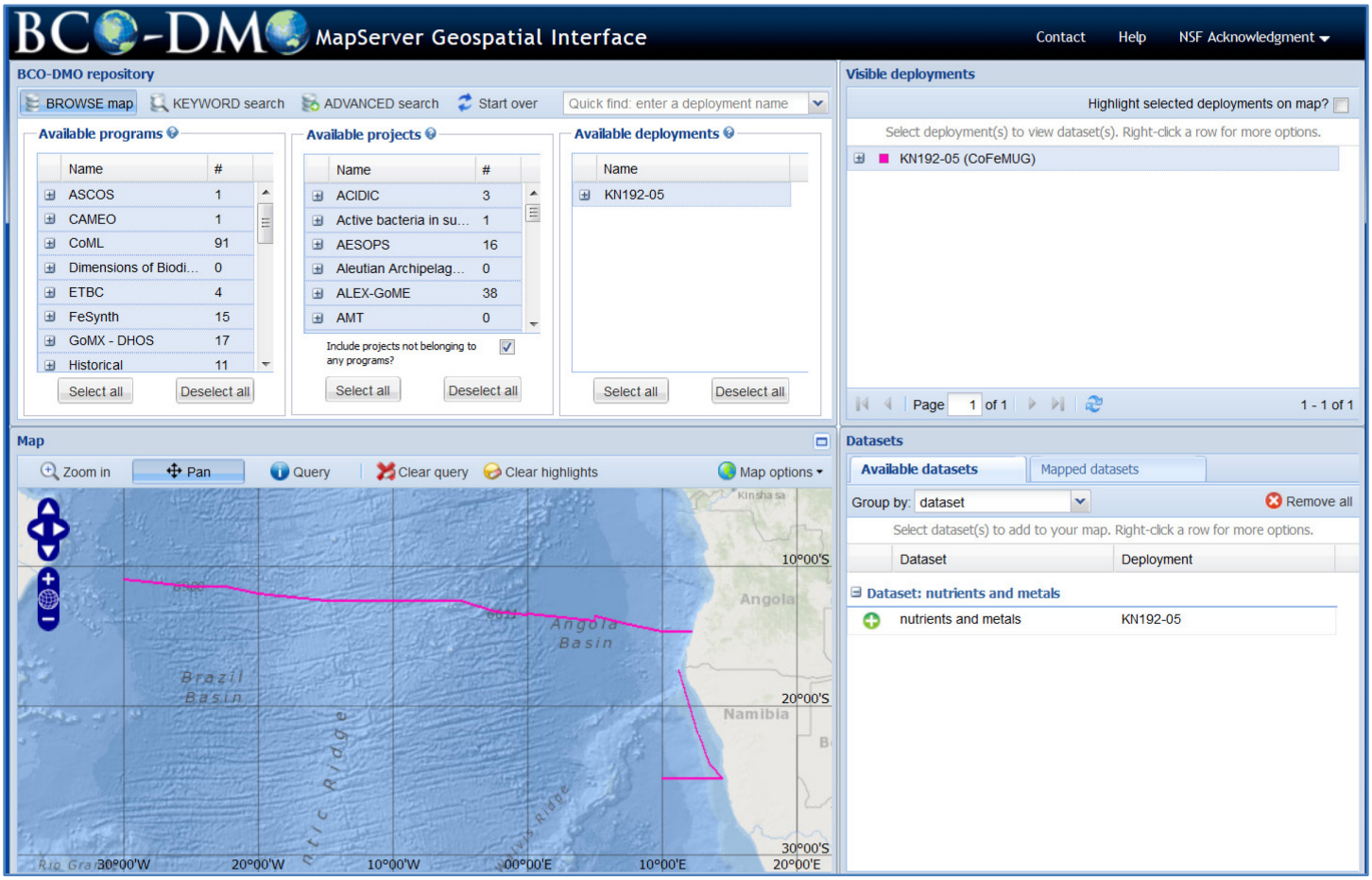
depth	ev_code	run_id	cast	date	year	mon	day	time	NO3_NO2	NO3_NO2_flag	PO4	PO4_flag	SiO4	SiO4_flag	NO2	NO2_flag	NH4	NH4_flag
11	TMR41	MS0822B	41	20071206	2007	12	06	1724	21.90	1	1.431	1	3.12	1	0.786	1	0.175	1
29	TMR41	MS0822B	41	20071206	2007	12	06	1724	22.92	1	1.579	1	3.79	1	0.866	1	0.292	1
70	TMR41	MS0822B	41	20071206	2007	12	06	1724	25.32	1	1.723	1	4.93	1	0.919	1	0.400	1
150	TMR41	MS0822B	41	20071206	2007	12	06	1724	29.09	1	2.056	1	11.35	1	0.696	1	0.200	1
190	TMR41	MS0822B	41	20071206	2007	12	06	1724	30.60	1	1.958	1	10.88	1	0.101	1	0.117	1
239	TMR41	MS0822B	41	20071206	2007	12	06	1724	32.94	1	2.095	1	11.57	1	0.043	1	0.133	1
299	TMR41	MS0822B	41	20071206	2007	12	06	1724	36.13	1	2.258	1	15.23	1	0.054	1	0.108	1
400	TMR41	MS0822B	41	20071206	2007	12	06	1724	39.26	1	2.525	1	18.44	1	0.017	1	0.142	1
600	TMR41	MS0822B	41	20071206	2007	12	06	1724	39.75	1	2.623	1	27.16	1	0.041	1	0.000	1

Explanation of buttons in the data system:

Directory	displays (returns to) the Data Directory listing for this cruise
Documentation	displays the supporting documentation for this dataset
Data Display	returns to data display from documentation display
Download & Other Operations	options for download, sub-setting and reformatting of data
Level 0	returns to level 0
Next Level	expands the data to the next level of detail
Flat Listing	displays one record per line of the **current** level of data

Now that we've seen the data, let's look at it on a map. Return to the previous browser window (or tab) with the display of dataset metadata.
URL: http://www.bco-dmo.org/dataset/3233

Click the Map It button to launch the MapServer GIS for this dataset. The browser opens a new window or tab, and the display should look similar to this for URL: http://mapservice.bco-dmo.org/mapserver/maps-ol/index.php?datasetId=3233



Some things to notice on the map shown on the previous page:

- The KN192-05 cruise track, with the map zoomed in to the area of the cruise.
- Up top, the MapServer is in "BROWSE map" mode.
- Because we launched the MapServer from the text-based data discovery system, the only "Available deployment" (in the right-most panel directly above the map) is KN192-05, the only deployment contributing data to this dataset.
- The "Map options" button (upper right-hand corner of the map panel) provides options for changing the map display (e.g. base map, projection, etc.) and printing the map. Note the small button to maximize the map window above the "Map options" button.

Once again, there are many ways to access additional information from this display:

- Clicking the + box to the left of the KN192-05 cruise ID in the "Available deployments" or "Visible deployments" panels, displays metadata about that cruise including a link to the cruise report if one is available.
- In the "Visible deployments" panel (in the top-right corner of the window), right click on the cruise ID to display available options for that cruise.

Down below (right) in the "Datasets" panel:

- Clicking on the + symbol in the green circle, or anywhere on the row with the dataset name (nutrients and metals), requests that this dataset be 'mapped'.

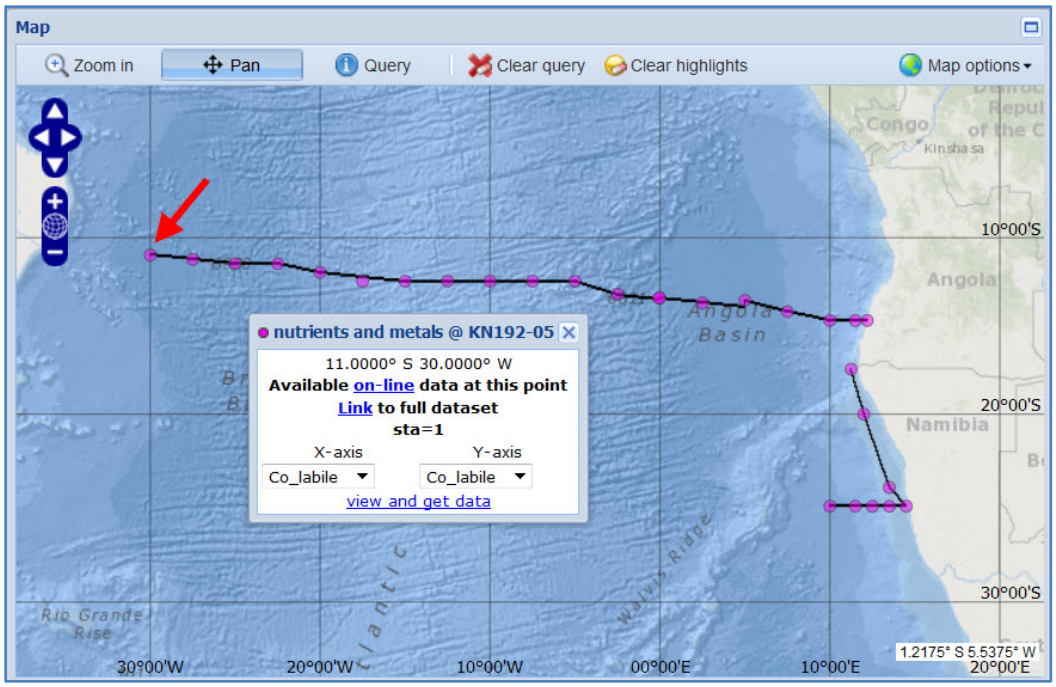
Try it … small colored dots will appear on the map, and the Datasets panel on the right has shifted to a display of "Mapped datasets", of which there is only one at the moment. The single mapped dataset is listed, with a color-coded dot, the dataset name (nutrients and metals), a number (28) that indicates the number of sampling locations included in this dataset, from the selected cruise (KN192-05).

Once again, we have several ways to get more information: Right-clicking anywhere on the text in the row with the dataset name pops up a menu with several options:

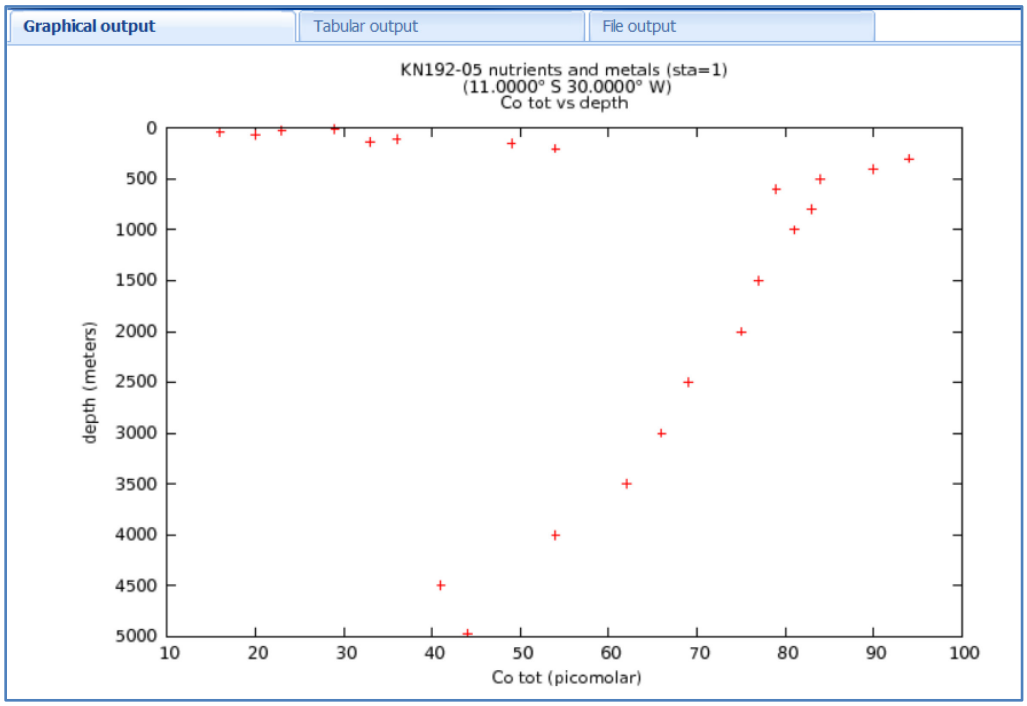
- "View/export mapped dataset" shows a tabular listing of the sampling locations, with options for data export;
- "View mapped dataset on-line" opens the full dataset online;
- "Choose a color" allows you to change the color of the dots on the map;
- "MapServer link to mapped dataset" gives you the ability to save/bookmark the MapServer link to this dataset (for later access)
- "Remove mapped dataset" removes this dataset from the map (removes the dots). Note that unchecking the box to the left of the dataset name temporarily removes the dataset. Checking the box again displays them.

The line underneath the dataset name shows more options/information for each dataset (different options are available depending on the data type).

The MapServer also provides the ability to generate 'quick view' plots of the data at a sampling location. On the map, select the dot for station 1 (the western-most point indicated by the red arrow in the following figure), to bring up a dialog box that offers: links to the data from the database and a way to select variables from the dataset to generate a 'quick view' X-Y plot. (If the dialog box ever gets in the way, just move it aside.)



Choose ‘Co_tot for the X-axis and ‘depth’ for the Y-axis. Click “view and get data” at the bottom of the box to generate the plot on the next page of total Cobalt concentration vs. depth.



Notice once again, several tabs at the top make it easy to view a tabular list of the data in the graph or download the data from the “File output tab”. Close the graph.

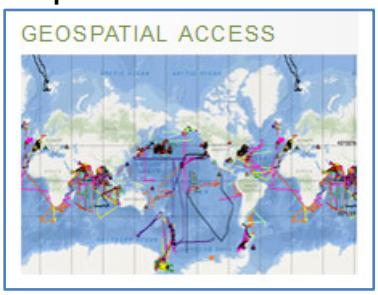
(Note: this dataset has the option to plot data from multiple stations, as indicated by the line stating “Plot multiple records by sta on one graph.” under the dataset name in the “Mapped datasets” panel. This function will be demonstrated in part 3 of this tutorial.)

Data access: MAP BROWSE

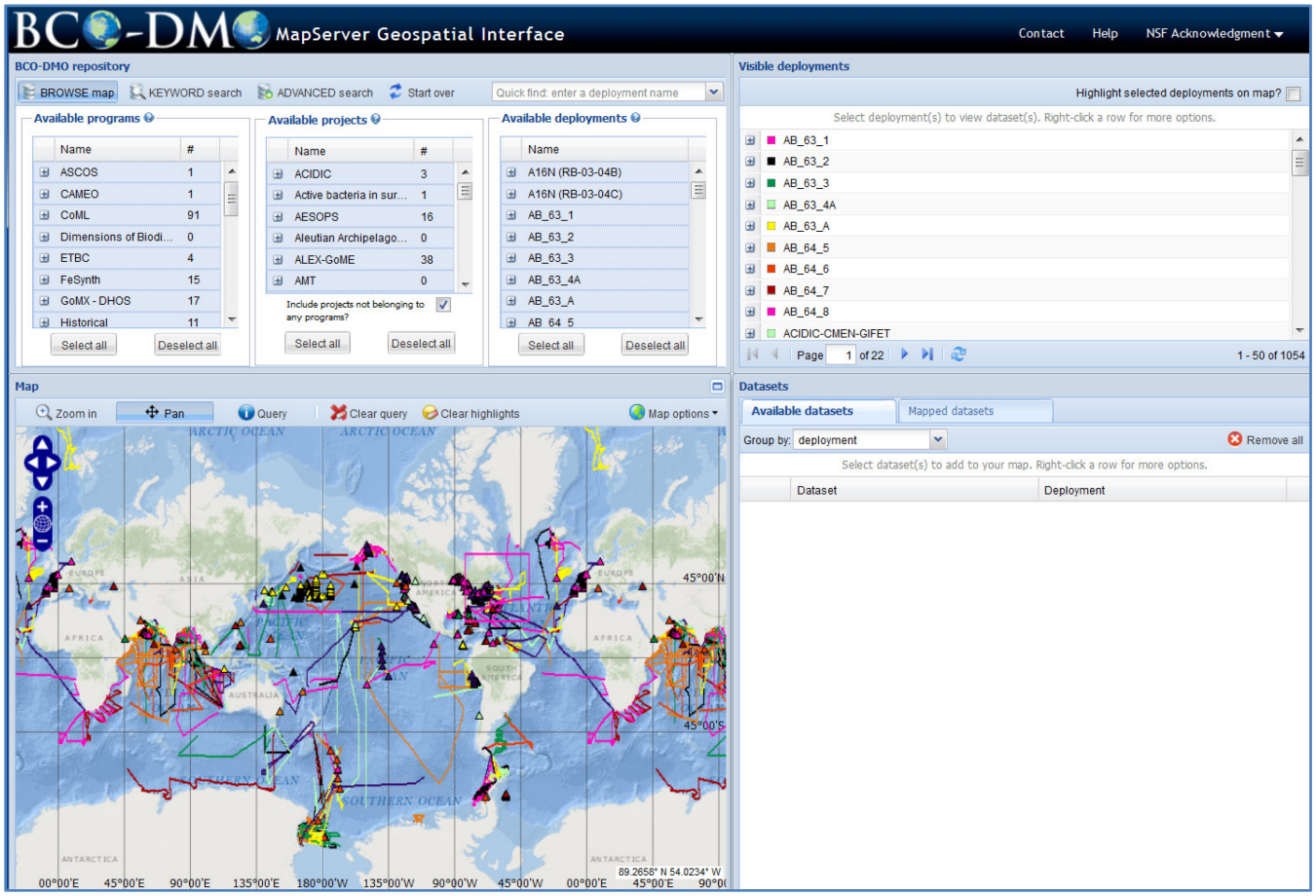
scenario 2: You are interested in data from a particular geographic region.

Go to: http://bcodmo.org/ the BCO-DMO home page

At the bottom of the DATABASE column on the left, click on the GEOSPATIAL ACCESS map.



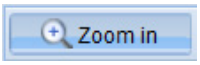
The MapServer system map showing all the deployments from the BCO-DMO database opens in a new window or tab.



A common way to use the MapServer GIS is to define a region of interest on the map.

For example: you are interested in phytoplankton blooms in the North Atlantic.

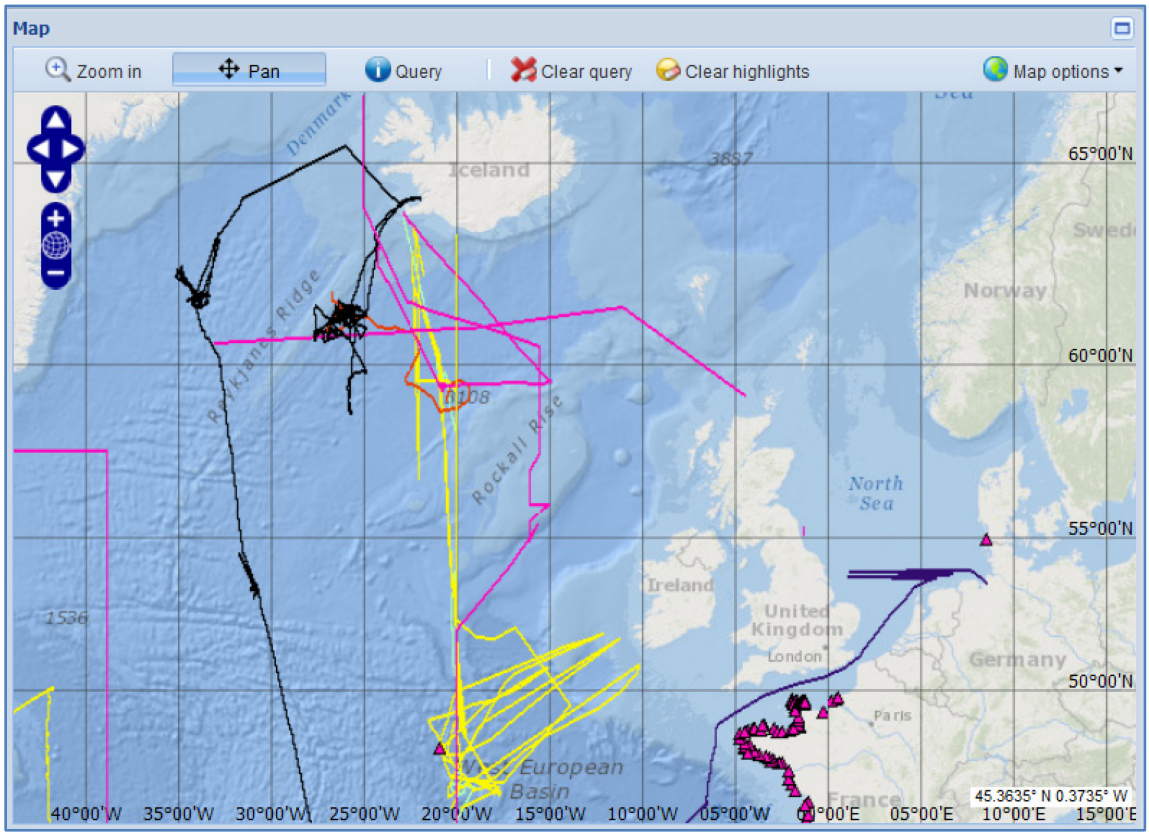
The colored lines represent cruise tracks; triangular symbols represent fixed sampling sites (e.g. mooring locations or time-series sites). Zoom in to the cruise track just west of Europe (between the Labrador and North Sea in the vicinity of Ireland at 50° N 20° W).

Click the "Zoom in"  button and then click and drag to define an area on the map.

Notice the available map positioning tools shown at right (indicated by the red arrow).

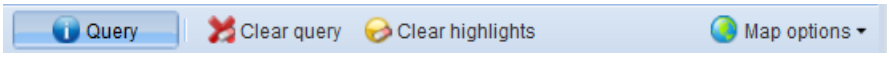
The default mode is Pan, and if using the "Zoom in" button you will need to put the "mapper" in Zoom mode each time if doing multiple zoom actions. Suggestion: once the area of interest has been selected using "Zoom in", use the zoom tools (+ and – buttons).

Use the various map positioning tools (shown above right) until your map looks something like the screen shot below.

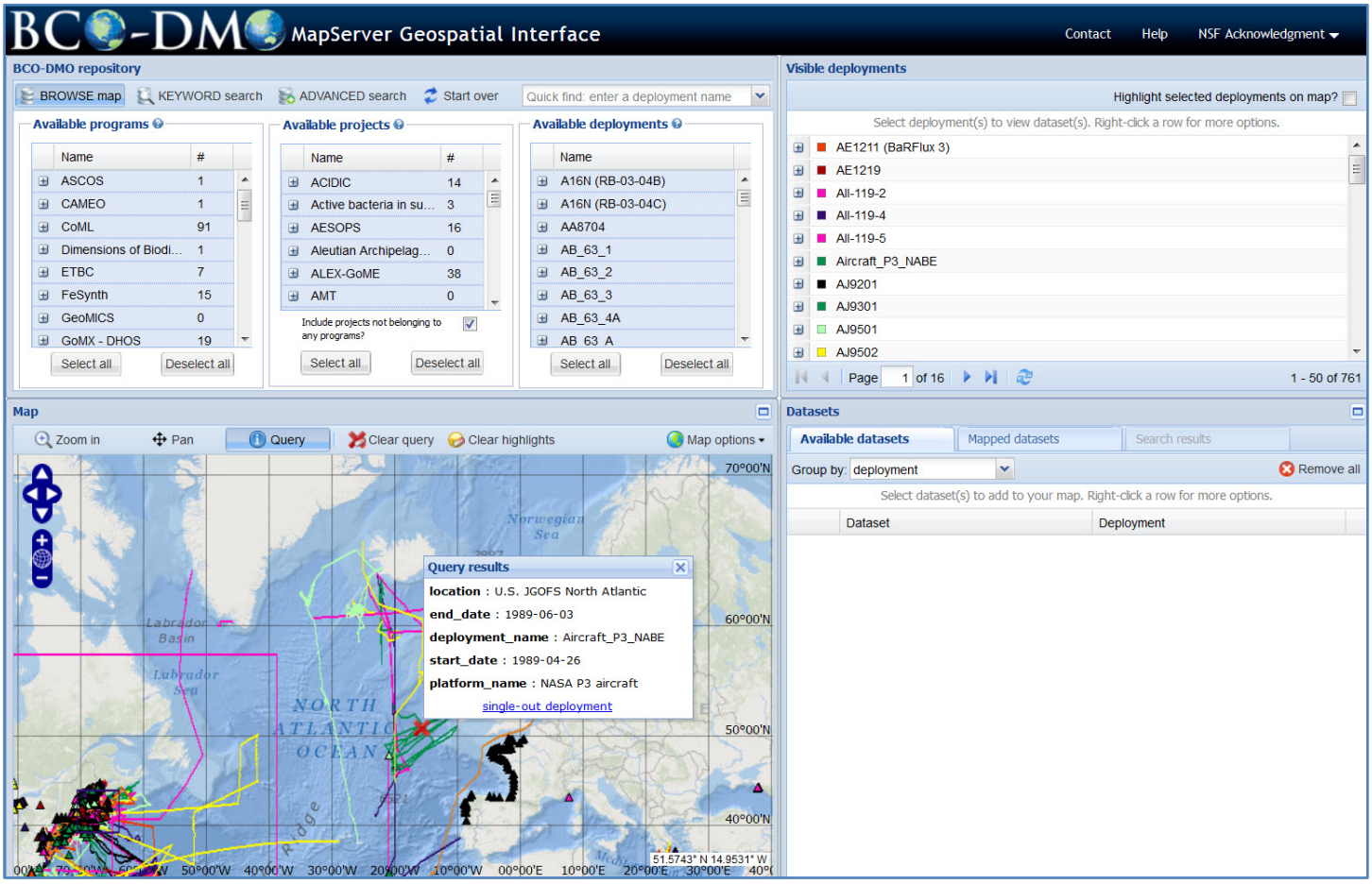


There is a lot of information displayed on this map and in the associated panels. We notice an assortment of cruise tracks (colored lines) and fixed sampling locations (colored triangles). The map system offers several tools for figuring out what is being shown on the map. The map controls appear in a tool bar above the map.

To find out what cruises these are, use the Query tool (click on the Query button now). Then click on a colored item on the map. For example, click on the cruise track southwest of Ireland (in the image below, this is the green line with the red X on it).



A new "Query results" window will pop-up on the map, identifying the deployment name, platform, location, and dates.

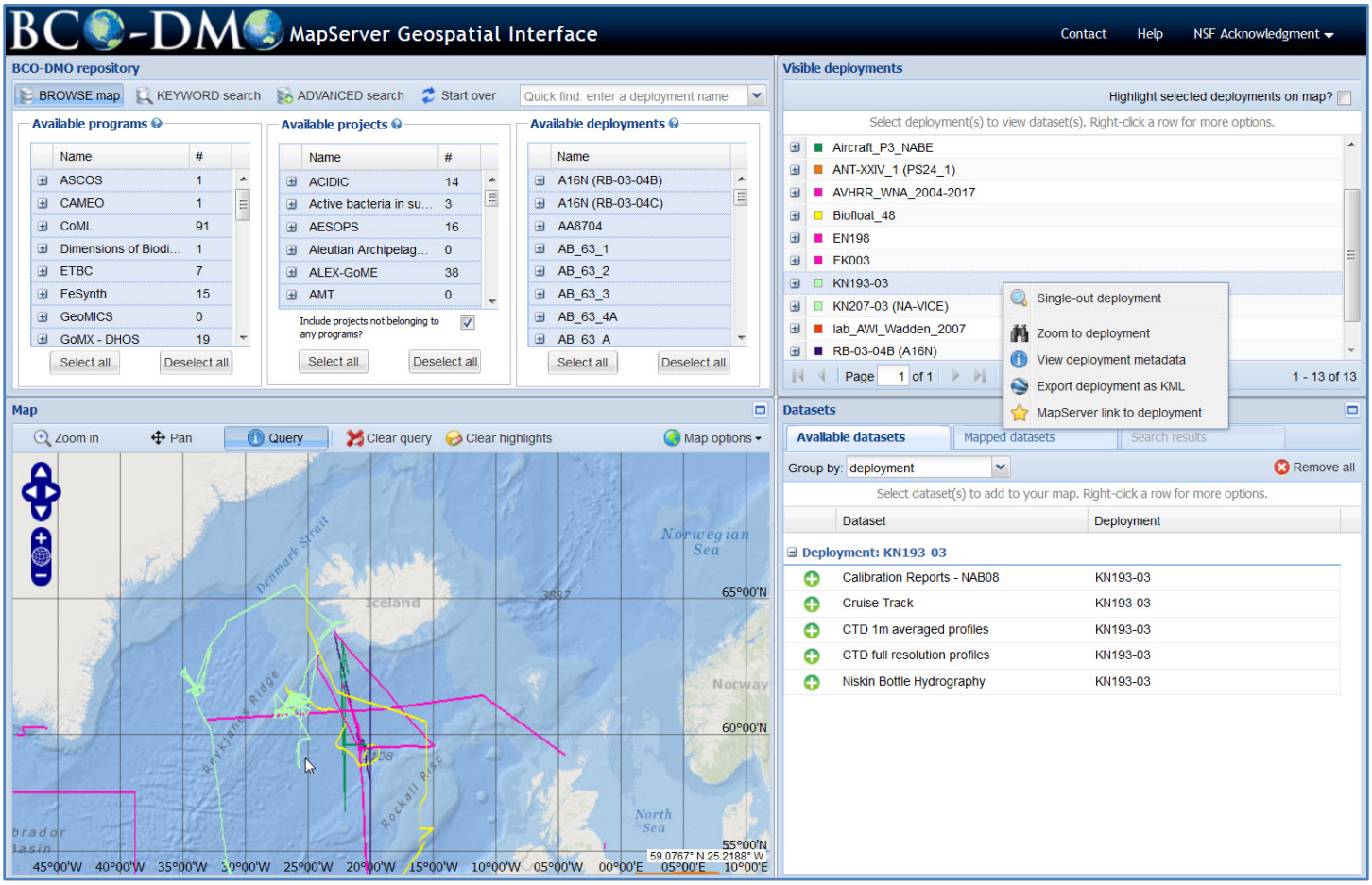


The results of the Query tool indicate that the selected navigation track is from a deployment called "Aircraft_P3_NABE". The results window also reveals that this track is from a NASA P3 aircraft flight. In the results box, you could click on "single-out deployment" to remove all other deployments from the map and show only the one identified.

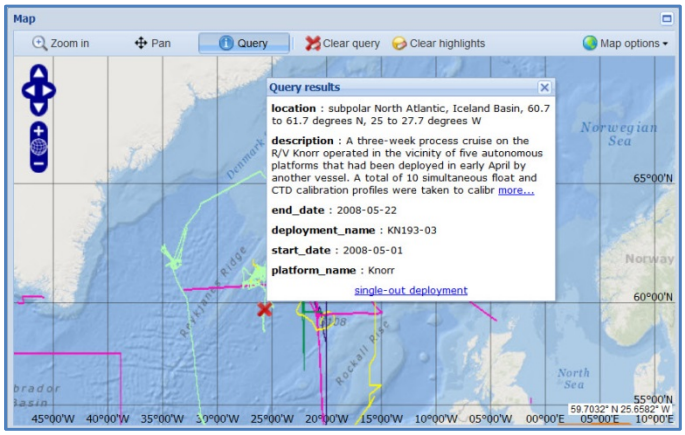
But let's find some data…

Use the “Zoom in” tool to zoom in closer to the area around Iceland.

Click the “Query” button and then click on the cruise track southwest of Iceland (in the image below, this is the light green line next to the arrow cursor on the map).



Click on the Query tool, then click on this deployment to identify it.

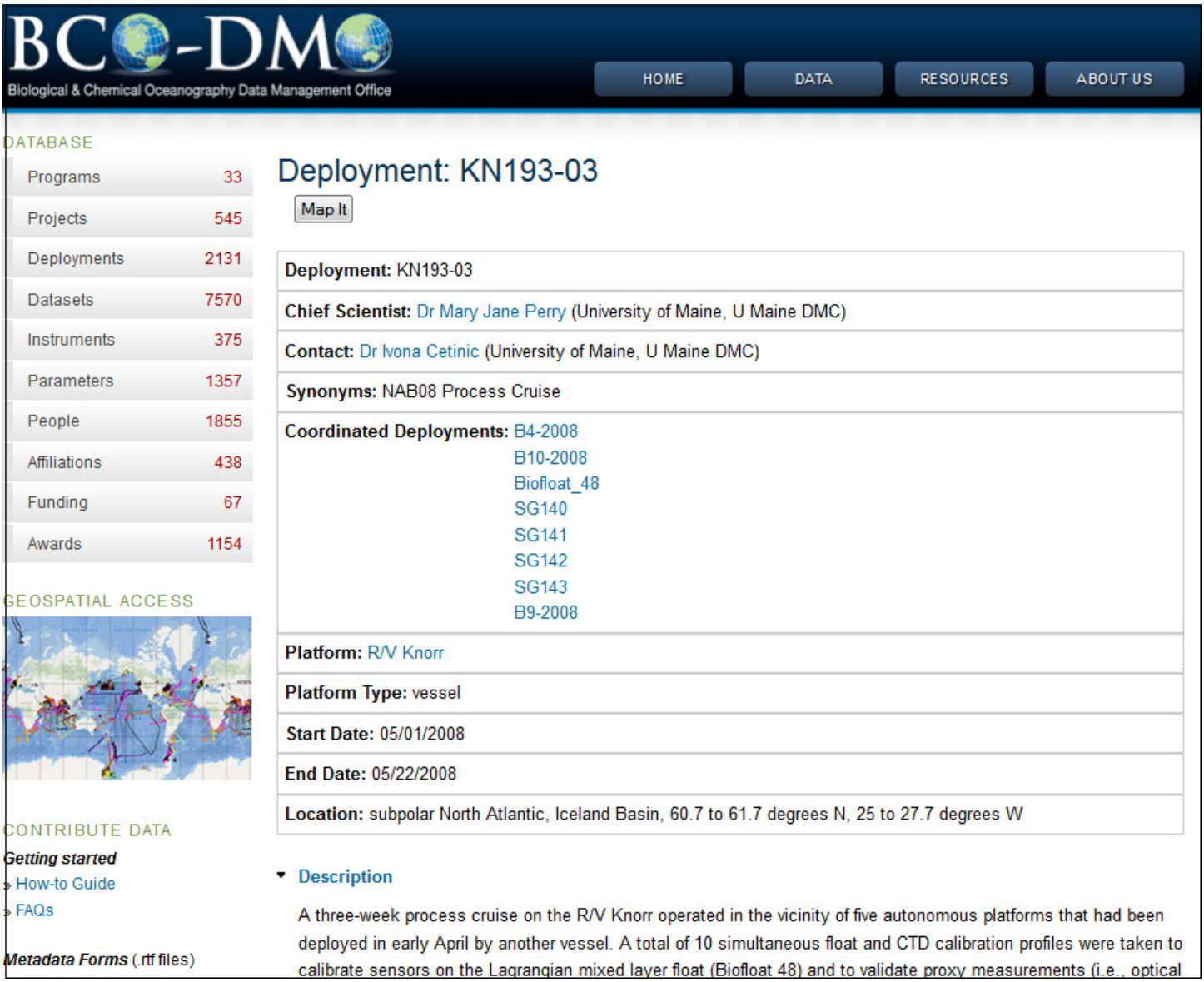


This deployment is identified as KN193-03, as shown in the image on the left.

Right click on the KN193-03 line in the Visible deployments panel (see image above), and select “View deployment metadata”.

This opens up a new browser window or tab with a view of the deployment metadata/documentation (shown on the next page).

View of KN193-03 deployment (cruise) metadata (documentation) retrieved by the MapServer from the BCO-DMO database.



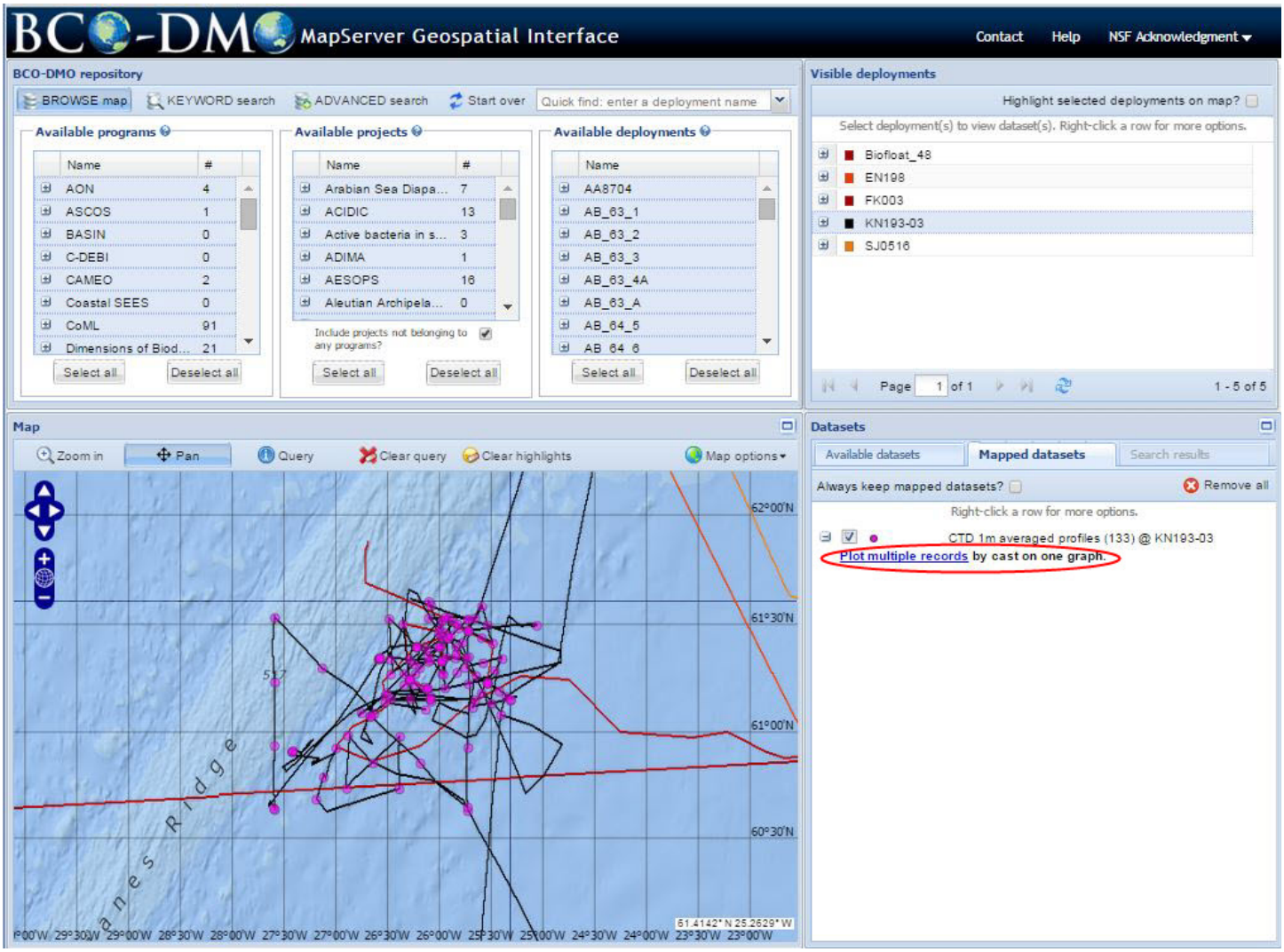
Return to the map (the browser window with this URL address: http://mapservice.bco-dmo.org/mapserver/maps-ol/index.php). Most likely this is in the browser tab that precedes the one you have been looking at with the deployment metadata (that looks like the screen shot above).

Now take the map out of Query mode by clicking on the “Pan” button (to the left of the “Query” button) and then click the “Clear query” button to remove the red X from the map.

Let’s see what data the 2008 North Atlantic Bloom Experiment (NAB 2008) investigators have shared from their KN193-03 cruise.

In the “Datasets” panel (in the lower right of the browser window), select the “CTD 1m averaged profiles” dataset from the list of available datasets from this cruise. Notice the dots that appear on the map, which represent the CTD casts.

Use the Zoom + tool to zoom in until the spatial extent looks similar to the image below (some additional dialog windows from the next step have already been opened and are visible in the image below).

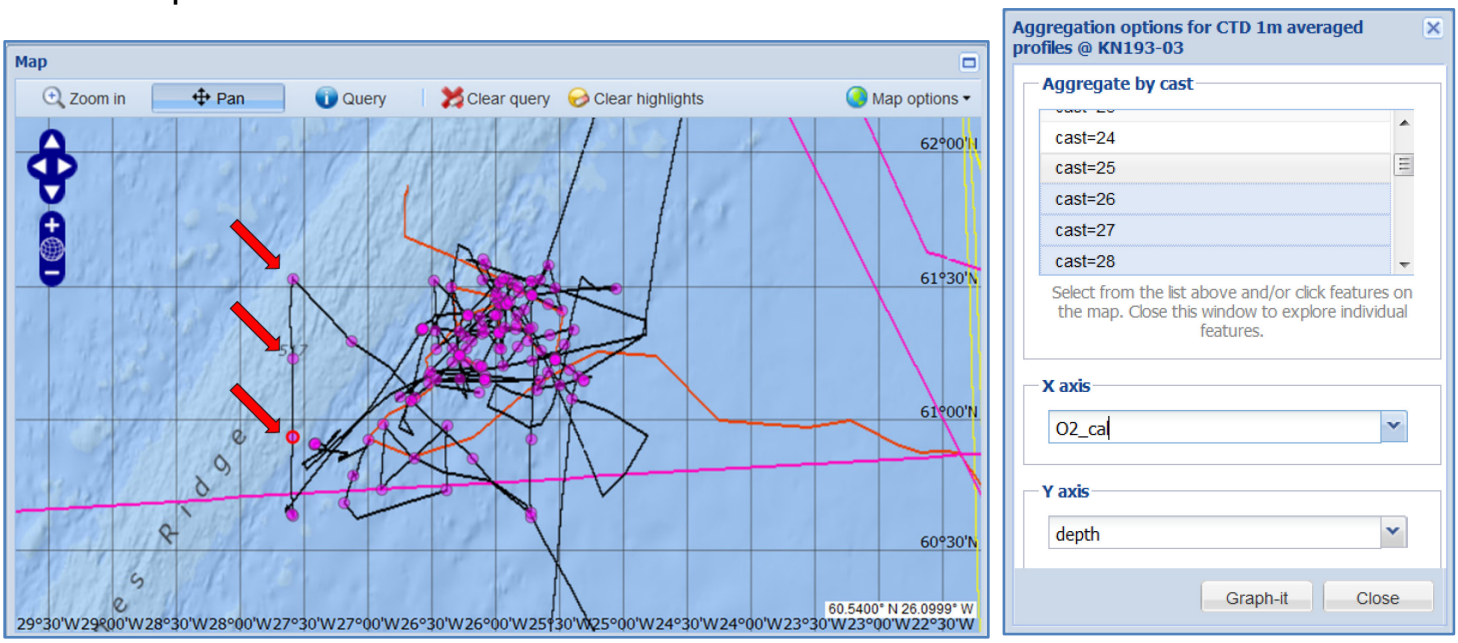


This CTD dataset provides the option of making a graph containing data from multiple casts. In the “Mapped datasets” panel, click where indicated to create an aggregate graph by cast (circled in the image above).

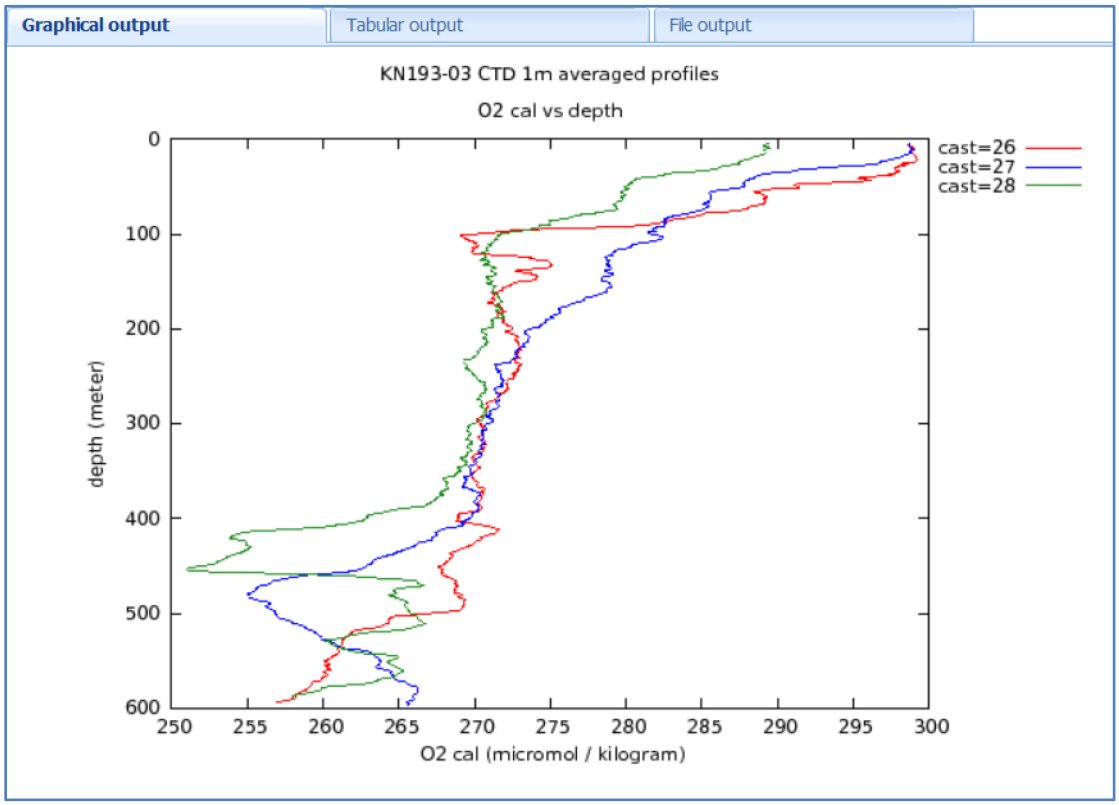
A window opens that allows you to choose the casts to include in the graph. Let’s make a graph of casts 26, 27, and 28. You can select stations in two ways:

1. Click on the dots on the map. Each selected dot will be highlighted in the aggregation options window; OR
2. Hold down the CTRL key and click on the station names in the aggregation options window.

Casts 26, 27, and 28 are indicated by the arrows in the image below. After selecting the stations (using either method), choose ‘O2_cal for the X-axis and ‘depth’ for the Y-axis. The multiple plot options window should like the image below. Then, click the “Graph-it” button.

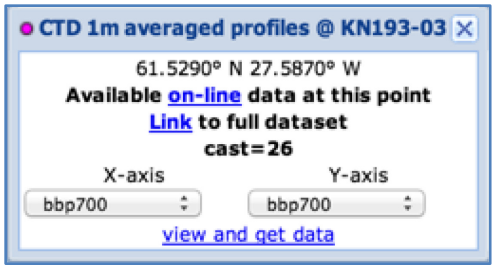


The resulting graph should look like the one below, where calibrated oxygen from 3 casts is plotted against depth. Each cast is represented by a different color, as indicated in the legend located in the upper right-hand corner.

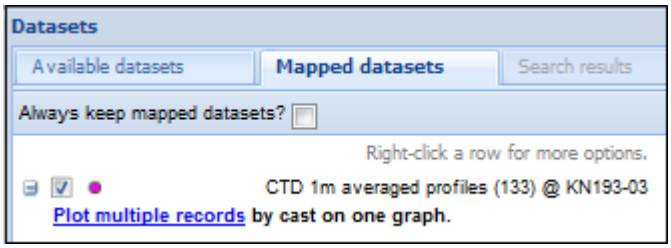


Close the graph and the aggregation options window.

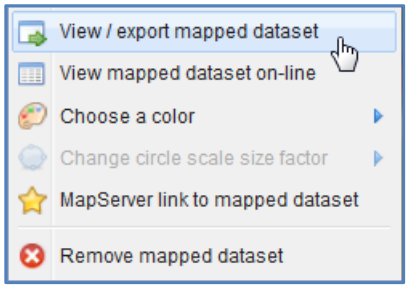
The data can be downloaded using several different methods:



From the map, one can view the data from any selected point ("Available on-line data at this point"), or one can view all the records in the full data dataset ("Link to full dataset").

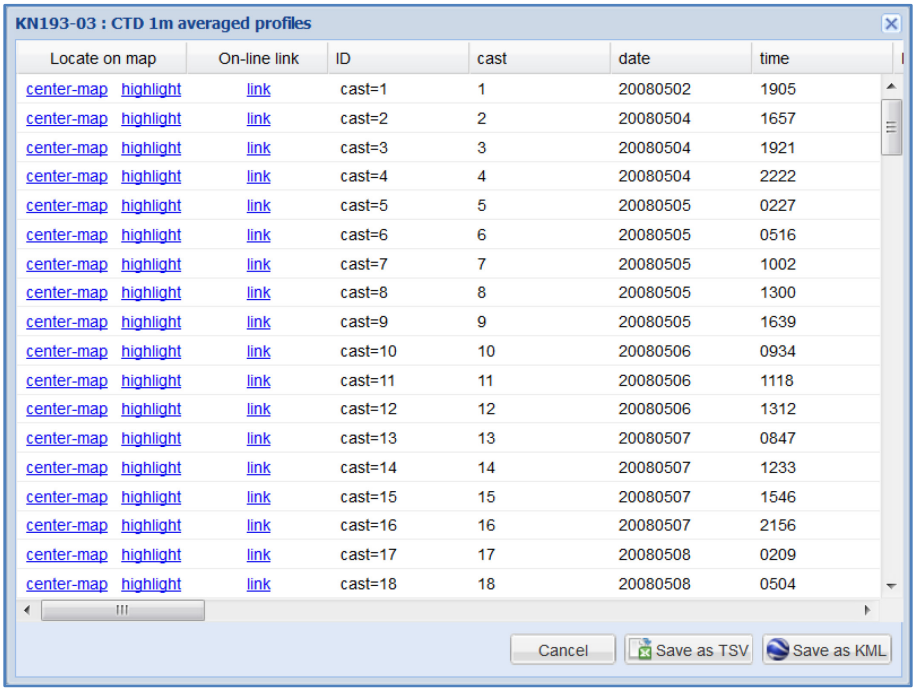


Another option is to right click on the dataset line in the Datasets panel (to the right of the map display).



Right clicking the dataset name brings up the dataset options menu, from which one can export the data. Choose "View/export mapped data".

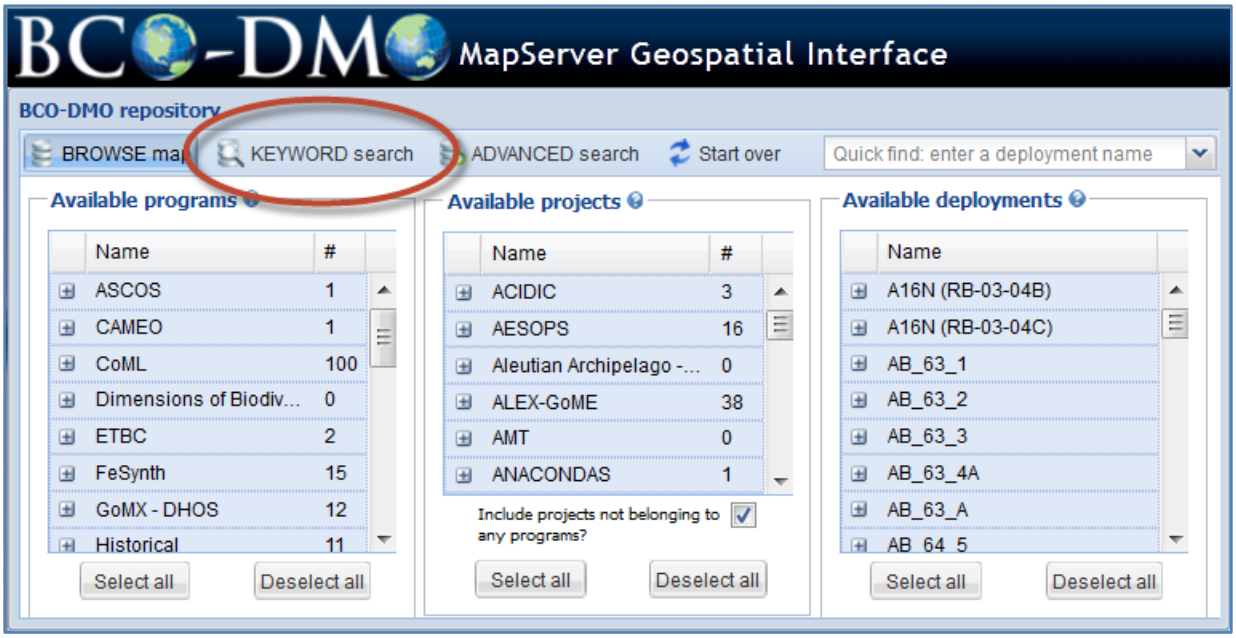
In addition to the data table display, one has the option to download the data as tab-separated values (TSV) or KML (Google Earth) formatted files.



Data access: MAP KEYWORD SEARCH

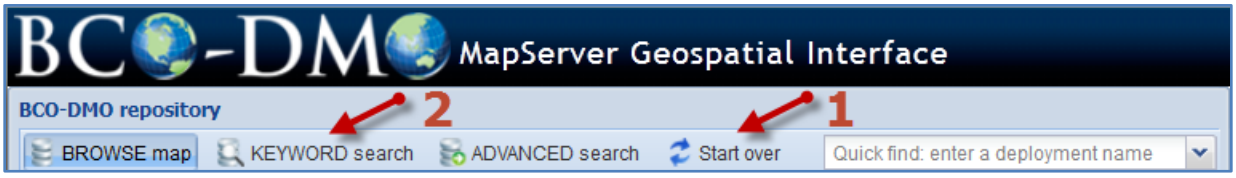
scenario 3: You are interested in data of a particular type from a particular geographic area.

The MapServer has a KEYWORD search that is useful for finding certain types of data.



To test out the KEYWORD search function,

1. Click “Start over” in the top tool bar above the search panels (to clear the map and restart with a fresh map), and
2. click the “KEYWORD search” menu item toward the upper left-hand corner of the browser.

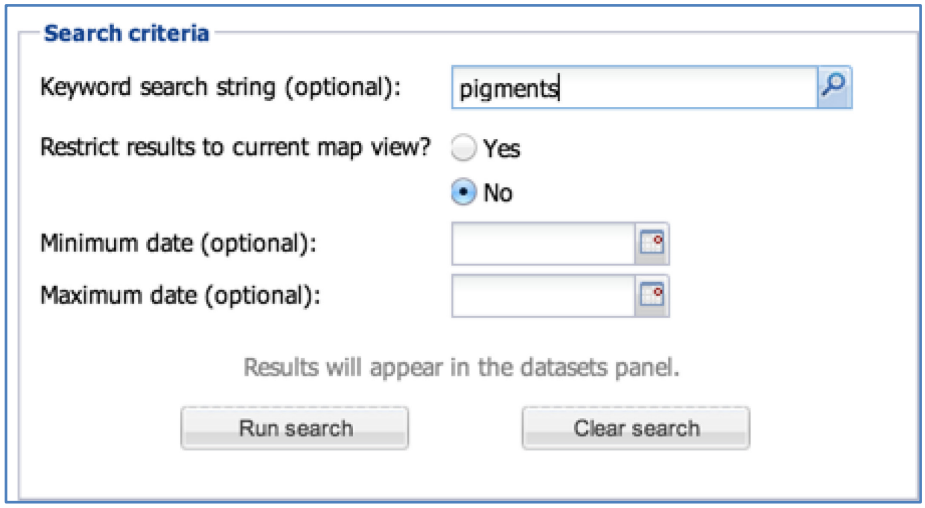


Enter something of interest in the “Keyword search string” text entry box.

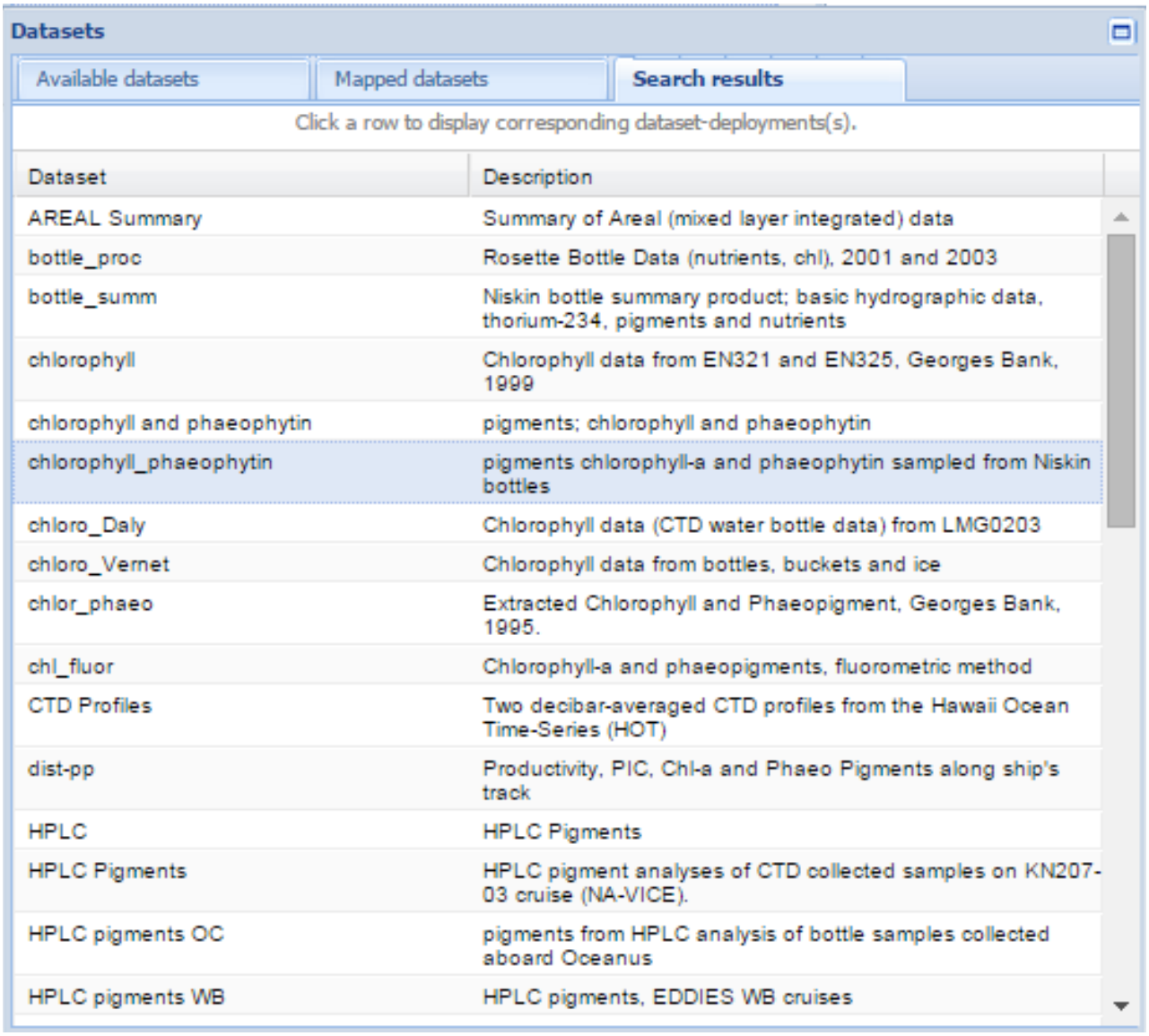
For example, type in “pigments”.

If you are interested in pigments from only a certain geographic area, you can zoom-in to that area on the map, then choose “Yes” next to “Restrict results to map?”

You may also specify a date range in the minimum and maximum date fields.



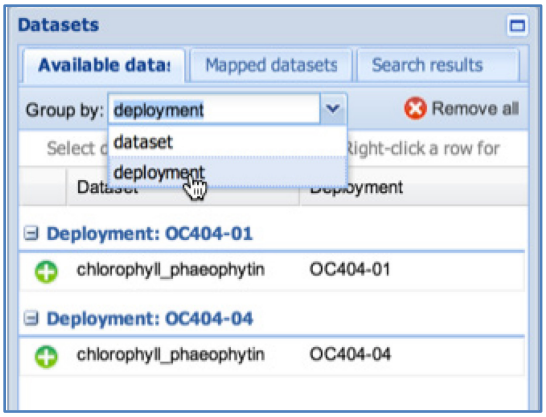
For now, let's just search for "pigments". Click the "Run Search" button. The results appear in the datasets panel. Select "chlorophyll_phaeophytin".



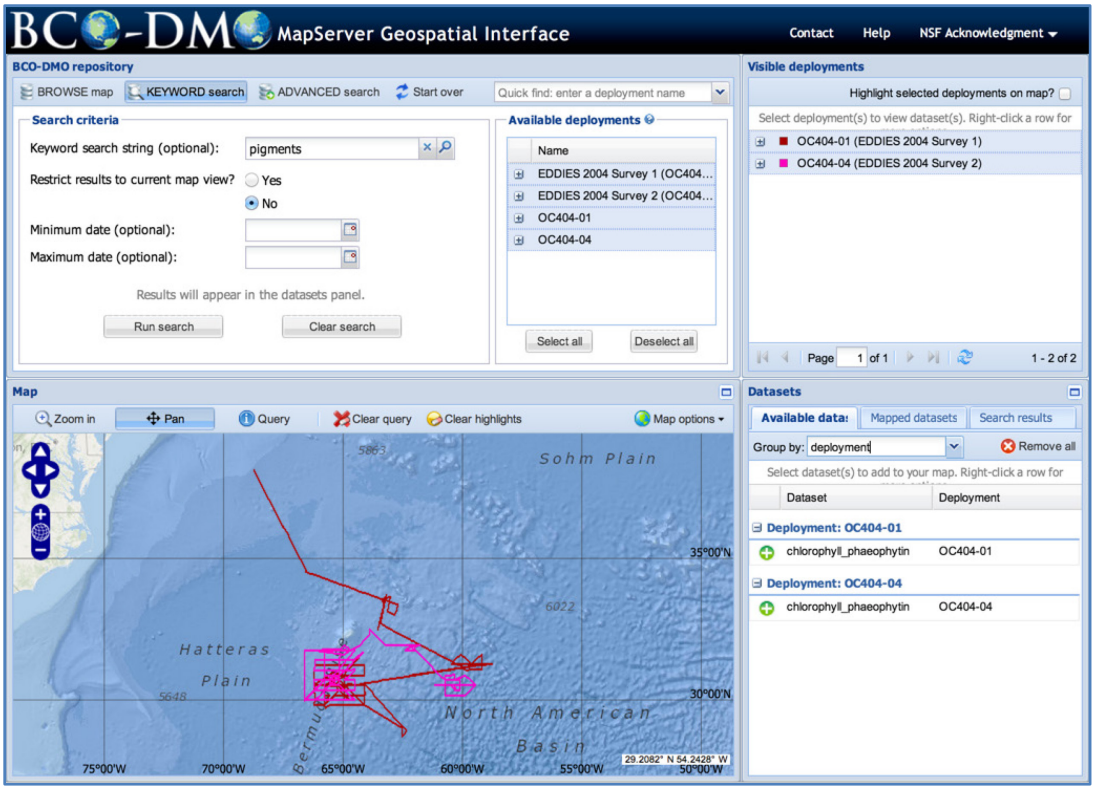
Dataset	Description
AREAL Summary	Summary of Areal (mixed layer integrated) data
bottle_proc	Rosette Bottle Data (nutrients, chl), 2001 and 2003
bottle_summ	Niskin bottle summary product; basic hydrographic data, thorium-234, pigments and nutrients
chlorophyll	Chlorophyll data from EN321 and EN325, Georges Bank, 1999
chlorophyll and phaeophytin	pigments; chlorophyll and phaeophytin
chlorophyll_phaeophytin	pigments chlorophyll-a and phaeophytin sampled from Niskin bottles
chloro_Daly	Chlorophyll data (CTD water bottle data) from LMG0203
chloro_Vernet	Chlorophyll data from bottles, buckets and ice
chlor_phaeo	Extracted Chlorophyll and Phaeopigment, Georges Bank, 1995.
chl_fluor	Chlorophyll-a and phaeopigments, fluorometric method
CTD Profiles	Two decibar-averaged CTD profiles from the Hawaii Ocean Time-Series (HOT)
dist-pp	Productivity, PIC, Chl-a and Phaeo Pigments along ship's track
HPLC	HPLC Pigments
HPLC Pigments	HPLC pigment analyses of CTD collected samples on KN207-03 cruise (NA-VICE).
HPLC pigments OC	pigments from HPLC analysis of bottle samples collected aboard Oceanus
HPLC pigments WB	HPLC pigments, EDDIES WB cruises

The map updates to show that these data have been reported from two deployments: OC404-01 and OC404-04. At this point, you can choose a different search result if this was not what you were looking for. Zoom-in to the cruise tracks (if needed).

The cruise tracks are displayed on the map, and the datasets available from both cruises are listed in the "Datasets" panel. Click on the 'Available datasets' tab. They are grouped by dataset by default. Choose "deployment" from the "Group by" drop-down menu to change the grouping. (This is more useful for cruises that have many datasets.)



The datasets have been discovered and at this point one can proceed to map the datasets, view the data, or make quick view plots.



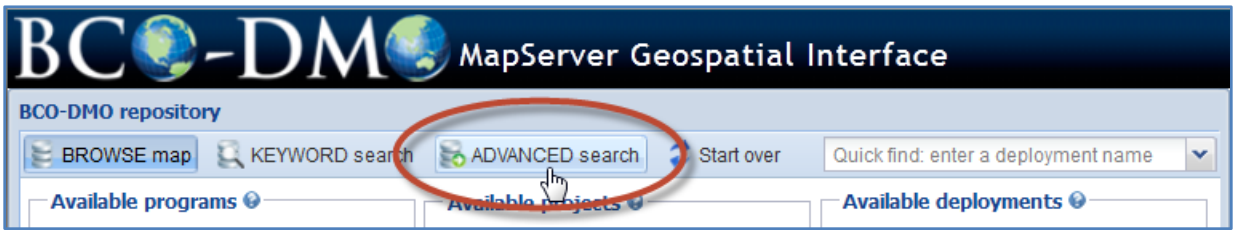
Data access: MAP SEMANTIC SEARCH

scenario 4: You have an idea what you are looking for, but you do not know the Program, Project, or Deployment name.

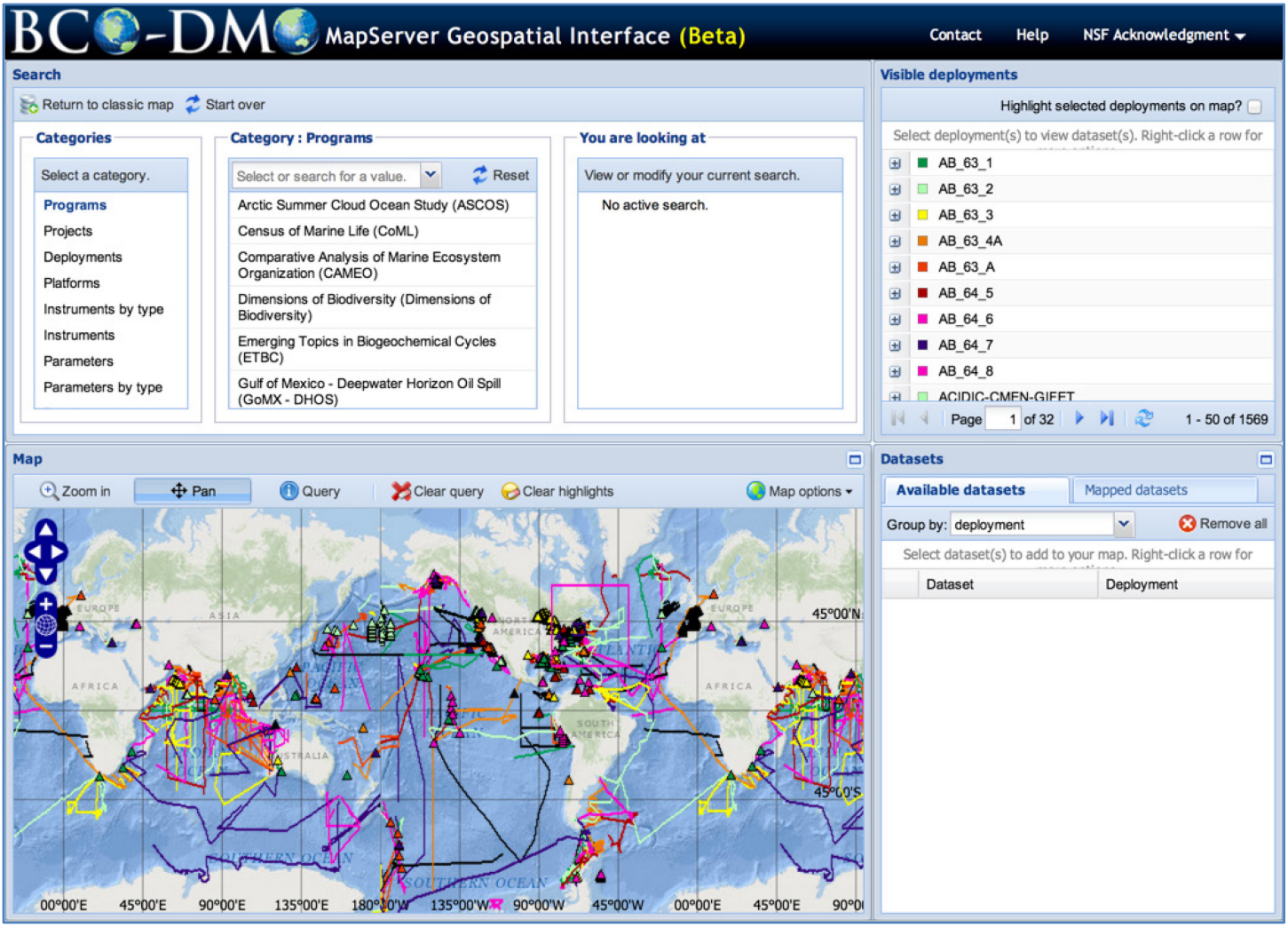
The MapServer ADVANCED search is a semantically-enabled, faceted search. It provides access to the same data as the "BROWSE map" and "KEYWORD search" modes, but the use of the search categories allows the user to construct their own hierarchical filter/search.

A beta version of this interface was released in May 2012, and is available as the ADVANCED search option (still as a beta version) from the MapServer.

Try it out …



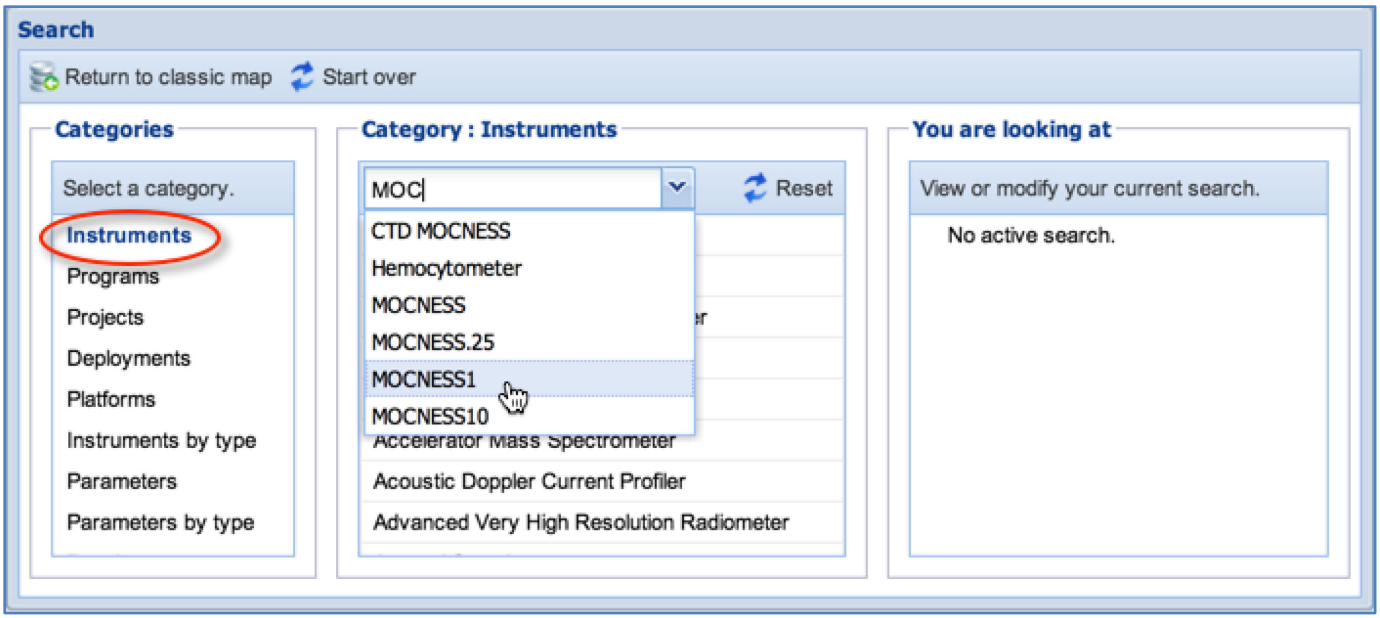
Click on the ADVANCED search button. The map interface will reinitialize, clearing any selections you have made previously; click "Yes" in the pop up dialog box to confirm this and to continue. The browser should look like the image on the next page.



The elements of the ADVANCED (semantic) search interface are similar to the MapServer layout you have seen in the previous scenarios in this tutorial.

The Search panel initializes with the Programs category selected by default. For this exercise, let's assume you know the type of instrument that collects the data you want to find. For example, you have read a paper about some data collected with a MOCNESS system using 1-m^2 nets, and the paper says the data are available from BCO-DMO.

In the Categories panel on the left, select the "Instruments" category.



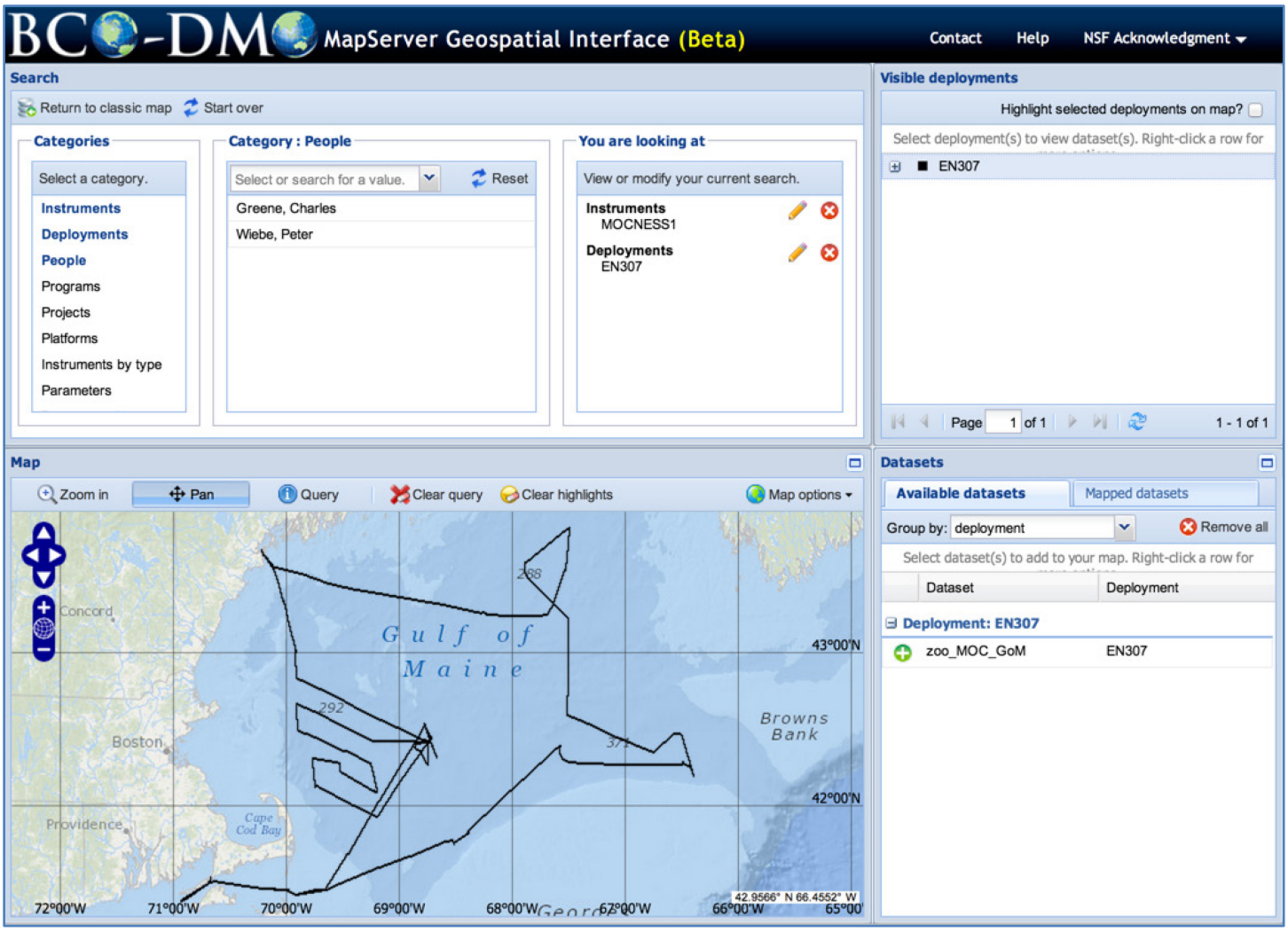
Enter an instrument name into center panel; for example, in the text entry box under "Category: Instruments", type "MOCNESS". We are interested in a MOCNESS with 1-m^2 nets, so select "MOCNESS1" from the list of matches that pop up.

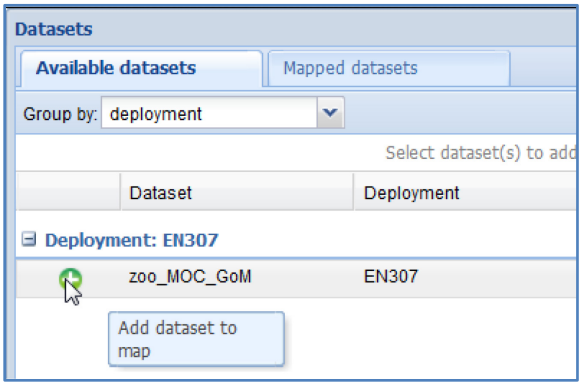
Observe how "MOCNESS1" is highlighted in the list of instruments, and the list of "Visible deployments" changes to reflect that you are only interested in data from a MOCNESS1 tow.

At this point you could further narrow the search by selecting a deployment on the map, but let's try choosing a deployment (cruise ID).

In the Category search panel on the left, click the "Deployments" category. Scroll down to EN307, or type it into the text field. Observe how the "Visible Deployments" panel on the right updates to display only deployment EN307.

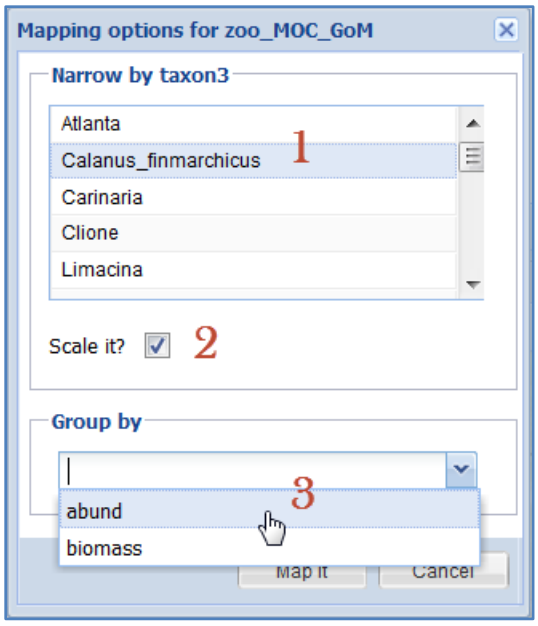
If you return to the Category search panel and click on the "People" category, you can see that the EN307 MOCNESS1 data were reported by Peter Wiebe and Charles Greene. If you click on EN307 in the "Visible Deployments" panel, you will see one dataset "zoo_MOC_GoM" listed in the "Available dataset" section. Your browser window should look similar to the following image.





To map the dataset, click on the "zoo_MOC_GoM" dataset from the "Available datasets" panel in the lower right.

A dialog box pops up with "Mapping options" for this dataset. This particular type of dataset has some custom mapping options.

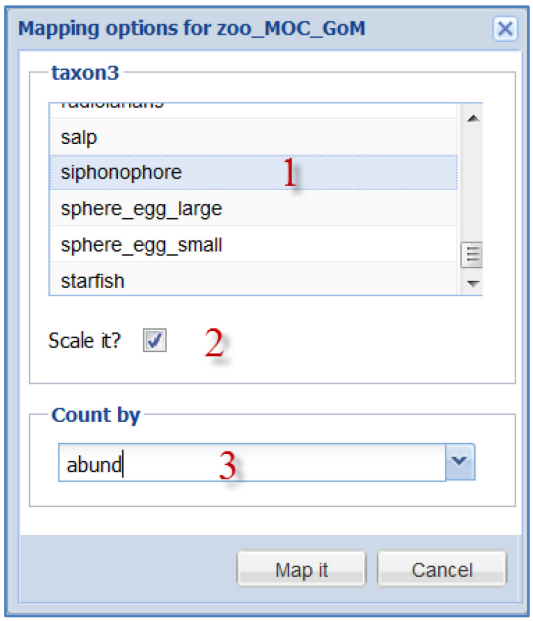


Select the options in the dialog box:

1. First, you must choose a taxon from the list. Select *Calanus finmarchicus*.
2. Select the "Scale it" option to display the results using symbol sizes scaled to indicate higher or lower values.
3. Last, you can choose "biomass" or "abund" (abundance) as the measurement by which to group the data. Select abund.

Click the "Map it" button to see the results.

The relative size of the circles indicates differences in the abundance of *Calanus finmarchicus* at the MOCNESS1 sampling locations on this cruise.

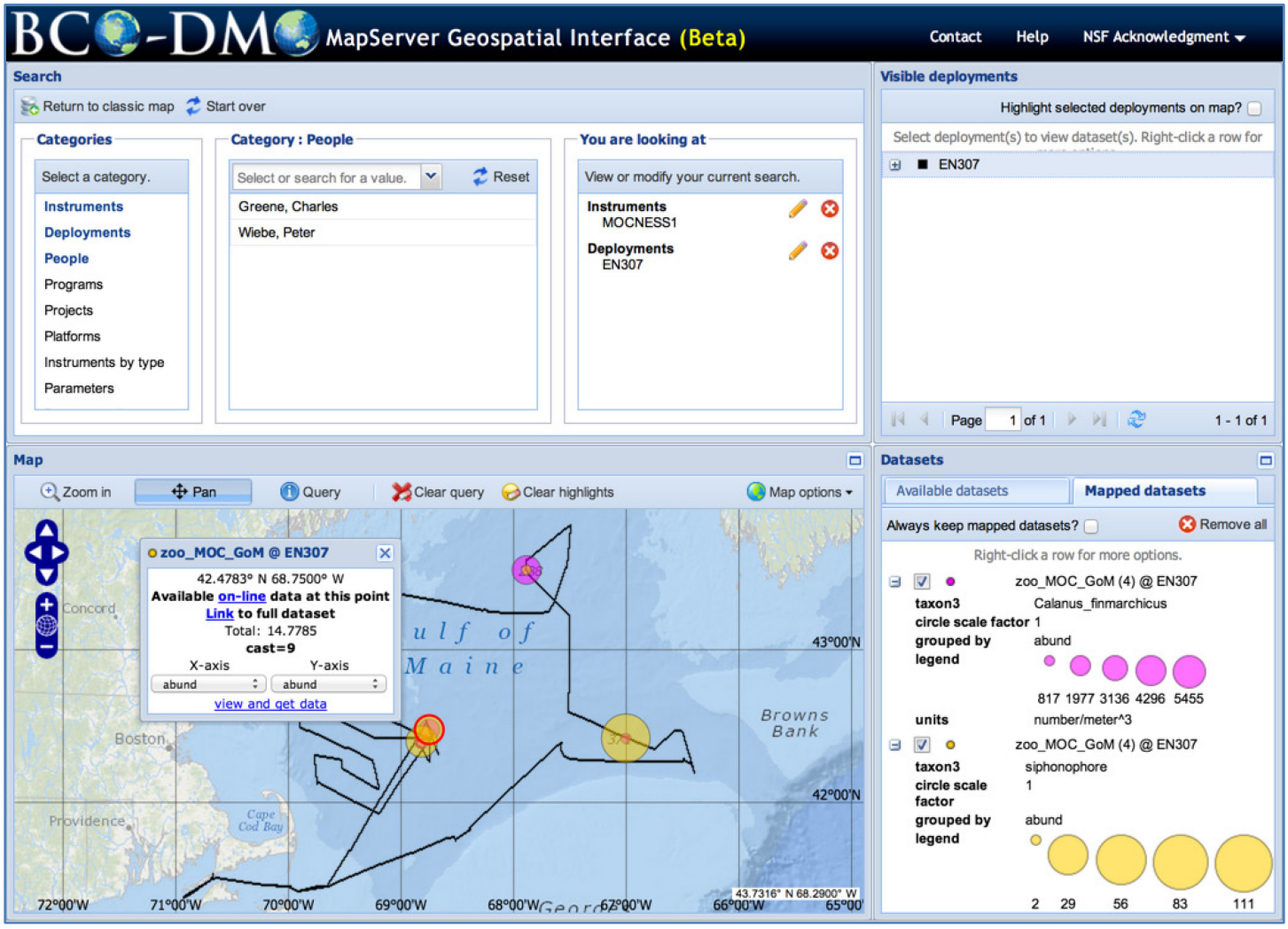


One can display multiple taxa from a MOCNESS1 dataset. Click on the "Available datasets" tab and then click on the zoo_MOC_GoM dataset once again. In the pop up dialog box, scroll to the bottom of the taxon list and select (1) siphonophore, (2) "Scale it", and (3) abund.

Click the "Map it" button to put the scaled symbols on the map indicating abundance of siphonophores from the MOCNESS1 tow data.

As with the other MapServer search modes, clicking or right-clicking on different elements in the display pops up dialog boxes that display either information about the item or additional options for that item (e.g. access to that selection in the database, or the option to do 'quick view' plots).

The final browser window looks something like this if you have clicked on a sampling location. To see data points covered by others, un-check the box in the Mapped datasets area located beside the minus sign. This removes visibility of that dataset.



Glossary of Terms

BCO-DMO	Biological and Chemical Oceanography Data Management Office http://bcodmo.org/
GIS	Geospatial Information System; a map system to display spatial data
MapServer	Open Source software for publishing spatial data and providing interactive mapping applications via the Web http://mapserver.org/
OCB	Ocean Carbon and Biogeochemistry research program http://www.us-ocb.org/
US GLOBEC	GLOBal Ocean ECosystems Dynamics research program http://www.usglobec.org/
US JGOFS	United States Joint Global Ocean Flux Study http://usjgofs.whoi.edu/

Acknowledgments

The MapServer interface is a custom implementation of the Open Source MapServer software developed at the University of Minnesota. The BCO-DMO MapServer system is the result of collaboration between programmer Charlton Galvarino (Second Creek Consulting, Columbia, SC) and BCO-DMO staff members. In particular, BCO-DMO marine biology data specialist, Dicky Allison has led this development effort. Funding for this work is provided by the National Science Foundation Division of Ocean Sciences and Division of Polar Programs.

Follow BCO-DMO:

info@bco-dmo.org

@BCODMO

https://www.youtube.com/user/BCODMO