

HYDROSEEK TUTORIAL

INTRODUCTION

HydroSeek is a data discovery tool that provides integrated federated search capability across multiple data sources (universities and federal agencies) that have WaterOneFlow web services through a single interface. It combines spatial, temporal and thematic aspects of search in order to make it possible to more effectively discover data despite heterogeneity issues among data repositories. It allows data discovery using keywords from an ontology which maps variables as defined by each data source to higher level searchable concepts. In short, HydroSeek is designed to provide a one-stop search center in which a user can query for specific data groups across a large data inventory regardless of the storage and description conventions used by the respective warehouses.

Currently supported data sources include:

- 1) National Data Sources: USGS National Water Information System (NWIS), EPA STORET (subset)
- 2) Regional Data Sources: Chesapeake Information Management System (CIMS), Texas Commission for Environmental Quality (TCEQ), regional subset (Chesapeake Bay) MultiSensor Precipitation Estimator (MPE), regional subset (Chesapeake Bay) National Air Deposition Program (NADP),
- 3) Ten local testbed sites distributed across the nation: Santa Fe Basin (FL), FerryMon (NC), Baltimore Ecosystem study (MD), Chesapeake Bay/ Susquehanna River (PA), Corpus Christi Bay (TX), Clear Creek (IA), Crown of the Continent (MT), Minnehaha Creek (MN), Little Bear River (UT), Sierra Mountains (CA).
- 4) State based data sources: IdahoWaters

A list of registered data sources that HydroSeek accesses can be found at the HIS Central website:

<http://hiscentral.cuahsi.org/>.

This tutorial introduces HydroSeek and shows how to download time series data.

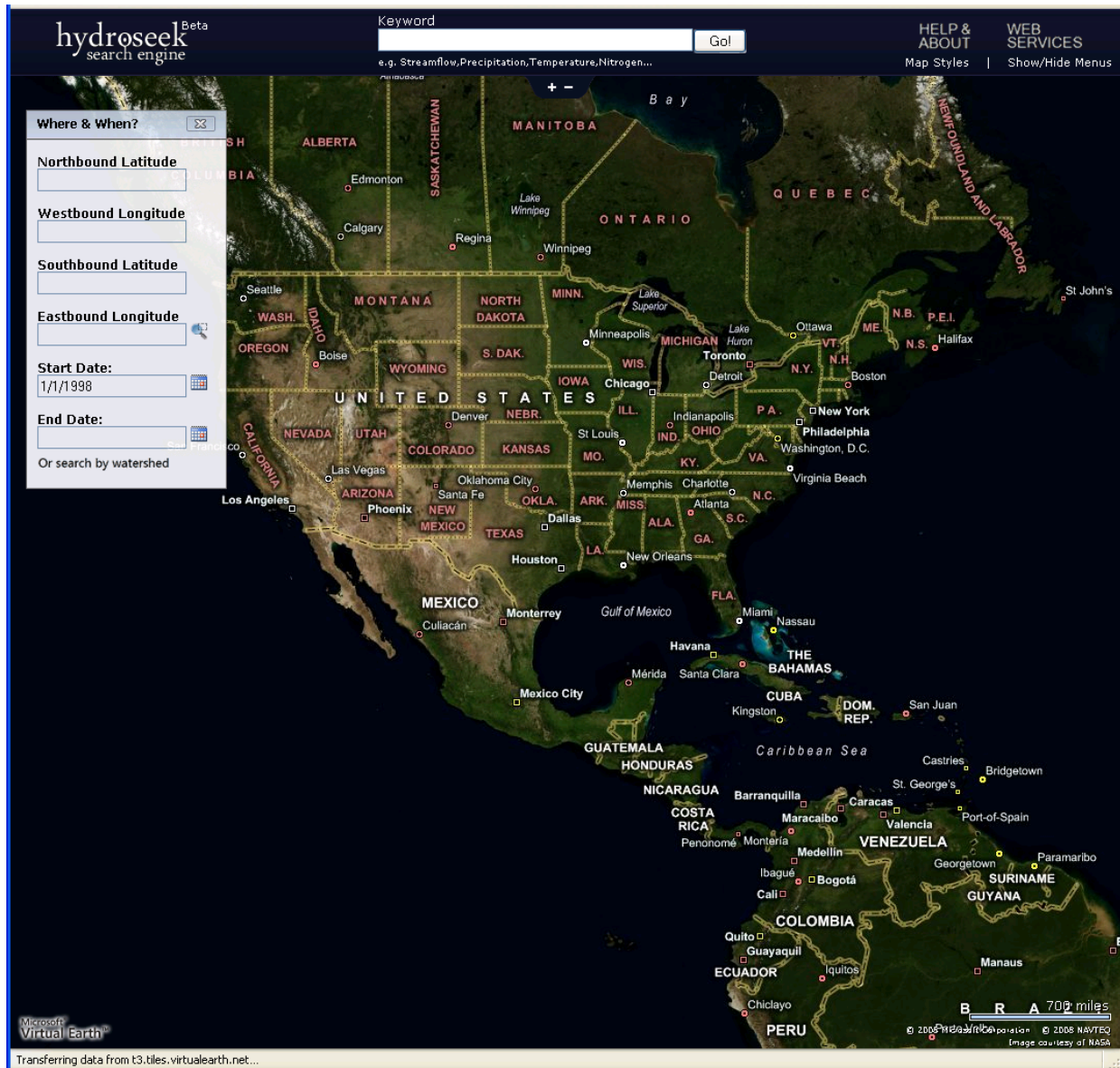
Links:

HydroSeek can be accessed at the following site:

<http://www.hydroseek.net/search>

PROCEDURE

1. Open a web browser and navigate to <http://www.hydroseek.net/search/>.

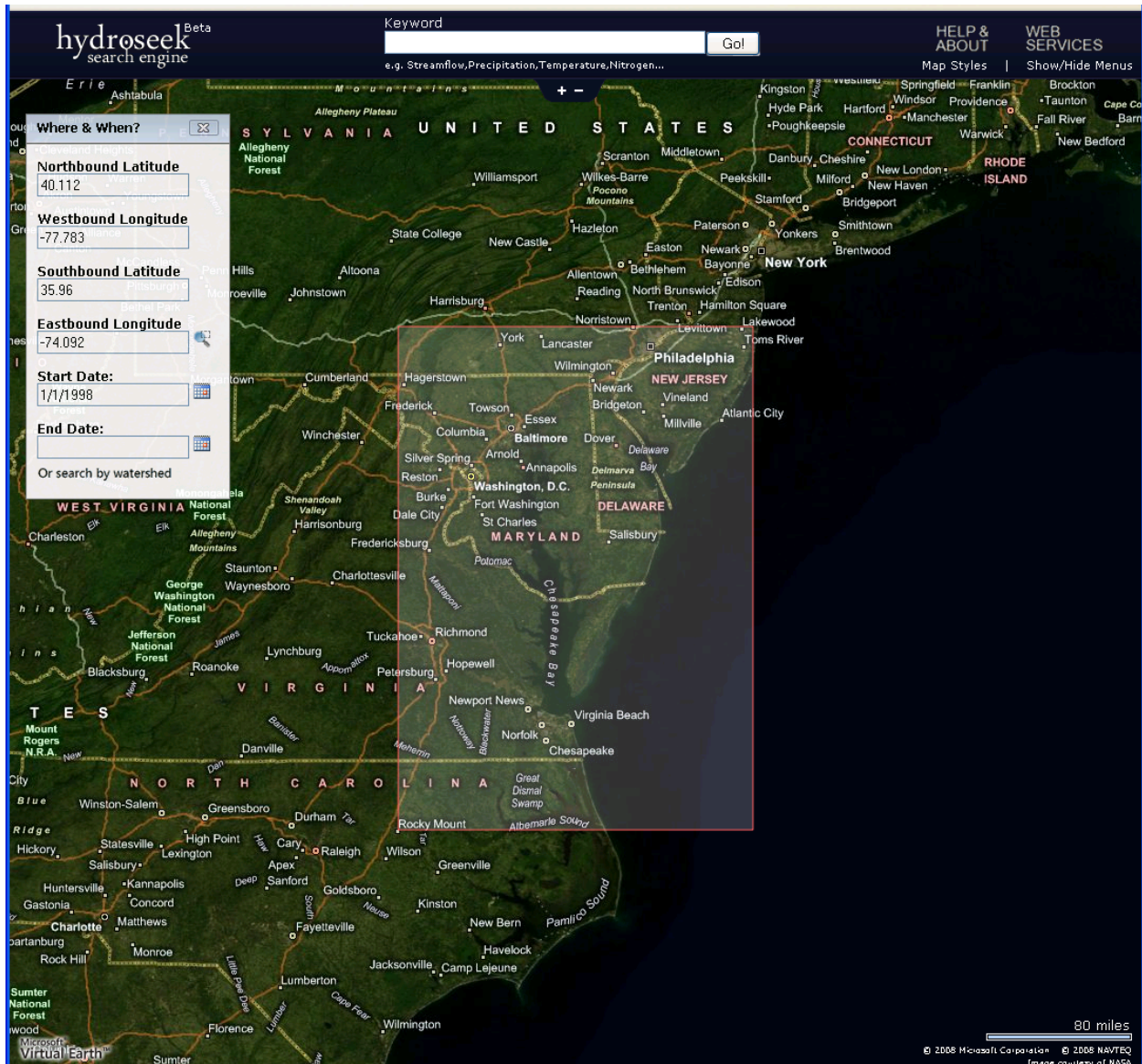


If you just click “Go!” HydroSeek will use its default settings to execute the search. The default settings comprise a search region around Chesapeake Bay (spatial) and the last 10 years (temporal) as well as an “unrecognized keyword” that will search for all data entries for the default region and time bracket.

2. In all likelihood, however, you would want to search with your own parameters. You can identify the spatial bounds using two methods:
 - a. Type in the longitude and latitude values directly or move the mouse over the small magnifying icon to the right of the “Eastbound Longitude” box and left click once. This activates the search bounding box definition via the mouse. Move your mouse to the upper left corner of the desired region and left click ONCE (and release afterwards), move to the lower right corner and again

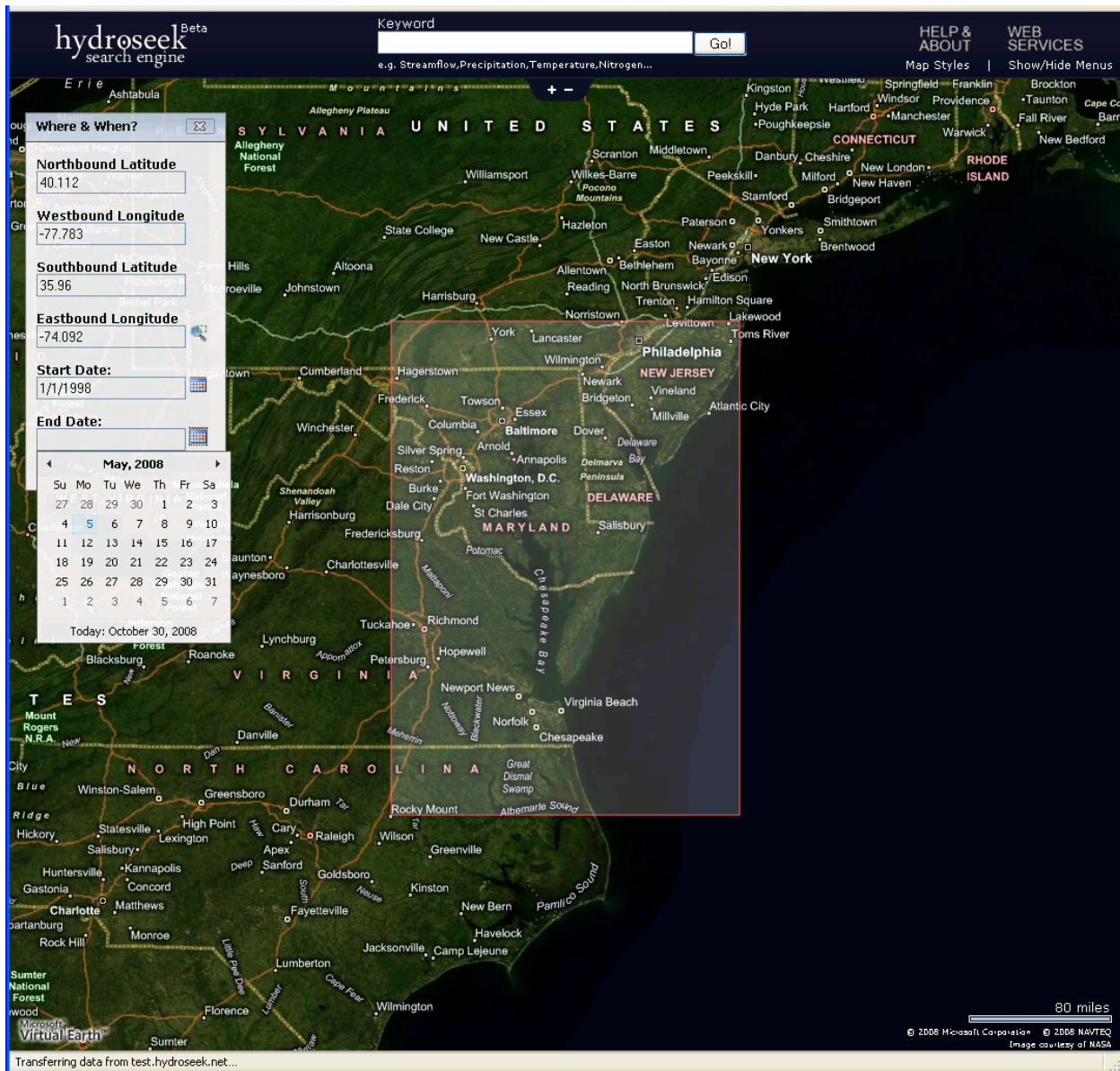
left-click ONCE. This defines the search bounding box and HydroSeek will also zoom the map to the region.

- b. Click the “Or search by watershed” in the “Where and When?” panel to the left. This will allow you to use a Watershed name to identify a search region (currently supports up to 8-digit HUC). There is an autocomplete feature, i.e., after typing a few letters HydroSeek will offer up what watershed names in its database matching those letters.

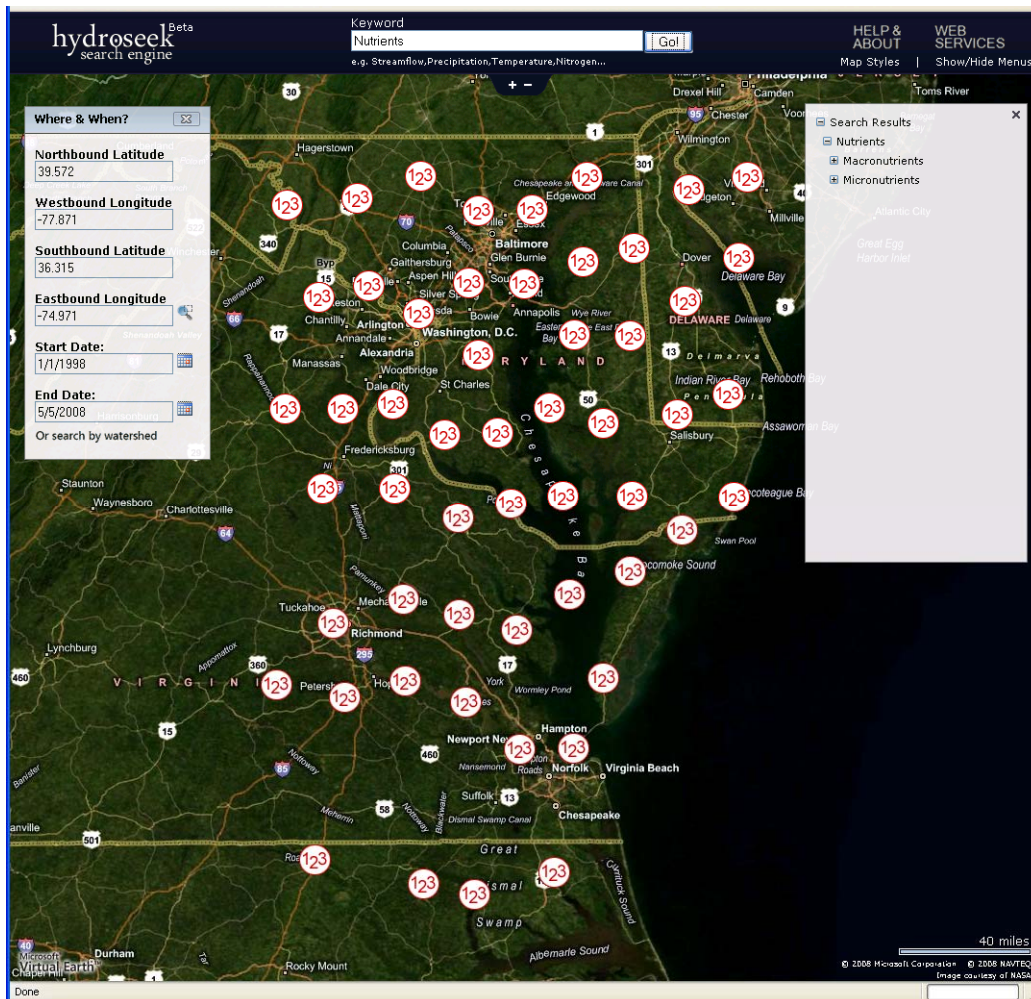


3. You will notice that a default start time of January 1st, 1998, has been entered. You can define your own time bracket by either typing different start and end dates (make sure to use the same format is displayed in the Start Date box), or using the calendar feature to the right of the date boxes. If there is a data point within the

specified time bracket HydroSeek will display the site icon even though the data set may not span the time bracket in its entirety.



4. Now we need a keyword to identify for which type of data to search. This makes use of the underlying search ontology which hosts concepts of various degrees of specificity. For example, you could use **Nutrients** (which is fairly general), **Nitrogen** (which is a little more specific), or **Nitrate (or Ammonia)** which is as specific a concept you could use. You could also use **Precipitation** or **Rainfall** which are synonyms for one and the same thing. HydroSeek understands these and regardless which one you use will spawn the proper search. Notice that there is also an autocomplete feature that is intended to help the user to find the "right" keyword. After you have inserted the keyword hit "Go!".



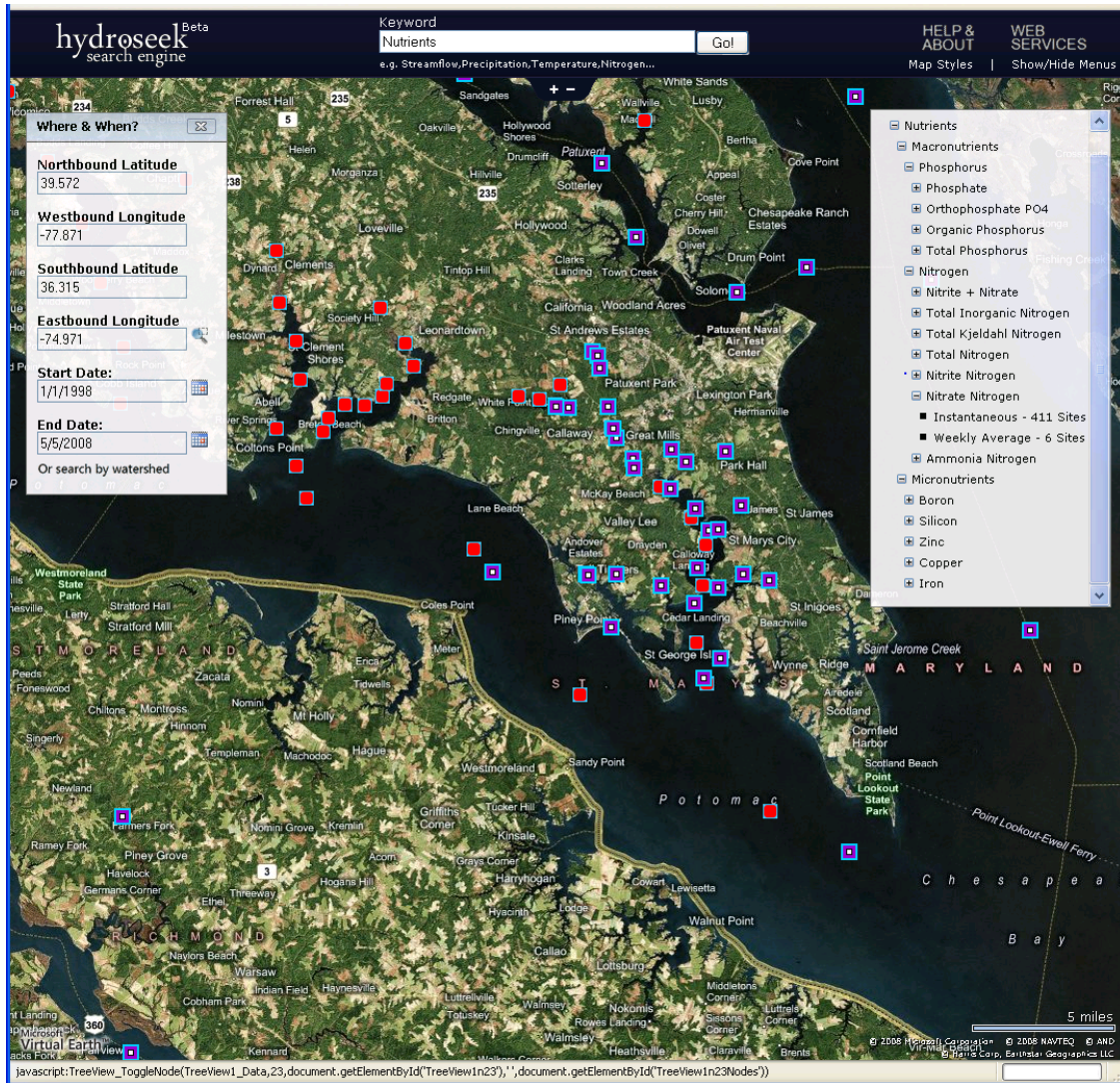
After hitting “Go!” and waiting for just a few seconds you will see something like the above image:

Notice the assembly of “123” icons ; these are “meta-stations”, i.e., each icon stands for a number of real sites. These icons are shown because the map interface can only display a limited number of icons at a time and also to prevent clutter in the user interface. You will need to use zoom in so you can see the actual sites.

Also notice the “Search Results” panel that has appeared on the right. As you can see, the results are returned in a classified fashion. This means that you can traverse the concept tree (the ontology) by expanding and collapsing the offered sub-concepts. The map will then change to display only those stations that actually have the data type just identified in the “Search Results” tree.

- In order to **zoom in** or **out** use either the “+ -” icon located at the top and center of the map, or use the scroll wheel on your mouse. Be careful not to zoom in too quickly as the GUI needs a little time to rebuild each zoom level. Zoom until the 123 icons disappear and you see the colored pins. Also notice in the figure below that the “Search Results” have been expanded to show **Phosphorus** and **Nitrogen** constituents as **Macronutrients** and **Boron, Silicon, Zinc, Copper, and Iron** as **Macronutrients**. You can click, for example, Nitrate Nitrogen-

Instantaneous and HydroSeek would only show the stations that hold this particular variable (notice; there should be 411 sites in the search region that have this type of data).



You will see only a subset of the sites found because of the current zoom level. But you can pan through the search region (hold down your left mouse button and drag the image) and inspect other places. In the image above there are apparently 2 data sources that have Nutrient data: the red squares are EPA STORET sites and the purple squares are Chesapeake Information Management System sites.

6. In order to learn more about the displayed sites, move your mouse over any one icon and observe a small pop up that invites you to take a look at the "Station Details". Left-click and see something like the following image.

The screenshot displays the hydroseek search engine interface. At the top, a search bar contains the keyword "Nutrients" and a "Go!" button. Below the search bar, a list of example keywords is provided: "e.g. Streamflow, Precipitation, Temperature, Nitrogen...". In the upper right corner, there are links for "HELP & ABOUT" and "WEB SERVICES", along with "Map Styles" and "Show/Hide Menus" options.

The main area features a map of the Potomac River region, showing various locations and data points. A "Station Details" popup window is open, displaying the following information:

- Site ID:** EPA:MDEDT04:XCE6420
- Site Name:** Breton Bay
- Latitude:** 38.2737
- Longitude:** -76.63425
- Elevation:** 0 m
- County:** ST MARY'S
- State:** MARYLAND

Below the site information, there are two dropdown menus for selecting variables:

- All variables measured at this station:** Carbon, organic
- Relevant variables measured at this station:** Nitrogen, inorganic

Each dropdown menu includes a "Download" icon (a floppy disk) and a "Data Cart" icon (a shopping cart). To the left of the map, a "Where & When?" sidebar provides search criteria for location and time, including Northbound Latitude, Westbound Longitude, Southbound Latitude, Eastbound Longitude, Start Date, and End Date. To the right, a "Nutrients" data cart is visible, listing various nutrient categories and their sub-categories.

You see some information about the site (in this case an EPA STORET site) and also two boxes that tell you about the data variables available at this station. The first box lists ALL variables that are measured at this station (even the ones you did not specifically ask for). The second lists variables that are connected to the search concept keyword you typed in earlier.

7. You can now either **download** the time series for a selected data variable directly (left-click the disk icon and get a compressed Excel file) or add the time series to the **data cart** for later download. You can move through the region and collect any data series you wish. Activate the data cart by clicking the "Show/Hide Menus" icon in the upper right. In the example shown below, two series each from an EPA site and a CIMS site have been added to the cart.

The screenshot shows the hydroseek search engine interface. At the top, the search bar contains the keyword "Nutrients" and a "Go!" button. Below the search bar, a list of example keywords is provided: "e.g. Streamflow, Precipitation, Temperature, Nitrogen...". The interface is divided into several panels:

- Where & When?**: A panel on the left with input fields for Northbound Latitude (39.572), Westbound Longitude (-77.871), Southbound Latitude (36.315), and Eastbound Longitude (-74.971). It also includes date pickers for Start Date (1/1/1998) and End Date (5/5/2008), and a search by watershed option.
- Station Details**: A central panel displaying information for the selected station:
 - Site ID: CIMS:LE2.3
 - Site Name: MOUTH OF POTOMAC RIVER; BOUNDARY BETWEEN CBS AND LE2; RIVER CHANNEL
 - Latitude: 38.0215149
 - Longitude: -76.3477249
 - Elevation: N/A
 - County: N/A
 - State: VIRGINIA
 Below this, there are two dropdown menus: "All variables measured at this station" (set to 'ACTIVE CHLOROPHYLL-A') and "Relevant variables measured at this station" (set to DISSOLVED INORGANIC NITROGEN).
- Data Cart**: A panel on the right showing a list of variables with their status:
 - EPA:MDEDAT04:XC6420 Carbon, organic **X**
 - EPA:MDEDAT04:XC6420 Nitrogen, inorganic **X**
 - CIMS:LE2.3 'ACTIVE CHLOROPHYLL-A' **X**
 - CIMS:LE2.3 DISSOLVED INORGANIC NI... **X**
 At the bottom of the Data Cart, there are "Download" and "Empty Cart" buttons.

The background is a satellite map of the Potomac River region, showing various locations and a 5-mile scale bar. The interface also includes a "HELP & ABOUT" and "WEB SERVICES" menu at the top right, and a "Map Styles" and "Show/Hide Menus" option at the bottom right.

Click "Download" and you will be asked to provide your delivery email. After a few minutes you should receive an email with a link to a zip file containing your data.

- If you would like to carry out a different search you can either zoom out and redefine a new search box and then go through the steps as outlined earlier or, in case you want to use the same region again but for a different variable, just retype a new keyword. For example, change the keyword to **Rainfall** and press ENTER. Change the keyword to **Streamflow**, and press ENTER.



There are three different types of rainfall data products. The first (yellow icons) are the MultiSensor Precipitation Estimator (MPE) NEXRAD based rainfall products from the NOAA National Weather Service. These are gridded data series; that is why the sites are aligned in a regular pattern. The results also tell you that these are daily averages. The blue triangle is a USGS NWIS site and the purple box is a National Air Deposition Program (NADP) site, both also collecting rainfall data.

This tutorial has demonstrated how to search for sites and variables from WaterOneFlow web services from a variety of data sources in a single map interface. For more information on HydroSeek, visit the HydroSeek page on the HIS website at <http://his.cuahsi.org/hydroseek.html>. For more on HIS in general, visit <http://his.cuahsi.org/>.