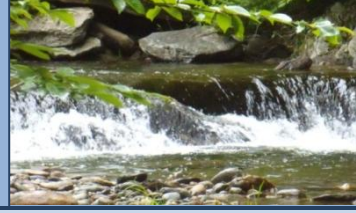
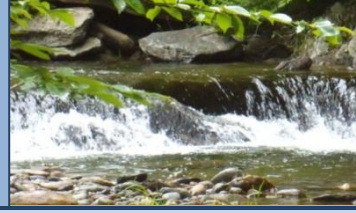


Module 3: Refining and Retrieving Data

Module 3



Refining and Retrieving Data

The Streams Project database holds all of the data from your field forms, the lab analysis of phosphorus, E.coli, Coliform and total suspended solids (TSS), and GIS-analysis of spatial data for each monitoring site and associated catchment area.

For your independent research you most likely don't want or need all of this data. This module helps you to do the following:

1) **Review the data in the database**

2) **Decide what data you need**

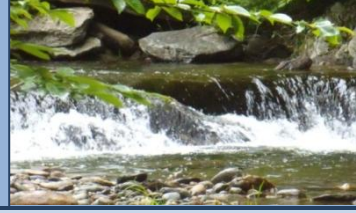
3) **Determine the time frame for which you'd like data**

4) **Determine the spatial extent for which you'd like data**

5) **Download the data you need from the Streams Project database**

6) **Refining your data**

Module 3



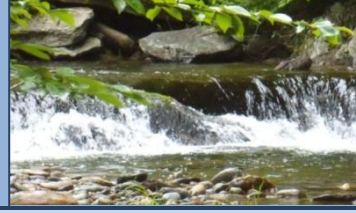
Refining and Retrieving Data

1) Review the data in the database

The database contains the following data discussed in module 2, available on the website in the form of “reports”:

- **Site Assessment Data**
- **Habitat Assessment Data**
- **Water Quality Assessment Data**
- **Macroinvertebrate Data**
- **E.coli and Coliform Lab Data**
- **TSS Lab Data**
- **Phosphorus Lab Data**
- **GIS Assessment Data**

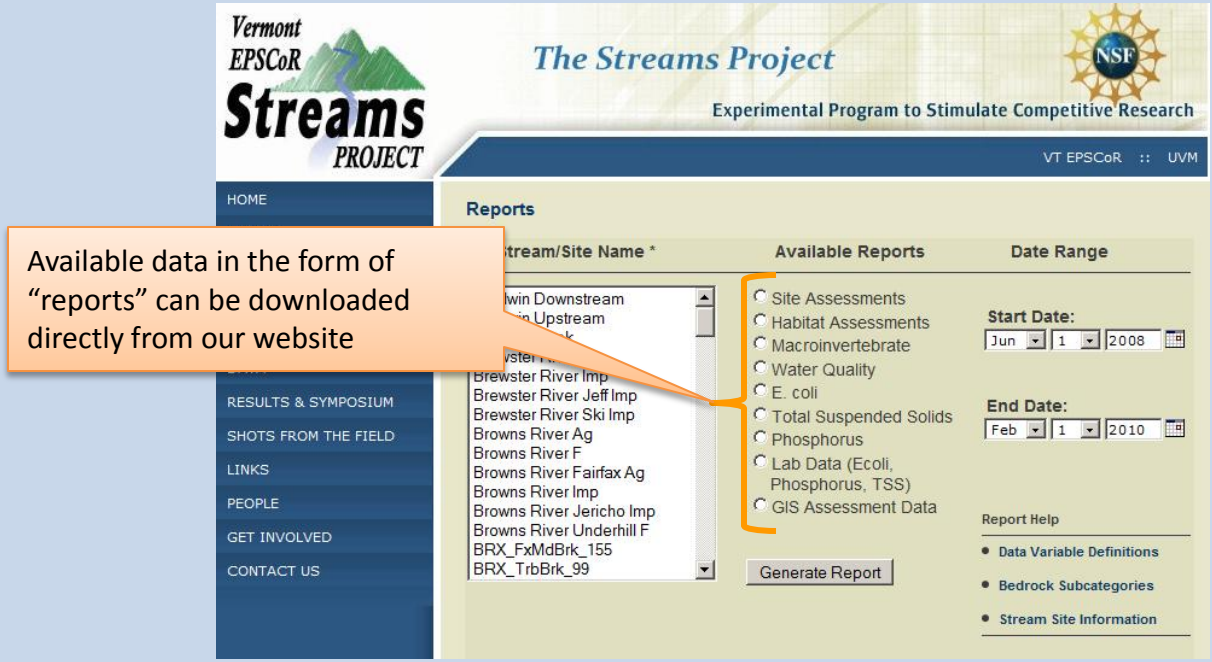
Module 3



Refining and Retrieving Data

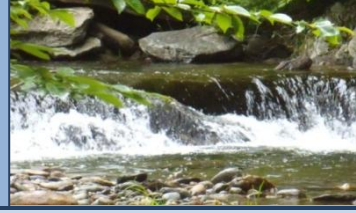
1) Review the data in the database

The following is the web interface through which you will download your data. All the data sources mentioned on the previous page are listed as “Available Reports”:



For more information about what is in each of these reports, download the “Data Variable Definitions” file in the bottom right hand corner of the data download webpage under “Report Help.” This file describes each field for each report including units of measurement.

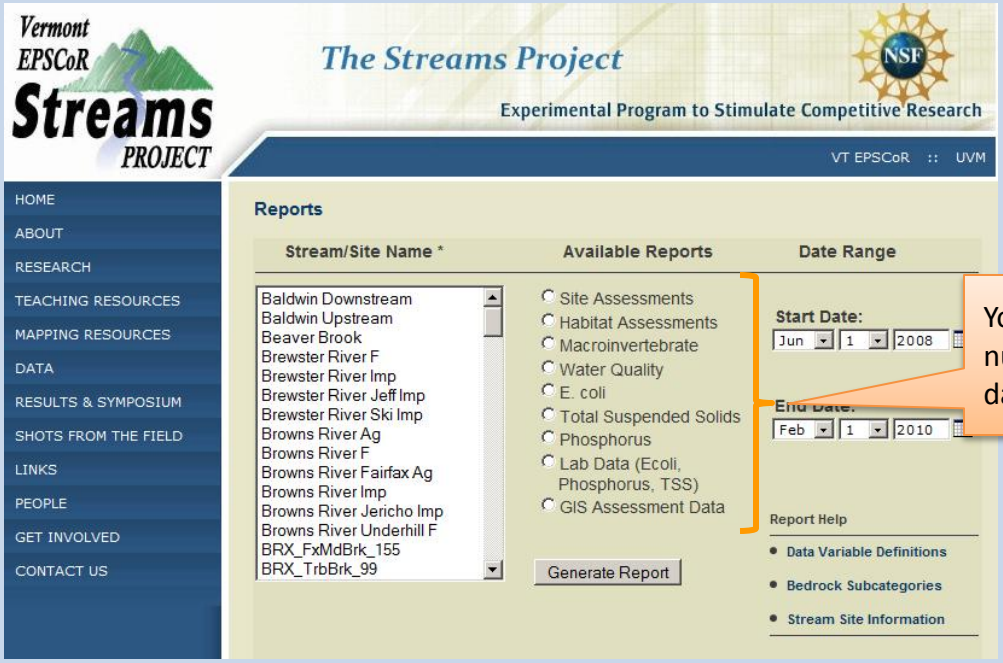
Module 3



Refining and Retrieving Data

2) Decide what data you need

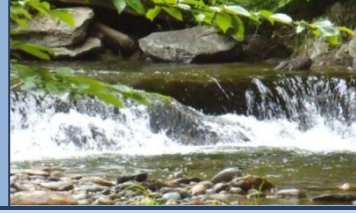
Most likely you will not need to download all of the available reports to do your data analysis. Think about your central question and in which report the data to answer this question is located.



You can download any number of reports to get the data you need

Again, For more information about what is in each of these reports, download the “Data Variable Definitions” file in the bottom right hand corner of the data download webpage under “Report Help.”

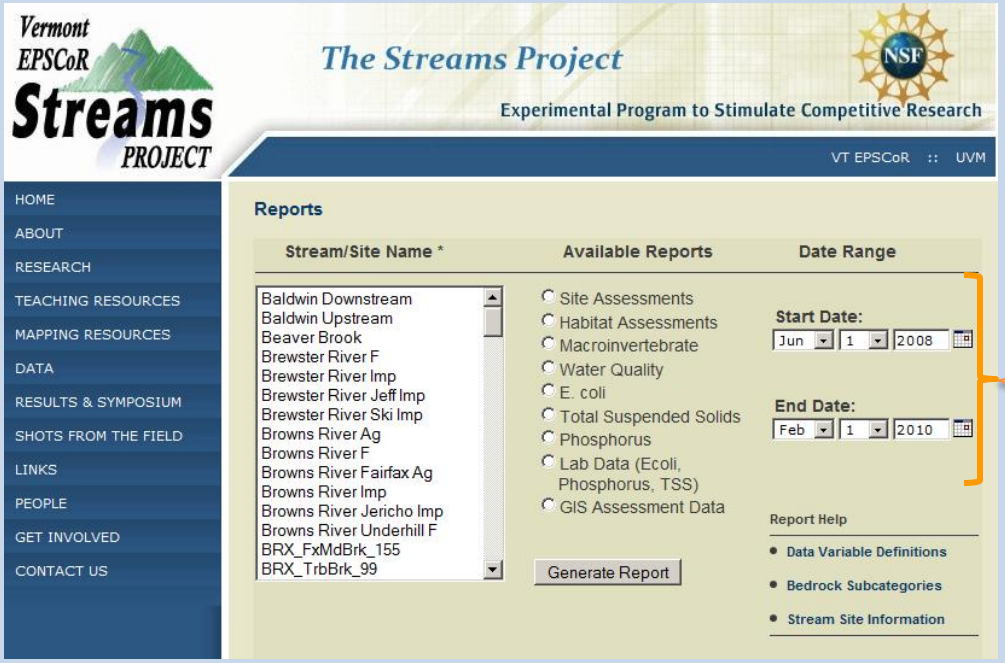
Module 3



Refining and Retrieving Data

3) Determine the time frame

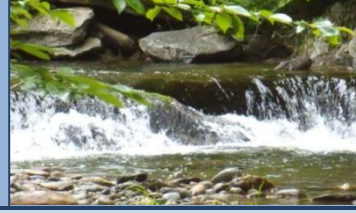
The database contains all data collected from June 2008 through the current date. As it is important to do throughout, review the question you are trying to answer – is a specific time interval best suited for answering this question?



You can specify the date for which you would like data here.

Most of this data was collected during June – December of each year. Think about if you want data from all years and all months or just a specific interval of time.

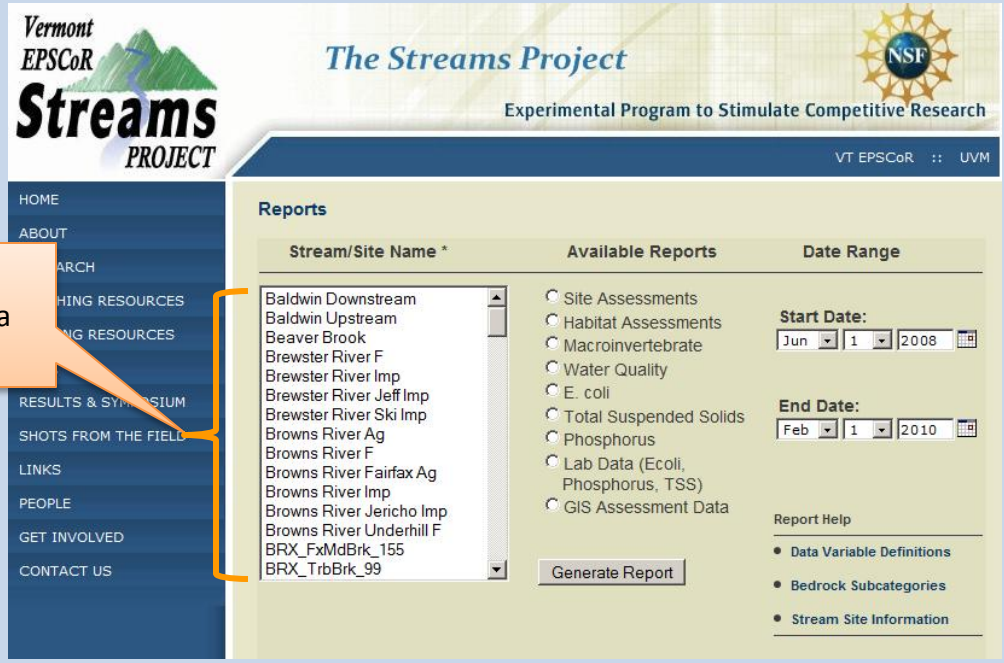
Module 3



Refining and Retrieving Data

4) Determine the spatial extent

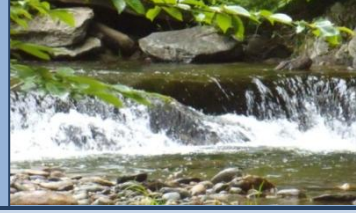
The database contains data from all streams sites being monitored in Vermont, New York, Connecticut and Puerto Rico. Use the Stream Sites Map to look up stream site names and help you narrow down the sites for which you would like data. Though you may decide you want to use all the data which is fine!!



Select which the sites for which you would like data here.

To select a stream site, highlight the name. To select all sites, highlight all the names. To select just a handful of sites, highlight the first and then hold down “Ctrl” while you highlight the remaining names.

Module 3



Refining and Retrieving Data

5) Download the data

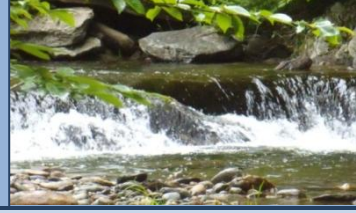
Now that you've reviewed the database and thought about the data you need, the time frame, and spatial extent for which you need this data, you are ready to download the data.

The following video will walk you through a quick example of how to download data using the data download website below:

WATCH DOWNLOAD VIDEO

Data download website: http://www.uvm.edu/~streams/?Content=pages/download_data8.inc

Module 3



Refining and Retrieving Data

6) Refining your data

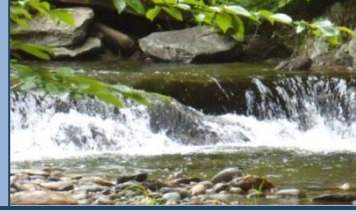
The data that you download from the website in many cases will not be ready for analysis as is. Please consider the following when preparing your data for analysis:

- You will note for E.coli, Coliform, TSS, and Phosphorus lab datasets, that all replicates numbers of data are included in the downloadable reports. You will have to refine these numbers, before using them in your analysis.

- You should be aware that there may be gaps because not everyone sampled on the same dates, so keep this in mind when you are compiling your dataset.

- Think about how your question might require that you structure or condense your data in a certain way. For example, if you want to regress land use and TSS, you might need to average your values of TSS for each site before running a regression analysis, as there will only by one land use value per site which is not dependent on the same time scale as TSS throughout your sample period.

Module 3



Refining and Retrieving Data

SUMMARY

- Data comes from field forms entered online, lab analysis on water quality samples gathered in the field, and GIS analysis done for the catchment area of each site.
- The following data is available as a "report" for download off the Streams Project website: Site Assessment Data, Habitat Assessment Data, Water Quality Assessment Data, Macroinvertebrate Data, E.coli Lab Data, TSS Lab Data, Phosphorus Lab Data, GIS Assessment Data
- Review your question to determine what data you need for your analysis
- Use the excel file Data_Field_Descriptions.xls to help you interpret the reports available for download.
- Consider the time interval for which you need data: all available data or a specific time period
- Consider the sites for which you'd like data: all available sites, or just a selection. Use the Stream Sites Map to help you determine which sites you'd like to focus on if not all.
- Download your data at http://www.uvm.edu/~streams/?Content=pages/download_data8.inc
- The data that you download is in a raw format, review guidelines in this module when refining your data for analysis