

Exercise 2: Organizing and Loading Data Into ODM

Jeffery S. Horsburgh¹
4-25-2012

Objectives

In this exercise you will learn how to organize and load data into an ODM database using the ODM Data Loaders.

Prerequisites

The computer that you will be using in today's exercise was set up ahead of time for the workshop. It has all of the software that you need to complete the exercises.

To complete this exercise on your own computer, you must first install the ODM Data Loader and ODM Streaming Data Loader Software and have completed Exercise 1 so that you have a blank ODM database set up in SQL Server that you can load data into.

Load Table Based Data into the ODM Database Using the ODM Data Loader

To demonstrate the capabilities of the ODM Data Loader, we will begin to populate your ODM database with information about the sites at which we have collected data, the variables that we have measured, the methods that we have used, etc. The ODM Data Loader was designed to load tables of data into an ODM database. Data tables to be loaded must conform to a set of acceptable templates that are documented in the ODM Data Loader software manual, which is available on the CUAHSI HIS Website (<http://his.cuahsi.org/odmdataloader.html>).

NOTE: Organizing your data so that it can be loaded into an ODM database is one of the most difficult steps in publishing your data using HydroServer. This process involves assembling all of the necessary information to populate the metadata within an ODM database and then formatting it into a set of files that can be loaded into the database. For the purposes of this exercise, we have done much of this work for you. You can look at the contents of the files that we have created to get an idea of the formats that the Data Loader can use.

For this exercise, a set of data files containing data for the Little Bear River Experimental Watershed has been prepared for you. Use the following steps to begin loading data into your ODM database:

1. Double click on the "workshopMaterials" folder on the desktop of your machine. Inside this folder, navigate to the "Wednesday/Horsburgh Data Files" folder. Inside this folder, you will find the data files for this exercise.
2. Open the ODM Data Loader by clicking Start → All Programs → CUAHSI HIS → ODMDL 1.1.6. The following window will appear:

¹ Utah Water Research Laboratory, Utah State University, Logan, UT 84322-8200, jeff.horsburgh@usu.edu

New Database Connection

Please select a Database:

Microsoft SQL Server

Server Address: (local)

Database Name: OD

Server User ID:

Server Password:

Test Connection Save Changes Cancel

3. Create a connection to your ODM database by supplying the information given in the previous exercise. Change your Server Address to “(local)”, change your Database Name to “OD”, change your Server User ID to “sa”, and change your Server Password to “Password123”. When your form looks something like the following, click the “Save Changes” button.

New Database Connection

Please select a Database:

Microsoft SQL Server

Server Address: (local)

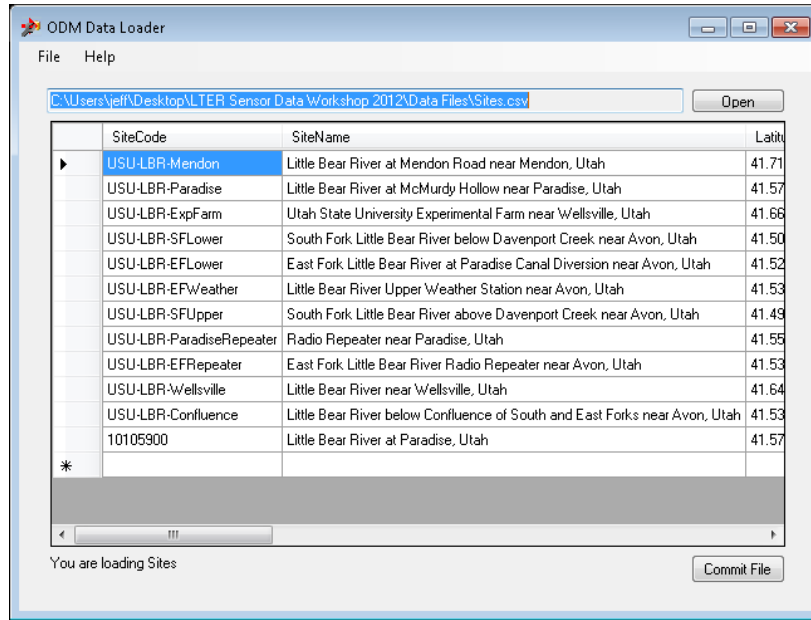
Database Name: OD

Server User ID: sa

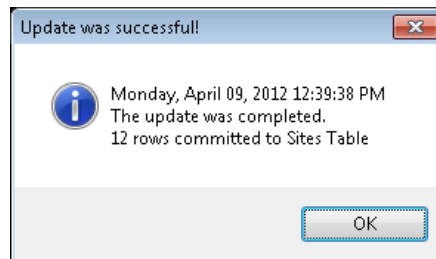
Server Password: Password123

Test Connection Save Changes Cancel

4. A small window indicating that your connection was successful will pop up. Click the “OK” button. In the form that opens, click the “Open” button to open a file dialog. Navigate to the “workshopMaterials/Wednesday/Horsburgh Data Files” folder on your computer’s desktop. Select the file called “Sites.csv” and then click “Open.” You will now see the file in the ODM Data Loader window.



- You will see at the bottom left of the form that the ODM Data Loader has figured out from the contents of the file that we are loading Sites into our ODM database. Click the “Commit File” button to load the sites into the ODM database. A small message window will pop up indicating the number of records that you have added to the database. Click the “OK” button to continue.



- Follow the steps above to load the “Methods.csv” table, the “Sources.csv” table, and the “Variables.csv” table.

NOTE: Now that you have added data to the Sites, Methods, Sources, and Variables table in your ODM Database, you can use SQL Server Management Studio to explore the contents of those tables. Although we have only loaded sites, variables, methods, and sources, you can load data into any of the tables in your ODM database using the ODM Data Loader, including DataValues.

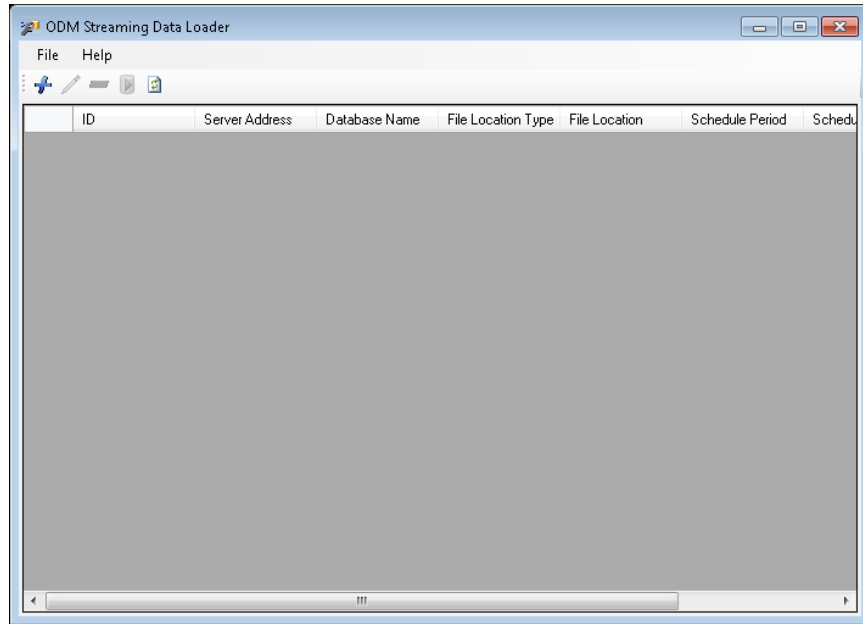
Load Data into the ODM Database Using the ODM Streaming Data Loader (SDL)

In this section, we will load some in situ sensor data collected in the Little Bear River into the ODM database using the ODM SDL. The SDL was specifically designed for loading data files that have one date/time column and multiple columns of data values, as is usually the case with datalogger files. This process will be simplified given that we have already loaded our Sites, Variables, Methods, etc. into our database.

The file that you will be loading contains continuous water quality data from one of the monitoring sites in the Little Bear River. It is a datalogger file created by a Campbell Scientific CR200 datalogger and contains 30 minute observations of battery voltage, water level, water temperature, dissolved oxygen concentration, specific conductance, pH, and turbidity.

Use the following steps to add the sensor data to your ODM database:

1. Open the ODM SDL by clicking Start → All Programs → CUAHSI HIS → ODM SDL 1.1.3 Config Wizard. The following window will appear:

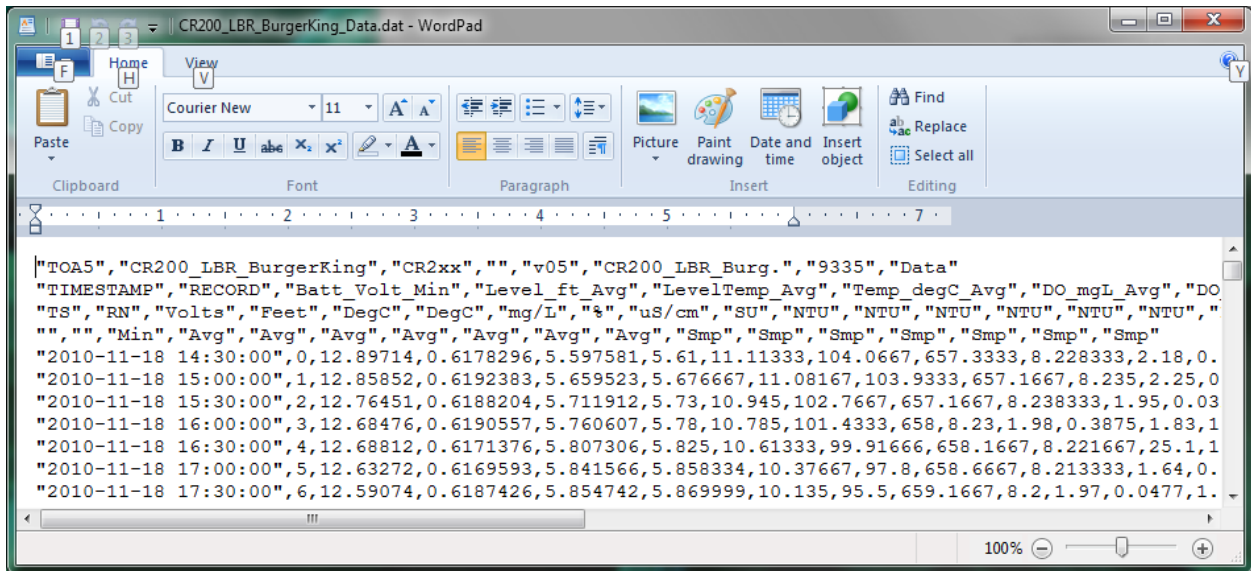


2. Click on the "Add" button (the blue plus sign icon at the top left of the form) to open the "Add New File" form. On this form, you will tell the SDL where the file is that will be loaded, which ODM database it will be loaded into, and some other information about the data file.

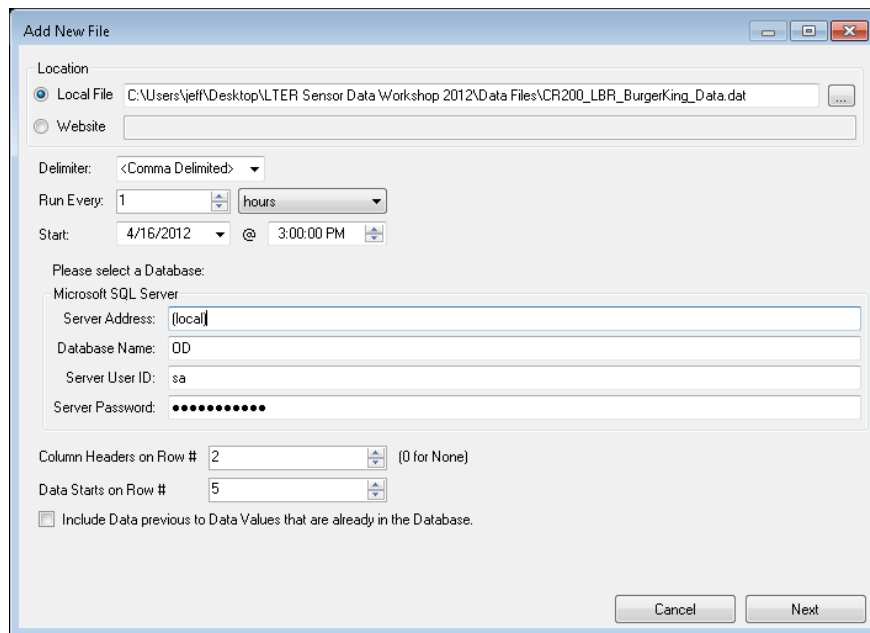
3. Click the button to the right of the “Local File” text input box to open a file browser. Navigate to the “workshopMaterials/Wednesday/Horsburgh Data Files” folder on your desktop and select “CR200_LBR_BurgerKing_Data.dat.” Click the “Open” button. You will notice that the path to your datalogger file is now shown on the form.

NOTE: The ODM Streaming Data Loader can load data from files stored locally on disk, from files available over a network share, or from a web or FTP site.

4. The file that you are loading data from is a .dat file from a Campbell Scientific CR200 datalogger. It is comma delimited and has a header at the top with information about the data. In the “Delimiter” drop down list, make sure the “Comma Delimited” is selected.
5. Next you will tell ODM SDL how often you want it to run. For this workshop, we will not be scheduling SDL to run automatically, but you can do that in the case where you are streaming new data into a datalogger file and want to load it periodically from the file to the ODM database. We will tell ODM SDL to run every 1 hour. Leave the starting date as it is.
6. Now we need to tell ODM SDL how to connect to your ODM database. Use the database connection information in the section above to create a connection to your ODM database.
7. ODM SDL also needs some information about the header of the file. The figure below shows the first few lines of the file that you are loading. The column headers are on row number 2, and the data start on row number 5. Set these values in the appropriate text boxes in the SDL window.

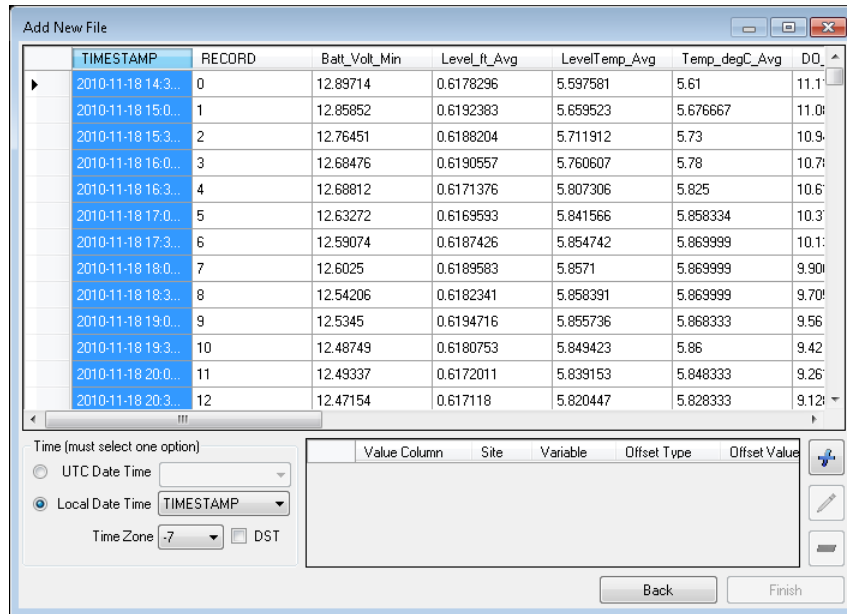


- The last option on this form is to tell SDL whether it should look for data in your data file that predate any data you have in the ODM database. Since we are starting with a blank database, we will leave this option unchecked. Your form should now look like the below. Click the “Next” button to continue.

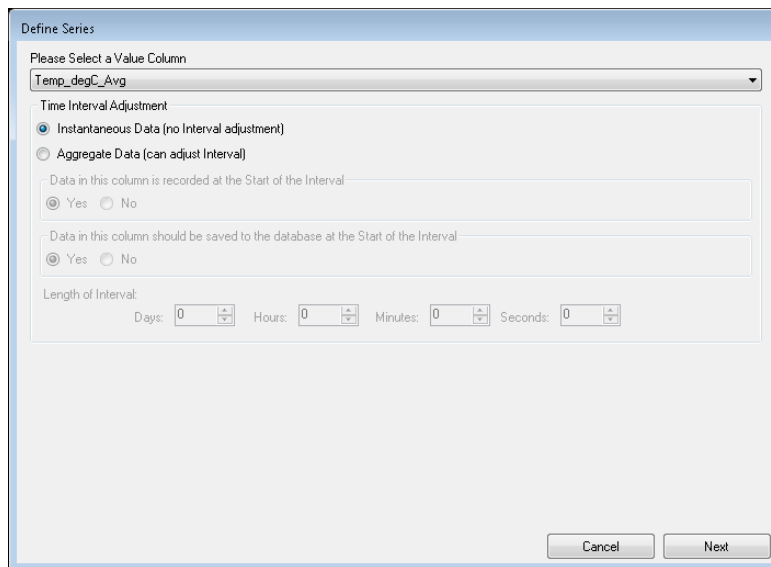


- You will now notice that the SDL has loaded the data file into a table on the form. The next step is to tell SDL which column contains the time stamp. Click the radio button next to “Local Date Time” at the bottom left corner of the form and then make sure that “TIMESTAMP” is selected in the drop down list. This is the name of the column containing the time stamp for the data values.

10. In the “Time Zone” drop down list, select “-7.” This datalogger is programmed to run on U.S. Mountain Standard time and does not change for Daylight Saving Time. Your window should now look like the following:



11. Now we need to map each of the columns of data to the appropriate site, variables, methods, etc. in the ODM database. To do this, we will select each column and map them individually. In the interest of time, we will skip to a couple of the water quality variables that might be interesting. Select the “Temp_degC_Avg” column by clicking on its header at the top of the window. Then click the “Add” button (the blue plus sign near the bottom right of the form). The following window will open:



12. The options on this window will allow you to change the way the time stamps of your data are recorded in the database. For example, if your datalogger records data at the end of each time

interval, but you want to record the data in the database at the beginning of the interval, you can set that up here. For this exercise, we will accept the default of no interval adjustment. Click the “Next” button to continue to the next window.

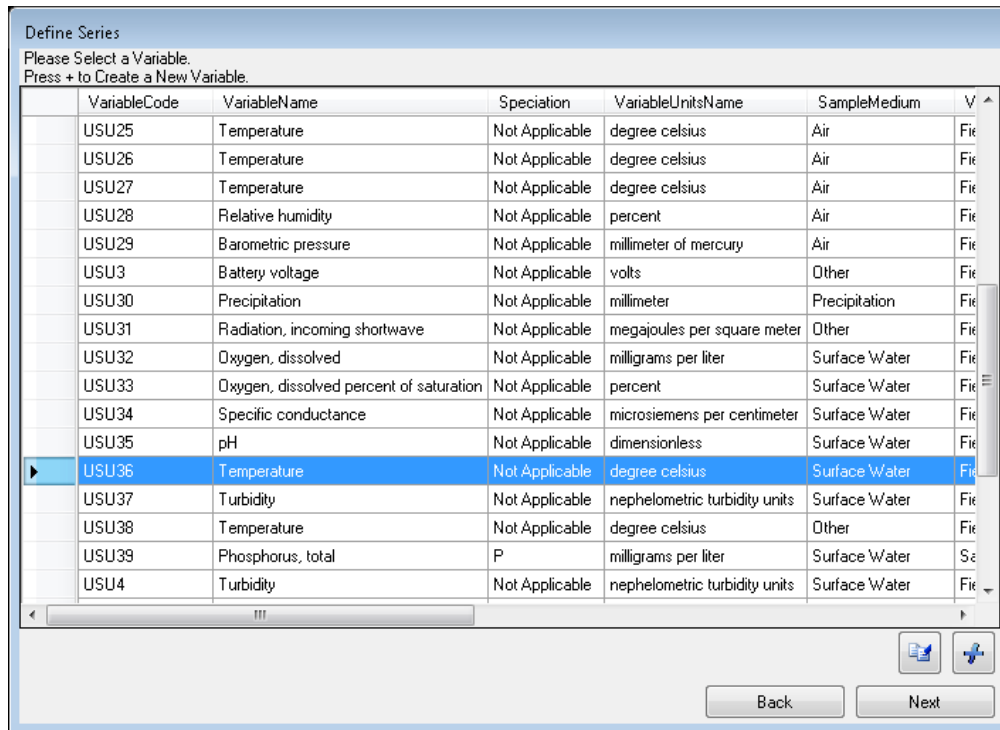
- These data were collected at the site having the site code “USU-LBR-Wellsville.” Since we loaded our sites previously, we can select it from the list. If the site were not already in the database, we could create it on the fly by clicking the “Add” button at the bottom right of the form (the blue plus sign). Select the “USU-LBR-Wellsville” site by clicking on its row in the table and then click the “Next” button.

The screenshot shows a dialog box titled "Define Series" with the instruction "Please Select a Site. Press + to Create a New Site." Below this is a table with the following data:

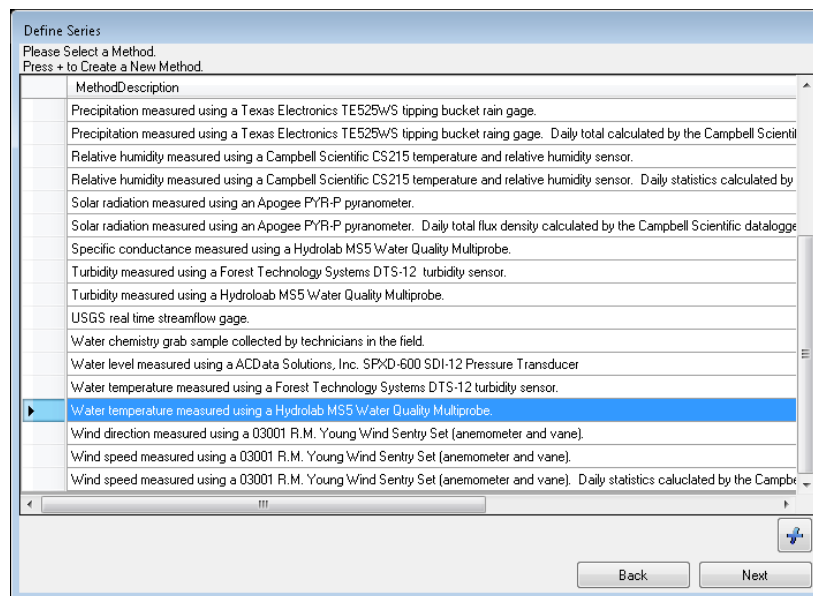
SiteCode	SiteName	Latitude	Longitude
10105900	Little Bear River at Paradise, Utah	41.5756	-111.8544
USU-LBR-Confluence	Little Bear River below Confluence of South and East Forks near Avon, Utah	41.536088	-111.830455
USU-LBR-EFLower	East Fork Little Bear River at Paradise Canal Diversion near Avon, Utah	41.529212	-111.799324
USU-LBR-EFRepeater	East Fork Little Bear River Radio Repeater near Avon, Utah	41.536	-111.806
USU-LBR-EFWeather	Little Bear River Upper Weather Station near Avon, Utah	41.535543	-111.805946
USU-LBR-ExpFarm	Utah State University Experimental Farm near Wellsville, Utah	41.666993	-111.890567
USU-LBR-Mendon	Little Bear River at Mendon Road near Mendon, Utah	41.718473	-111.946402
USU-LBR-Paradise	Little Bear River at McMurdy Hollow near Paradise, Utah	41.575552	-111.855217
USU-LBR-ParadiseRepeater	Radio Repeater near Paradise, Utah	41.556736	-111.854455
USU-LBR-SFLower	South Fork Little Bear River below Davenport Creek near Avon, Utah	41.506518	-111.815079
USU-LBR-SFUppper	South Fork Little Bear River above Davenport Creek near Avon, Utah	41.495409	-111.817993
USU-LBR-Wellsville	Little Bear River near Wellsville, Utah	41.643457	-111.917649

At the bottom of the dialog box, there are two buttons: "Back" and "Next". A blue plus sign icon is also visible in the bottom right corner of the table area.

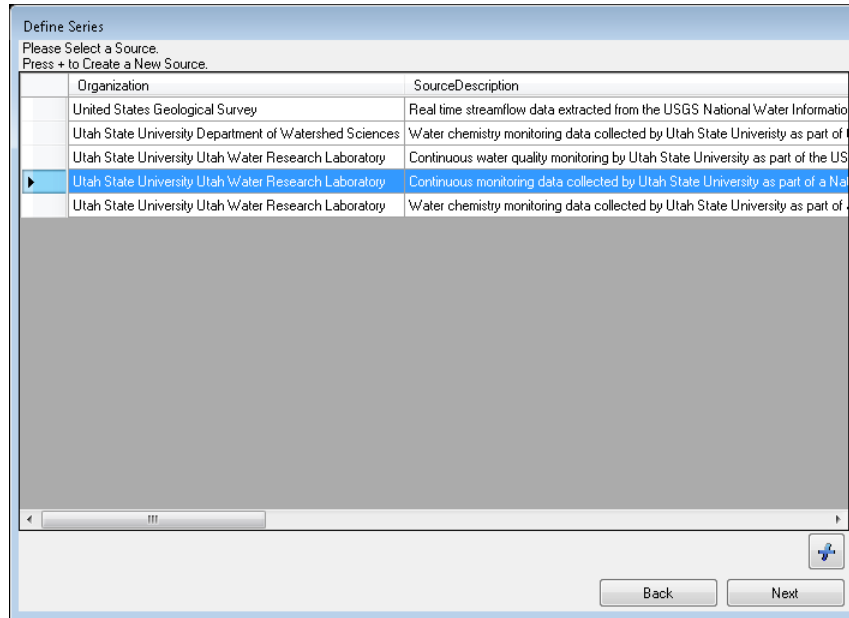
- This particular column in the datalogger file contains 30 minute average water temperature values. Again, we already loaded our variables, so we need to find the right one in the list. Scroll down and select the variable with VariableCode “USU36” (see the first column in the table) by clicking on its row in the table. You will notice if you look at the attributes of this variable that it represents field observations of temperature, in degrees Celsius, measured in surface water, with a time support of 30 minutes, and that the values are averaged over the time support. Again, if we hadn’t already loaded our variables, we could create a new variable on the fly here by clicking the “Add” button. Click the “Next” button to continue to the next window.



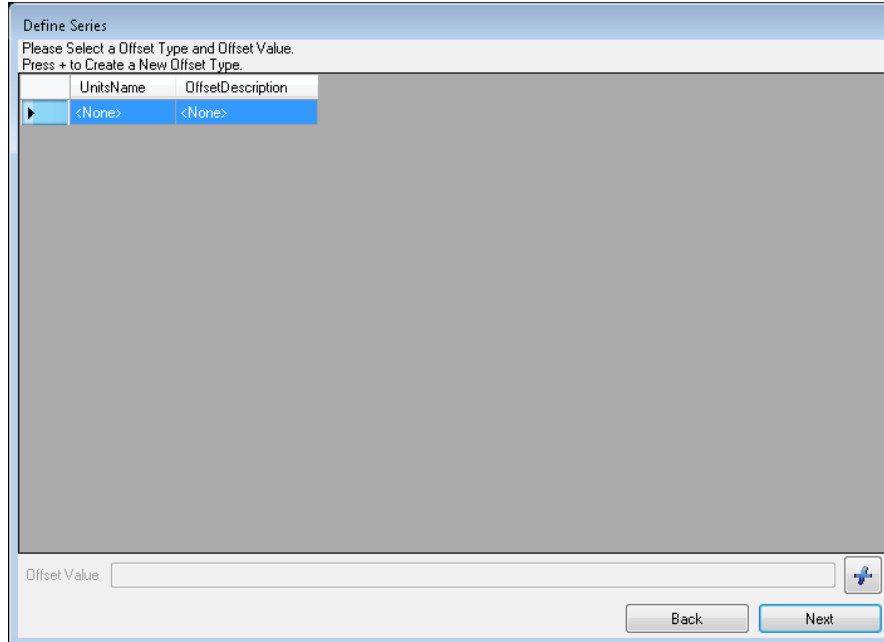
15. The temperature values were measured using a HydroLab MS5 Water Quality Multiprobe. Find and select the appropriate method in the list of methods (it's near the bottom of the list). Click the "Next" button to continue to the next window.



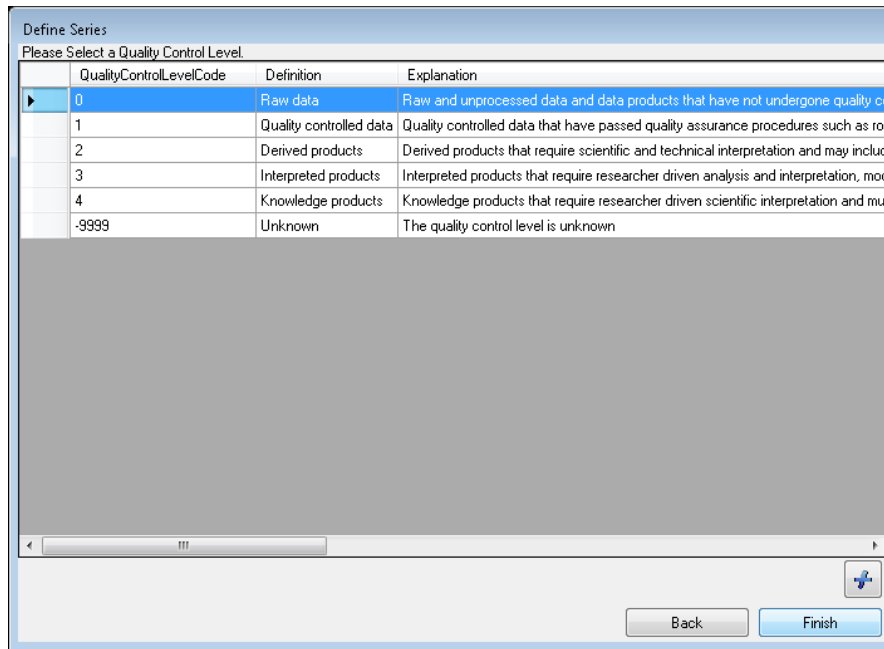
16. Now we need to select the source of the data from the list of sources in the database. These data were collected by Utah State University as part of an NSF-funded project, which corresponds to the fourth item in the list of sources. Select the appropriate source from the list and click the "Next" button to continue to the next window.



- On the offsets form, you can define an offset for your data in the case where it was collected with a constant offset – e.g., an air temperature sensor at a fixed location above the ground, or a soil moisture sensor at a fixed location below the surface of the soil. The offset is not really relevant for the water temperature data, so we will accept the default of “None” and click next to move to the next window.



- On the final screen of the Wizard, you will need to select a QualityControlLevel for your data. These are raw sensor data streaming in from the datalogger in the field, so we will select a QualityControlLevel of 0 (Raw Data) for these data. Select the appropriate record in the table and then click the “Finish” button.



19. Congratulations! You have now fully mapped one of the data columns in your file to be loaded into ODM. You will notice that a new record has been added to the table at the bottom right of the SDL window for the column that you just mapped.

NOTE: This process can be a bit laborious for large data files with many columns. However, it only has to be done once, and once a file is mapped, the SDL can load new data from that file to your ODM database on demand or on a schedule that you set.

20. If you want to, you can map one or more additional columns of data from the file to be loaded into the database using the same process. The following are the details for a couple more columns that you can map. ***If you are short on time, you might want to skip mapping additional columns of the file.***

Data File Column Name: DO_mgL_Avg

SiteCode: USU-LBR-Wellsville

VariableCode: USU32

MethodDescription: Dissolved oxygen measured using a HydroLab MS5 Water Quality Multiprobe.

SourceDescription: Continuous monitoring data collected by Utah State University as part of a National Science Foundation funded test bed project.

Offset: None

QualityControlLevelCode: 0

Data File Column Name: DO_Perc_Avg

SiteCode: USU-LBR-Wellsville

VariableCode: USU33

MethodDescription: Dissolved oxygen measured using a HydroLab MS5 Water Quality Multiprobe.

SourceDescription: Continuous monitoring data collected by Utah State University as part of a National Science Foundation funded test bed project.

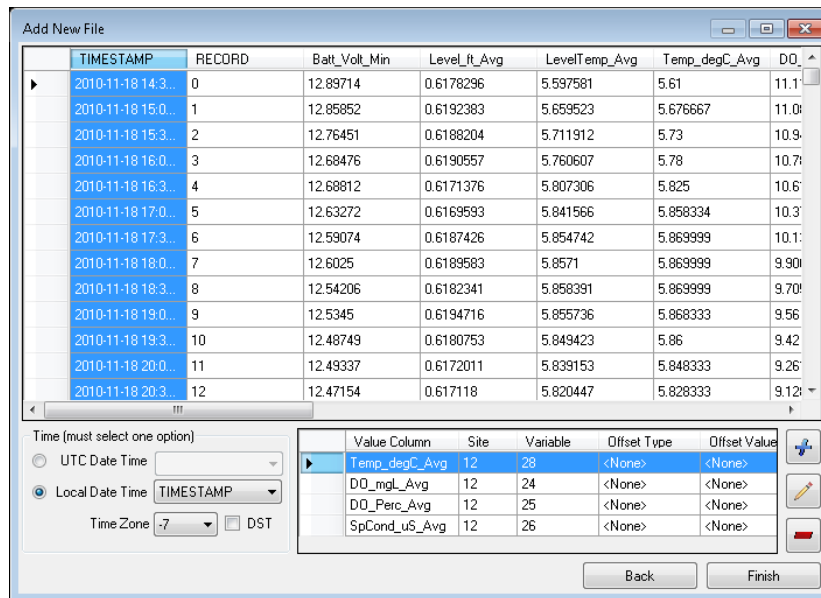
Offset: None
QualityControlLevelCode: 0

Data File Column Name: SpCond_uS_Avg
SiteCode: USU-LBR-Wellsville
VariableCode: USU34

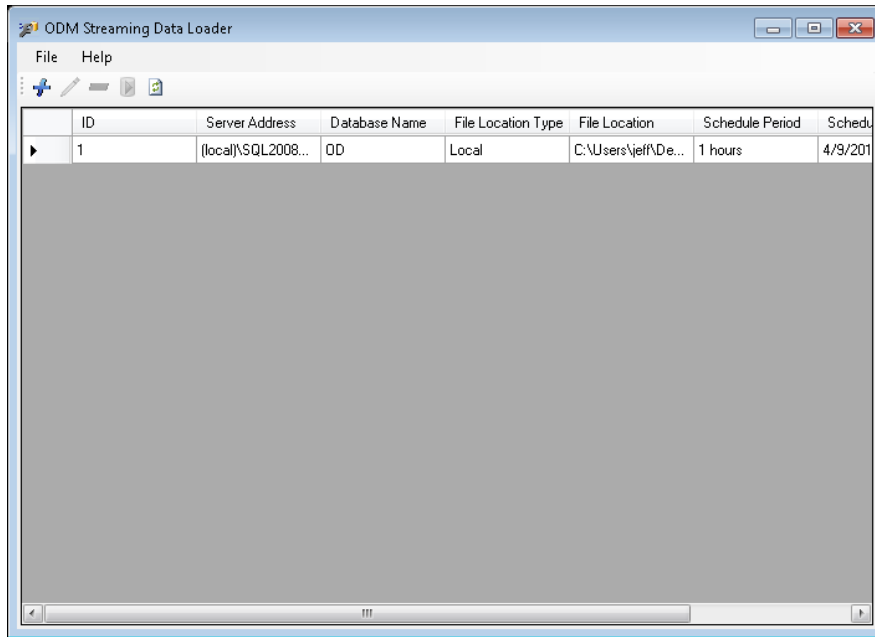
MethodDescription: Specific conductance measured using a HydroLab MS5 Water Quality Multiprobe.
SourceDescription: Continuous monitoring data collected by Utah State University as part of a National Science Foundation funded test bed project.



Offset: None
QualityControlLevelCode: 0

21. Now that you have mapped one or more columns in your data file to be loaded into your ODM database, your screen should look something like the following.



22. Click the “Finish” button on the “Add New File” Window to return to the main SDL window. It should now look something like the following. There will be a new record in the table for the file that you just configured.



23. You can now manually execute the SDL by making sure that the file that you just configured is selected in the table view (click on the row) and then clicking on the execute button on the toolbar . This will load all of the data for the columns that you just mapped into your ODM database. While the SDL is running, you will notice a small database icon that pops up in your Windows task manager. . When the database icon disappears, SDL has finished loading your data.

NOTE: The mappings that you just input to the SDL are stored on your computer's hard drive in a configuration file. The SDL will open this configuration file each time you open the software. You can add new files or edit existing files at any time. You can also schedule the SDL to run automatically on a predefined schedule in the case that your data files are receiving new data from the field on a regular basis. Instructions on how to do this, as well as advanced functionality available in the ODM SDL are described in the SDL software manual, which is available on the CUAHSI HIS website (<http://his.cuahsi.org/odmsdl.html>).