PUBLISHING DATA WITH THE TEXAS HYDROLOGIC INFORMATION SYSTEM

Presented in Austin, TX

May 27, 2009

by:

Tim Whiteaker (twhit@mail.utexas.edu)
Center for Research in Water Resources
The University of Texas at Austin

Distribution

Funding and Acknowledgements

Funding for this workshop was provided by the Texas Water Development Board. In addition, much input and feedback has been received from the CUAHSI Hydrologic Information System development team. Their contribution is acknowledged here.

Table of Contents

INTRODUCTION

This document provides steps to complete the hands-on training portion of a workshop that teaches how to publish water observations data using the Texas Hydrologic Information System (HIS). This involves loading data into an Observations Data Model (ODM) database, and then exposing the data in a secure and standard way online via WaterOneFlow web services. Technologies such as ODM and WaterOneFlow were originally developed for the CUAHSI HIS, and have been adopted for use with Texas HIS.

For background information on CUAHSI or HIS, please refer to presentation materials provided at the workshop, or the HIS website at http://his.cuahsi.org.

Why publish data online?

Sometimes people ask what the motivation is for using HIS to publish data. When the HIS team interacts with HIS users, here are the most common reasons those users give for why they want to publish their data:

- Academics
  - Recognition of work
  - Data publication is mandated by the funding agency
  - To support science in the US and promote collaboration
- Agencies
  - Standardizing data access (both internally and externally)
  - Time savings in developing a publication system
  - Public benefit with publication
  - Return on investment – people can get the data themselves without requiring a "middle-man"
  - Get all the state data "together"

If you have other reasons for publishing data, please let us know! Fulfilling your needs is a primary driver in the future development of Texas HIS.

GOALS OF THE WORKSHOP

This workshop seeks to introduce you to the **HIS data publication process**. As there are numerous avenues for publishing data, which often depend on a given user's available software and system setup, this workshop does not seek to teach every technique that can be used for publishing data. Rather, the workshop communicates the basic concepts of data publication, which can then be applied by you to fit your specific needs or environment. After completing the workshop, you should be able to:

- Understand what kind of data can be stored in an Observations Data Model database
- Translate your observations data and metadata into terminology used by the Observations Data Model
- Load your data into an Observations Data Model database
- Publish your data with a WaterOneFlow web service
- Access your data in a number of ways using HIS software:
  - Direct database connection with ODM Tools
  - Direct web service access with HydroExcel
- Learn more about HIS using the HIS website at http://his.cuahsi.org

WORKSHOP REQUIREMENTS

Computers and sample data were prepared ahead of time for the workshop. However, all HIS software is free, and so the configuration below could also be applied to your own computer if you had licenses for the commercial software used, such as the operating system. The system outlined below closely resembles the HIS Server Lite system, as described at http://his.cuahsi.org/hisserver.html#lite.

Hardware:

- PC, 4 GB RAM, 73 GB hard disk, Dual 2.8 GHz Xeon Processors

Software:

- Windows XP Professional SP3
    - Microsoft Internet Information Services (IIS) — comes as part of XP Professional or 2003 Server, but is not installed as part of the default operating system install, and so usually must be installed separately. Once IIS is installed, if "Default.aspx" doesn't appear as one of the Document tabs for the properties of Default Web Site in IIS, add it.
- .NET Framework 2.0 SP2, 3.5 SP1 (free)
- Microsoft SQL Server 2005 Express (free)
    - Be sure to install the SQL Management Studio
    - Install with these options:
        - Install as Local Service
        - Install with Mixed Mode Authentication
        - Set up an "sa" (system administrator) account with a password that you will remember
- HIS software (free)
    - ODM Data Loader 1.1 - http://his.cuahsi.org/odmdataloader.html
    - ODM Tools 1.1 - http://his.cuahsi.org/odmtools.html
    - HydroObjects - http://his.cuahsi.org/hydroobjects.html
- Optional software (not required to run HIS Server)
    - Microsoft Office 2007
    - Google Earth 5.0 (free) - http://earth.google.com/download-earth.html

Data (located in C:\HIS Training\):

- Raw data files of water quality time series
- Metadata text file describing the data
- Solution files, for reference or for use if workshop steps cannot be completed successfully

User (You):

- Basic knowledge of how to operate a computer and use the internet
- Familiarity with Microsoft Excel
- Very basic notions of database concepts such as the terms "table" and "field"
- Rudimentary understanding of hydrology and hydrologic data

About the Workshop Data

In this workshop, you will publish time series of water quality data measured in several Texas bays from 1987 to 1995. Measurements are taken at various depths below the water surface and sporadically in time. The data include measurements of salinity, dissolved oxygen, conductivity, temperature, and pH.

Each workshop computer has been installed with "raw" data files of the water quality observations described above. Some modifications have already been made to the raw files to facilitate data loading. We wanted you to get a sense of how to transform data, without overburdening you with easy-but-tedious data operations. The raw data files include site locations, time series of water quality observations, and metadata. These files are located in **C:\HIS Training\RawData**. During the workshop, you will transform the raw data files so that they are in a form that can be loaded into an Observations Data Model database. As a contingency plan in case you are unable to complete the data transformation process, transformed files have been generated for you, and are located in **C:\HIS Training\SolutionFiles\TransformedData**.

Review of HIS Data Publication Tools

The CUAHSI Hydrologic Information System provides web services, tools, standards and procedures that enhance access to more and better data for hydrologic analysis. HIS software is free and available on the HIS website at http://his.cuahsi.org. A variety of HIS software applications have been built to serve several types of users and scenarios, from data users to data publishers to educators and developers.

HIS Server

In this workshop, you'll be playing the role of the data publisher. This means you have some hydrologic observations data that you've collected, and you'd like to publish that data on the web in a standard way so that others can easily access and use it. To facilitate data publication, HIS offers **HIS Server** (http://his.cuahsi.org/hisserver.html), which is really just a bundle of HIS software designed for data publication that operates on a Windows Server or Windows XP computer.

HIS Server comes in two flavors: HIS Server Standard and HIS Server Lite. **HIS Server Standard** is integrated with ArcGIS Server and SQL Server Standard, which means you must have licenses for those commercial software in order to run HIS Server Standard. While HIS Server Standard offers more advanced analysis and publication capabilities, **HIS Server Lite** does not require commercial licenses beyond the operating system and is generally easier to implement. For this workshop, you'll use a computer that is basically HIS Server Lite, with the addition of Microsoft Office 2007 purely for the convenience of working with the workshop data. HIS Lite uses the Observations Data Model for data storage, and a WaterOneFlow web service for data publication.

Observations Data Model

The **Observations Data Model (ODM)** (http://his.cuahsi.org/odmdatabases.html) is a data model for the storage and retrieval of hydrologic observations in a relational database. ODM stores data and sufficient ancillary information (metadata) about the data values to provide traceable heritage from raw measurements to usable information allowing them to be unambiguously interpreted and used. A relational database format is used to provide querying capability to allow data retrieval supporting diverse analyses. To learn all the details of ODM, read the design specifications document on the website linked above.

Data can be loaded into an ODM database using a number of tools, including free HIS software. For loading static data files for what is generally a one-time process, the free **ODM Data Loader** is used. These data files are usually the result of a study or project that has been completed and will not need periodic updating. For data that are continuously updated, such as data streaming in from sensors in the field, use the free **ODM Streaming Data Loader**. For more complex data loading tasks, SQL Server Integration Services is one of many software packages up to the task. However, these software packages are typically not free and require significant training to learn how to use them. For this workshop, you'll be using the ODM Data Loader.

The ODM Data Loader reads input files that are formatted much like the tables in ODM. For example, if you want to load site locations into ODM, you could prepare a spreadsheet called "sites.xls" with column headings that use (roughly) the same names as fields from the Sites table in ODM (names are not case sensitive). In some cases, you can load data for more than one ODM table from a single input file by simply appending additional columns to the data in the file. This prevents you from having to create an input file for every table in ODM, which would be quite tedious since relational databases tend to have many associations across many tables. A document describing the input format required by the ODM Data Loader can be found at http://his.cuahsi.org/odmdataloader.html.

Once data are loaded into an ODM database, you can examine the data using the free **ODM Tools**. Or, if you have knowledge of SQL, you can write your own queries within SQL Management Studio. If the data look good, you can publish the data with a WaterOneFlow web service.

WaterOneFlow

A challenge in querying and interpreting data from disparate data sources is that each data source not only has its own method for asking for data, but also its own format for delivering the requested data to the user. **WaterOneFlow** overcomes this by providing a single query interface and a standard output format called WaterML, which is an XML language for the communication of water data. WaterOneFlow is a web service, which facilitates automated and programmatic access to the data. This is an advancement beyond simply publishing data on a web page, which can require complicated and often error-prone screen scraping and parsing.

A WaterOneFlow web service is available that hooks directly into an ODM database to publish data from that database. However, WaterOneFlow web services can also be written to support internal data formats other than ODM. This means that no matter what data storage mechanism you choose to use, you can still publish your data in a standard way with WaterOneFlow.

Registration with Texas HIS

Once your data are published, there's still the issue of data discovery. How do people learn about your data? That's where Texas HIS comes in. By registering your web service with http://waterdatafortexas.org, you'll be participating in a unified system for access to Texas data. At the time of this workshop, this online system is still under development, so you won't be registering your web service with during the workshop. However, an overview of the system will be demonstrated to the audience by the workshop facilitators.

HydroExcel

Now it's time to briefly play the role of the data user. Once data are published with HIS, how do people access the data? To help data users get started with HIS, several free applications or application extensions are available on the HIS website, geared towards application environments most commonly used by hydrologists such as Microsoft Excel. **HydroExcel** is an Excel spreadsheet customized with macros for accessing data from a WaterOneFlow web

service. In the workshop, you'll use HydroExcel to verify that you have successfully published your data with WaterOneFlow.

WORKSHOP OUTLINE

The workshop begins with presentations and demonstrations by the Texas HIS team to familiarize the audience with Texas HIS. Contact the workshop administrator to check for availability of these materials. The hands-on training portion of the workshop leads the audience through the data publication process with these key steps:

1. Translate raw data for loading into ODM.
2. Load data into an ODM database.
3. Expose database content online via WaterOneFlow web service.
4. Test the web service with HydroExcel.

Translating and Loading Data into ODM

To load data into ODM, you'll be using a tool called the ODM Data Loader. The ODM Data Loader loads data from comma delimited files (.csv) or Microsoft Excel 2003 files (.xls) that have a one row header that uses ODM field names in the header, followed by the data in subsequent rows. When loading data from Excel, the data should be located in a worksheet that has the same name as the file. More about these data formats can be found in the documentation for the data loader available at http://his.cuahsi.org/odmdataloader.html. Generally, the fields in the input files conform to the table structure of ODM, with some flexibility for specifying alternative information for database generated IDs. You'll see how this works during the workshop.

Most likely, your data are not exactly in the same format as what the ODM Data Loader is expecting. For example, instead of using the ODM terminology "SiteCode", you may call the unique ID for each of your observation sites a "USGS_ID". Also, you may need to do some leg work to look up information such as the horizontal datum associated with the latitude and longitude coordinates of your site, which is one of the pieces of information that ODM requires. Translating your data and metadata into ODM terminology may seem a bit tedious, but this exercise is actually very valuable as it helps you to fully understand ODM as well as your own data, and in the end you will have a database that richly describes your data.

In the interest of time, much of the data translation work has been performed for you. However, some items still remain untranslated and are just dying to have a talented hydro hero like you perform the transformation and save the day!

Installing a Blank ODM Database

An understanding of your own data as well as the Observations Data Model is essential before attempting to transform your raw data into inputs for the ODM Data Loader. To help you in this process, ODM includes some predefined terms called **controlled vocabularies** that you can choose from when populating its tables. Let's grab an ODM database from the HIS website and see some of these terms for ourselves.

1. In a web browser, navigate to http://his.cuahsi.org.
2. Follow the links to download a blank ODM database.
    a. Under Data Publication on the right, click **ODM Databases**.
    b. Click the link to download the **ODM 1.1 Blank SQL Server Schema Database**.
3. Unzip the contents of the downloaded file into the **HIS Training\Database** folder. The download includes both a database file (.mdf) and a log file (.ldf) that tracks transactions made to the database.

> **Note**
> You might notice the designation "1.1" with some of the HIS downloads. This refers to the 1.1 version of the software or sample data, which has slight improvements over the original 1.0 version that was initially released.

You'll now attach this database to SQL Server, and eventually you will load the workshop data into this database.

Tip
SQL Server is the database software installed on the workshop computers. "Attaching a database" to SQL Server basically means letting SQL Server know about your database so that its software can work with it.

4. Start **SQL Server Management Studio**. (Start | All Programs | Microsoft SQL Server 2005 | SQL Server Management Studio Express)
5. Click **Connect** to log into SQL Server.

Tip
SQL Server Management Studio is an application that lets you manage, view, and execute queries on your databases.

6. In the Object Explorer on the top left, right click **Databases** and click **Attach** (Figure 1).

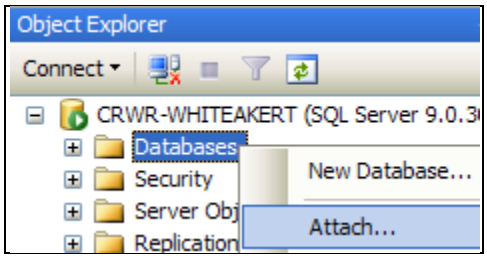

Figure 1 SQL Server Management Studio is used to attach databases to SQL Server

7. Click **Add**.
8. Navigate to **HIS Training\Database**, select **OD.mdf** (Figure 2), and click OK.

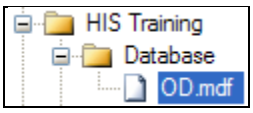

Figure 2 Locating the blank ODM database to attach

9. Change the "Attach As" name to **BayWQ** (Figure 3), and click **OK**.

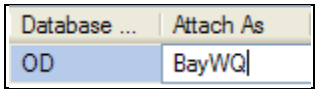

Figure 3 You can assign a database name of your choice in SQL Server

10. In the Object Explorer, click the plus sign to expand **Databases**. You should now see your BayWQ database.
11. Expand **BayWQ**, and then **Tables**, to see a list of tables in this blank ODM database.
12. Right click the **DataValues** table and click **Open Table** (Figure 4). This table stores time series values. Each row stores a single datetime, a single value, and metadata about that value. The table is currently blank, but you'll fix that later on!

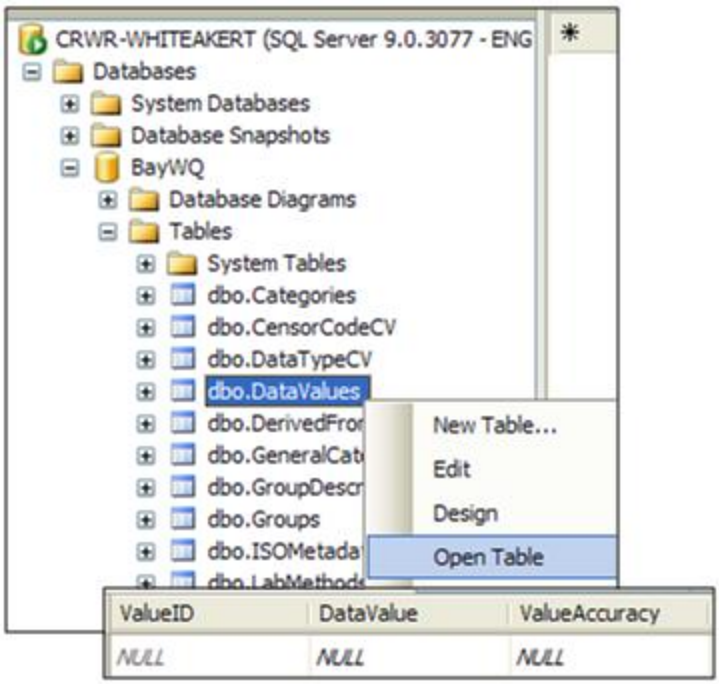

Figure 4 The blank ODM database has all of the ODM tables already created, with most of them empty and ready for data

13. Notice that there is a field in this table called **SiteID**. Rather than repeat the latitude, longitude, and other site information with every row in the DataValues table related to a given site, ODM keeps the database compact by using the SiteID to locate a matching row in the Sites table where the site details are stored. These types of **relationships** between tables are used extensively throughout ODM, leveraging the power of a relational database.
14. Click the X near the top right of the DataValues table to close it (Figure 5). Be sure not to click the X in the blue title bar for the application, or you will close Management Studio!

Figure 5 Closing the table will guarantee that it is refreshed in the view after loading data

15. Open the **Sites** table to see the information describing each site. **Close** the table when you are finished looking at it.

So far, the tables that you've been looking at are empty. Next, you'll open some tables that have already been populated with values that will be useful as you load data.

16. Open the **VariableNameCV** table. The letters "CV" at the end of the table name indicate that this is a controlled vocabulary table. Only terms from this table are used to describe variable names. This helps to standardize the terminology used to describe data across multiple ODM databases (Figure 6).

Term	Definition
Nitrogen, particulate organic as N	Particulate Organic Nitrogen as N
Nitrogen, total as N	Total Nitrogen (NO3+NO2+NH4+NH3+Organic)
Nitrogen, total kjeldahl	Total Kjeldahl Nitrogen (Ammonia+Organic) as N
Oxygen, dissolved	Dissolved oxygen
Oxygen, dissolved percent of saturation	Dissolved oxygen, percent saturation
Oxygen, dissolved, transducer signal	Dissolved oxygen, raw data from sensor
Peridinin	The phytoplankton pigment Peridinin
pH, filtered	pH, filtered
pH, unfiltered	pH, unfiltered
Pheophytin	Pheophytin (Chlorophyll which has lost the central Mg ion) is a degradation product of Chlorophyll
Phosphorus, inorganic as P	Inorganic Phosphorus as P
Phosphorus, organic as P	Organic Phosphorus as P

Figure 6 The VariableNameCV table is a controlled vocabulary of terms to use when naming variables

Tip
A level of data integrity is enforced through the use of controlled vocabularies (CVs) within ODM. If a field uses a CV, then only terms from that CV can be entered into that field. This way, the data use consistent terminology. If there is a CV term that you need which is not already in ODM, you can add it. Just remember to add the term to the CV table first, and then load your data.

17. Open the **SampleMediumCV** table. This table has terms used to describe the sample medium in which a measurement applies.
18. Continue opening and browsing the tables of ODM. When you are ready to move on, **close the tables** and **minimize SQL Server Management Studio**. You can leave Management Studio open as we'll use it again later.

Now that you've had a hands-on introduction to ODM, let's get to know the raw data that you'll be loading into it.

GETTING TO KNOW YOUR DATA

Let's now take a look at the raw data for the workshop, available on your computer at **C:\HIS Training\RawData**. These are data files from the fictitious organization called Texas Center for Water Resources. What? Fictitious? Never fear. The data that you'll be working with is derived from real measurements, but we're hiding the true source to emphasize that the data are purely for workshop use.

You should see three files in the RawData folder:

- **sampling_sites.csv** – Comma-separated text file with site locations for all sampling sites
- **observations.csv** – Comma-separated text file with the time series of water quality for all sites and variables
- **metadata.txt** – Text file with metadata describing your data

Sites

Open **sampling_sites.csv**. Each row in this file represents a single sampling site, and includes the following information:

- **Site ID** – Unique internal identifier for a site
- **Name** – Name of the site
- **LATITUDE** – Latitude of the site in decimal degrees
- **LONGITUDE** – Longitude of the site in decimal degrees

Time Series

Open **observations.csv**. Each row in this file represents a single water quality measurement at a particular site at a particular point in time. The spreadsheet includes the following information:

- **Site ID** – Unique internal identifier for a site
- **Variable** – Name of the item being measured
- **Datetime** – Date and time that a measurement occurred
- **Value** – Value of the variable at the given date and time
- **Offset** – feet below the water surface

Metadata

Open **metadata.txt**. If this data provider wanted to share the data, the information in the sites and time series files alone may not be enough to fully describe the data. Therefore, metadata such as the content of metadata.txt are often provided alongside the actual data files.

In this case, the metadata are a bit sparse, but you can find information about the nature of the study, the variables involved, and the data source. You'll use this information to help load the data into HIS.

Transformed Output

From these files describing sites, time series, and metadata, you will create transformed files that are formatted for loading into an ODM database using the ODM Data Loader. The transformed files that you will create are:

- Sites
- Variables
- DataValues

You'll also be manually entering information about the data source, just to give you a feel for how to make manual edits to a database. These are basically the minimum pieces of information you need to describe your data in ODM. Of course, there are plenty of additional types of information you can load into an ODM database to more fully describe your data, but for the purposes of this workshop, you'll just load the above items.

Before performing the transformation, it's imperative that you familiarize yourself with the structure of the Observations Data Model (http://his.cuahsi.org/odmdatabases.html), and the requirements for input files to the ODM Data Loader (http://his.cuahsi.org/odmdataloader.html). Review the information in "Review of HIS Data Publication Tools," the workshop presentation materials, and the online materials linked above for more information. The following sections describe the transformation procedure.

CREATING A DATA SOURCE RECORD

If you've made it this far, then you've downloaded and attached a "blank" ODM database to SQL Server. This database actually has a few tables already populated with common items like units of measure and spatial reference systems, and then it's up to you to fill in the rest with your data.

The first item you'll input is information about the data source. ODM has a Sources table where this information will go, and since there is only one data source for your workshop data, it's a good chance for you to add a record in the Sources table manually so you can see how that process works.

1. Restore **SQL Server Management Studio**. If you already closed it instead of minimizing it earlier, then start it up, connect, and then expand the databases list until you see the tables in the BayWQ database.
2. In the **BayWQ** database, **Open** the **Sources** table.
3. **Scroll** through the fields in the table. You see fields such as Organization, ContactName, and Address. Further to the right you'll see Citation and MetadataID. Citation stores a journal-style reference entry, so that when people use your data, they can cite it properly. MetadataID relates to a record in the ISOMetadata table.
4. Open the **ISOMetadata** table. Here you can provide more detailed metadata about your data source. Notice that there's already an entry with a MetadataID of 0 for "Unknown," for when you don't have additional metadata to enter.
5. **Close** the ISOMetadata table.
6. In the **Sources** table, input information about the data source (Figure 7), most of which you can find in the **metadata.txt** file. The rest, such as Citation, you can create on your own. You can input information by typing directly in the table. Just click where you see "NULL" and start typing away! A summary of the information to input is provided in Table 1.

Table 1 Information to Add to the ODM Sources Table

Field	Value
Organization	Texas Center for Water Resources
SourceDescription	Texas Coastal Water Quality Information
SourceLink	http://data.tcwr.gov
ContactName	Eric Schmersh
Phone	512-555-0233
Email	EricSchmersh@tcwr.gov
Address	123 Road Avenue
City	Austin
State	TX
ZipCode	78712
Citation	Texas Center for Water Resources. (1987-1995). "Texas Coastal Water Quality Information."
MetadataID	0

State	ZipCode	Citation	MetadataID
TX	78712	Texas Center fo...	NULL

Figure 7 Details about Sources or other information can be input into an ODM database by directly typing the information in the appropriate table

7. After you have finished typing the information, press **ENTER** on your keyboard to confirm the edit. Those little red exclamation marks will go away, indicating that your new record was successfully entered (Figure 8).

SourceID	Organization	SourceDescription
1	Texas Center fo...	Texas Coastal ...
NULL	NULL	NULL

Figure 8 Many fields in ODM, such as SourceID, are implemented as autonumber fields, which means the database fills them in for you. In this figure, the database assigned our new record for our data source a SourceID of 1.

8. Make a note of the **SourceID** that was automatically assigned for the record you just entered. You'll use that number later on when you load your time series values.
9. **Close** the Sources table.

Ahhh, good old manual database entry. That wasn't so bad now, was it? Yikes! But there are still 49 sites and thousands of time series values to load. That's a lot of manual entry! Time to enlist the services of the ODM Data Loader.

Next, you'll create files that the ODM Data Loader will use to load data into the database. You'll start with a file for sites, then a file describing the variables, and finally a file for the actual time series values. Basically, you first load everything to fully describe your time series values, and then load the values last. The reason for this is the ODM Data Loader checks each and every time series value to make sure it is properly described by related tables in your database.

Creating a Sites File

Information about your sampling sites is contained in the **RawData\sampling_sites.csv** file. You'll transform this to a **TransformedData\sites.csv** file. The fields in the transformed file include:

- SiteCode
- SiteName
- Latitude
- Longitude
- LatLongDatumSRSName
- SiteState

These are some of the field names the ODM Data Loader expects to see when loading information about sites. The first four fields match very well with the data from sampling_sites.csv. Note that you will be adding a couple of fields that aren't in sampling_sites.csv: LatLongDatumSRSName and SiteState. The ODM requires the datum to be stored with the latitude and longitude coordinates of a site. Luckily, you can find that data in the meatadata.txt file. The metadata indicates that the datum used is NAD83. There is already a record for this datum in the SpatialReferences table of ODM (Figure 9). The record has a SpatialReferenceID of 2, and an SRSName value of NAD83.

SpatialReferenceID	SRSID	SRSName	IsGeographic	Notes
0	0	Unknown	False	The spatial refer…
1	4267	NAD27	True	*NULL*
2	4269	NAD83	True	*NULL*
3	4326	WGS84	True	*NULL*

Figure 9 The NAD83 datum is among the list of coordinate systems in the ODM SpatialReferences table

Being a relational database, the ODM Sites table is expecting a numerical datum ID (in this case, the number 2) to accompany each site record. However, it's easier for us humans to interpret text rather than numbers during the data translation process, which is why the ODM Data Loader allows you to use "NAD83" instead of the number 2 to refer to your datum. The ODM Data Loader will make sure the LatLongDatumSRSName refers to an SRSName in ODM's SpatialReferences table before finalizing the data loading operation. This is one of the advantages of using the ODM Data Loader for loading data – it performs some quality control and maintains integrity of relationships between the tables of ODM during loading.

You'll be following this procedure of referencing information in your ODM database a lot as you figure out how to translate your data to ODM. A quick summary of the procedure is:

1. You need to know how to describe some aspect of your data, such as the datum.
2. You check the ODM database to see if a table already exists to describe that item.
3. You find a match for your item in the ODM table, and use the matching term from ODM as you build your translated data files for eventual loading into the ODM database.

If you don't find a matching item in ODM, you can add it. There are even methods for adding items to the master list (i.e., controlled vocabularies) maintained by CUAHSI HIS, although updating those lists is beyond the scope of this workshop.

Oh yeah, I almost forgot about that second field that you're adding: SiteState. All of these data are for Texas bays, so you'll assign "Texas" as the state for all sites.

Details of how to map from the raw file to the transformed file are below.

Table 2 Mapping Raw Data to Sites

Transformed Field	**Raw Data Field (from sampling_sites.csv)**	**Notes**
SiteCode	Site ID	

SiteName	Name	
Latitude	Latitude	
Longitude	Longitude	
LatLongDatumSRSName	n/a	Use “NAD83” from ODM SpatialReferences table
SiteState	n/a	Use “Texas”

Tip
The ODM Data Loader ignores case in the field names, so Longitude and LONGITUDE are both valid field names.

Tip
If you have any trouble creating the transformed files, you might find it helpful to refer to the solution files in **HIS Training\SolutionFiles\TransformedData**.

To create the transformed sites file:

1. Open **HIS Training\RawData\sampling_sites.csv** with Excel. You can probably do this by double clicking on the filename in Windows Explorer.
2. Save the file in the **HIS Training\TransformedData** folder as **sites.csv**. Be sure to select **CSV (Comma delimited) (*.csv)** from the **Save as type** drop down box as you save the file.

ODM Data Loader Best Practice – Use CSV Files
The ODM Data Loader can work with both comma delimited (.csv) files and Microsoft Excel 2003 (.xls) files. However, the author has found that sometimes Excel cell formatting can cause an incorrect interpretation of the data. Therefore, the author recommends saving the transformed files as comma delimited text files, which contain no instructions about how data should be formatted.

3. After saving, click **Yes** if prompted to keep the workbook in CSV format.

Transforming the sites file will be very easy. You’ll start by renaming some fields, and then add a couple of new fields. Note that you must not misspell any of the field names, or else the ODM Data Loader will not recognize the field.

4. Rename the following fields:
    a. Site ID to SiteCode
    b. Name to SiteName
5. Add a field called “LatLongDatumSRSName” (without quotes) and calculate all values to be “NAD83” (without quotes).
6. Add a field called “SiteState” (without quotes) and calculate all values to be “Texas” (without quotes).
7. **Save** the file. If prompted about keeping the workbook in CSV format, click **Yes**.
8. **Close** the file. If prompted about saving changes to the file, click **No**. You just saved them, so you should be fine.

Great job! Creating the sites file was a snap since the raw data already had a sites file to begin with. In addition to sites, ODM also keeps a table of variable descriptions. Next you'll create a variables file that defines the variables associated with your time series values.

CREATING A VARIABLES FILE

Your data represents several water quality time series variables. Some information about these variables can be found in the metadata.txt file. Often you'll have to do a bit of legwork to fill in the rest, in order to fully describe your data in ODM.

As a brief summary, measurements for salinity, dissolved oxygen, conductivity, temperature, and pH were taken sporadically in time and at different depths below the water surface.

The fields in the variables file that you will create include:

- VariableCode
- VariableName
- Speciation
- VariableUnitsName
- SampleMedium
- ValueType
- IsRegular
- TimeSupport
- TimeUnitsName
- DataType
- GeneralCategory
- NoDataValue

VariableName, Speciation, SampleMedium, ValueType, DataType, and GeneralCategory must all conform to terms in ODM controlled vocabularies. This can actually make your life easier because all you have to do is pick the CV term that best describes your variable.

The ODM Data Loader will use the abbreviation for the variable units name to match the variable to a unit in the Units table. This is another example of how the ODM Data Loader performs integrity checks on the data during the loading process.

Details of how to supply values in the transformed file are shown in the table below. The example values are for the dissolved oxygen variable.

Table 3 Mapping Raw Data to Variables (for Dissolved Oxygen)

Transformed Field	Value	Notes
VariableCode	DO	The term used for dissolved oxygen, from metadata.txt
VariableName	Oxygen, Dissolved	From VariableNameCV
Speciation	Not Applicable	From SpeciationCV – no speciation applies to dissolved oxygen
VariableUnitsName	milligrams per liter	From UnitsName field of Units table, matches units from metadata.txt
SampleMedium	Surface Water	From SampleMediumCV
ValueType	Field Observation	From ValueTypeCV
IsRegular	FALSE	These are instantaneous measurements made irregularly through time
TimeSupport	0	Use "0" when instantaneous measurements are recorded
TimeUnitsName	Day	The value doesn't really matter for instantaneous data, as long as it matches text in the UnitsName field of the Units table
DataType	Sporadic	From DataTypeCV
GeneralCategory	Water Quality	From the GeneralCategoryCV
NoDataValue	-9.99	From metadata.txt

To create the variables file:

1. In the interest of time, a file named **variables.csv** has already been created for you in the **HIS Training\TransformedData** folder. This file has records for four out of five variables. You'll add the fifth variable, dissolved oxygen. Open the file with Excel.
2. In row 6, fill in the values for the dissolved oxygen variable. For reference, see Table 3.
3. **Save** and **close** the file.

Nice job! Two out of three input files for the ODM Data Loader have now been created. Before you work on the last file, let's go ahead and load the other data into ODM.

LOADING SITES AND VARIABLES INTO ODM

Before a single time series value is loaded into an ODM database, you must have already loaded information that describes the time series, e.g., the sites, sources, and variables data that you just finished preparing. As you load time series values, the ODM Data Loader will look for the other information related to the time series to make sure it is being appropriately described. If something is missing, the ODM Data Loader will pop up a friendly message basically saying something to the effect of, "Hi there. I see you're loading data, but you haven't told me what the data represent or how to describe it in the database." This is one way the data loader helps to ensure the integrity of your data.

So without further ado, let's run the ODM Data Loader on those transformed files that you've been working on.

1. Open the **ODM Data Loader**. (Start | All Programs | CUAHSI HIS | ODM DL 1.1)

2. Enter the connection information for the blank database you attached earlier (Figure 10). As a reminder, you attached a database named BayWQ to your local SQL Server, which is the free "SQLExpress" edition for the workshop computers. You'll use the SQL administrator account, "sa", to connect.
    a. Server Address: (local)\SQLExpress
    b. Database Name: BayWQ
    c. Server User ID: sa
    d. Server Password: [password provided to workshop participants]

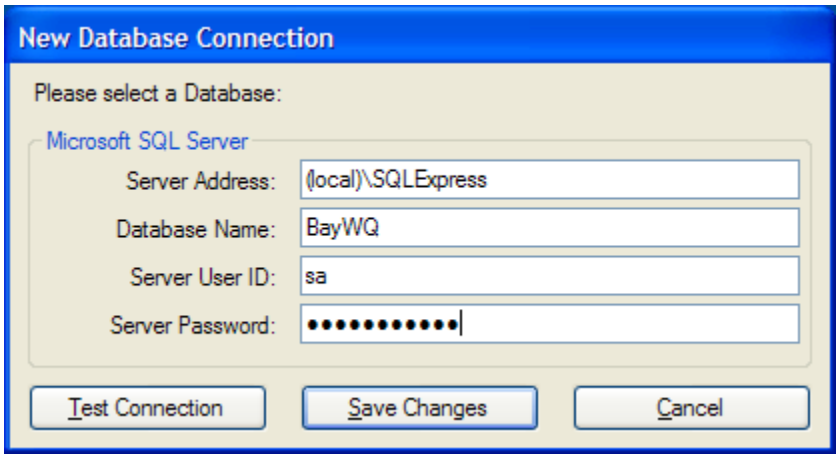

Figure 10 Connecting ODM Data Loader to the database

3. Click **Save Changes**.
4. Dismiss the message indicating that the connection was successful.

With the connection set, you are now ready to open the transformed data files and commit them to the database.

5. Click **Open**.
6. Navigate to and open the **HIS Training\TransformedData\sites.csv** file. The ODM Data Loader previews the file (Figure 11). As indicated in the bottom left corner, the application has recognized that you are loading sites information. It does this based on the field names in the input file you selected.

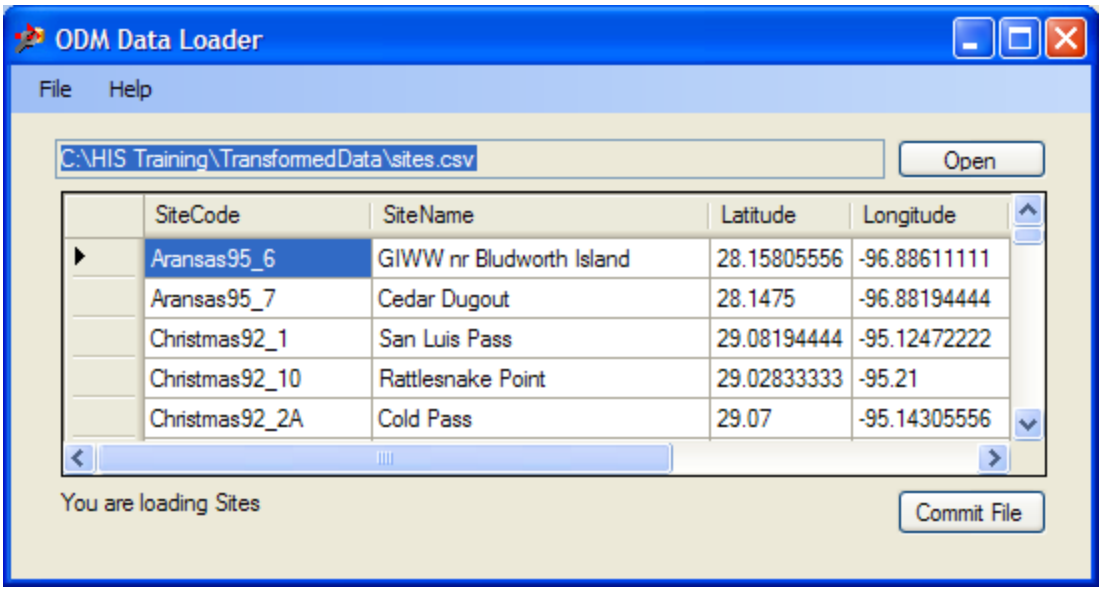

Figure 11 Loading sites into ODM

7. Click **Commit File** to write the records to the database. After a moment, a message box appears indicating success (Figure 12).

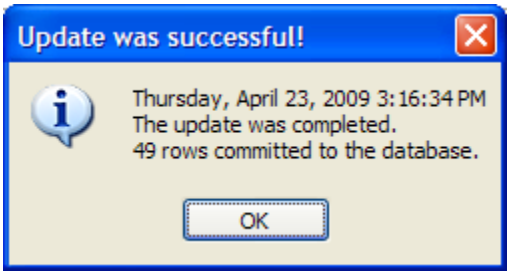

Figure 12 If data loading succeeded, the ODM Data Loader lets you know with a message box

8. Dismiss the message box.
9. Repeat steps 5-8, this time loading **variables.csv**.
10. **Minimize** the data loader when it has finished.

You'll use the data loader again in a moment to load the time series values. For now, take look inside the database to see that data were loaded successfully.

11. Restore **SQL Server Management Studio**. If you closed it, please open and connect to it again.
12. In the **BayWQ** database, open the **Sites** table.

You should now see your sites in the table (Figure 13). If the table is still blank, try closing and reopening the table to refresh it.

SiteID	SiteCode	SiteName	Latitude	Longitude	LatLongDatumID
1	Aransas95_6	GIWW nr Bludw...	28.15805556	-96.88611111	2
2	Aransas95_7	Cedar Dugout	28.1475	-96.88194444	2
3	Christmas92_1	San Luis Pass	29.08194444	-95.12472222	2

Figure 13 Sites information successfully loaded into ODM

Notice the LatLongDatumID field. The ODM Data Loader automatically used the abbreviation for the datum that you provided in the transformed sites file and matched it up with the datum ID from the SpatialReferences table in ODM.

13. Open the **Variables** table to see the result of data loading.

With this information in the database, you can now load the time series values.

Creating and Loading a Data Values File

The actual time series of groundwater levels is stored in raw form at **HIS Training\RawData\observations.csv**. Now that you've loaded metadata about the time series into your ODM database, you're ready to load the time series values themselves.

The fields in the transformed data values file that you will create include:

- SiteCode
- VariableCode
- LocalDateTime
- UTCOffset
- DataValue
- OffsetValue
- OffsetUnitsName
- OffsetDescription
- SourceID
- MethodID
- QualityControlLevelID
- CensorCode

Many of these fields do not have an equivalent in the observations.csv file. The original data source may not have conceived that their data would be fully described in an ODM database! For some of the additional fields, you will find matching terms or IDs in the ODM database. For example, the ODM Data Loader will match the SiteCode and VariableCode to items in the Sites and Variables tables which have already been loaded. You'll use the SourceID from the Sources table that matches the record you added earlier. Similarly, you can look in the Methods and QualityControlLevels tables to get an idea of what to enter for MethodID and QualityControlLevelID. For offsets, you'll fully describe them in the transformed file, which will force the ODM Data Loader to create new records for those items in the OffsetTypes table of ODM.

Details of how to supply values in the transformed file are shown below.

Table 4 Mapping Raw Data to Data Values

Transformed Field	Raw Data Field (from observations.csv)	Notes
SiteCode	Site ID	
VariableCode	Variable	
LocalDateTime	Datetime	
UTCOffset	"-6"	From metadata.txt, all values are in Central Standard Time, which is six hours behind Coordinated Universal Time (UTC), hence the value of -6 for the UTC offset
DataValue	Value	
OffsetValue	Offset	
OffsetUnitsName	"international foot"	From UnitsName field in Units table of ODM, matching information from metadata.txt
OffsetDescription	"Depth below water surface"	From metadata.txt
QualityControlLevelID	"-9999"	"-9999" indicates an unknown QC level, defined in the QualityControlLevels table of ODM
MethodID	"0"	Method information was not included in metadata.txt, so you'll use a MethodID of 0 to indicate that no method is specified, as defined in the Methods table of ODM
SourceID	e.g., "1"	This should be the value of the SourceID for the source data you entered earlier
CensorCode	"nc"	From CensorCodeCV table of ODM. "nc" is the default value meaning "not censored."

Like the raw sites file earlier, you'll start with the raw time series file and modify it accordingly.

To create the data values file:

1. Open **HIS Training\RawData\observations.csv** with Excel.
2. Save the file in the **HIS Training\TransformedData** folder as **datavalues.csv**. Be sure to save the file as a comma delimited file.
3. Rename the following fields:
    a. Site ID to SiteCode
    b. Variable to VariableCode
    c. Datetime to LocalDateTime
    d. Value to DataValue
    e. Offset to OffsetValue

4. Add a field called “OffsetUnitsName” (without quotes) and calculate all values to be “international foot” (without quotes). Likewise, for the remaining columns, assume that the column names and values should be input without quotes.
5. Add a field called “OffsetDescription” and calculate all values to be “Depth below water surface”.
6. Add a field called “UTCOffset” and calculate all values to be “-6”.
7. Add a field called “QualityControlLevelID” and calculate all values to be “-9999”.
8. Add a field called “MethodID” and calculate all values to be “0”.
9. Add a field called “SourceID” and calculate all values to be the SourceID that was generated when you created the source information in the database earlier.
10. Add a field called “CensorCode” and calculate all values to be “nc”.
11. **Save** and **close** the file.

Tip
Quality control levels provide some confidence as to the amount of quality control performed on a dataset. A quality control level of zero (0) indicates raw data, while a quality control level of one (1) indicates quality controlled data. The level of quality control for individual data values is not available for the workshop dataset, so you’ll use a value of -9999 to indicate “unknown”. For more information on quality control levels, see the ODM design specifications document at http://his.cuahsi.org/odmdatabases.html.

Now all that’s left is to load the data values into ODM.

12. Restore the **ODM Data Loader.** If you closed the program, please reopen it and connect to the BayWQ database.
13. Open the **datavalues.csv** file.
14. **Commit** the data.
15. When the data loader finishes, close the program and view the results in the DataValues table in SQL Server Management Studio.

That’s it! You’ve now finished loading all of the raw data into an ODM database. If you are familiar with SQL, you can now write queries to play with other tools in SQL Server to work with the data. But just in case you are not a SQL pro, HIS has developed software called the ODM Tools, which can be used to query and plot graphs of data in an ODM database.

Working with ODM Data

Now that data are loaded into an ODM database, how do we analyze it? Lucky for us, the ODM Tools are specifically designed to query and visualize data within an ODM database. You will use the ODM Tools to examine the contents of the database that you have just created.

The ODM Tools software is a free download from the HIS website at http://his.cuahsi.org/odmtools.html. The tools have already been installed on the workshop computers.

To examine your data with the ODM Tools:

1. Open the **ODM Tools**. (Start | All Programs | CUAHSI HIS | ODM Tools 1.1)
2. Enter the connection information for your database as you did with the ODM Data Loader, and click **Save Changes**.
3. Dismiss the message indicating that the connection was successful.

The ODM Tools application opens with three tabs visible: Query, Visualize, and Edit. The Query tab is selected by default. On this tab, you specify various filters to search for time series in the ODM database. Take a moment to review the query options.

Our data are fairly uniform in nature, i.e., they all have the same data source, value type, sample medium, etc. This limits the kinds of interesting queries we can perform with our data. But we can still query by site, variable, number of observations, or time period.

As each workshop computer's data are different, you are free to query by whatever parameters you choose. However, to make sure that each participant gets at least one data series returned, let's query for data at a site as described below.

4. Click the check box to **Query by Site** (Figure 14). This enables you to choose a site from the site list.
5. **Select Rattlesnake Point** in the site list by left clicking on it (Figure 14).
6. Click the check box to **Query by Variable** (Figure 14). This enables you to choose a variable from the variable list.
7. **Select Temperature** in the site list by left clicking on it (Figure 14).
8. Click the **Query** button in the bottom right corner (Figure 14).

The results of the query are shown at the bottom of the application window. You should see a single item there, which represents temperature for the site that you selected.

9. **Right click the data series** (Figure 14)**.** In the context menu, notice that you have options for plotting graphs, editing the data, viewing and exporting metadata, and exporting the data series itself as comma or tab delimited text.
10. In the context menu, click to **View MetaData** (Figure 14).

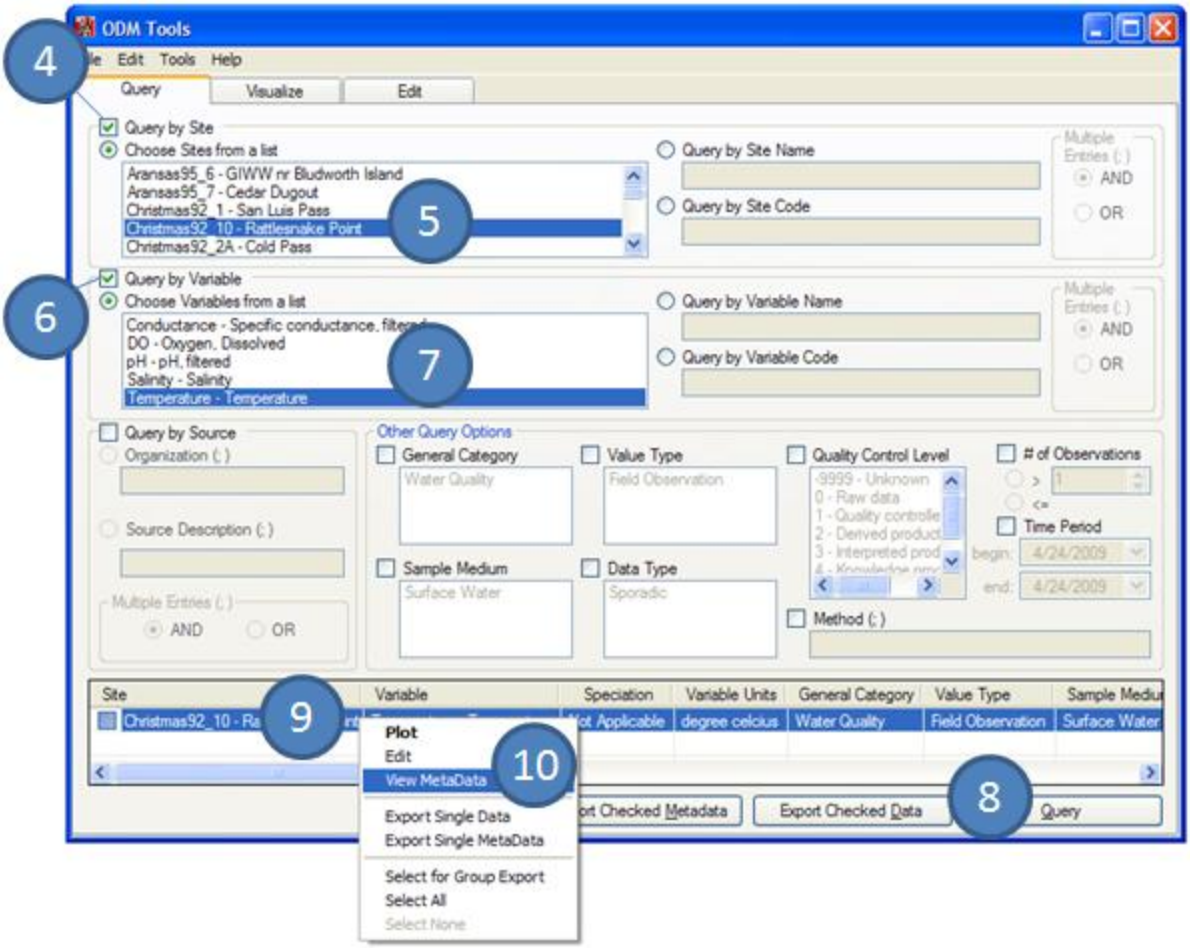

Figure 14 Querying data with the ODM Tools

Metadata for the time series are extracted from the database and transformed to XML (Figure 15).

```
<VariableInformation>
  <VariableCode>Temperature</VariableCode>
  <VariableName>Temperature</VariableName>
  <Speciation>Not Applicable</Speciation>
- <VariableUnits>
    <UnitsName>degree celcius</UnitsName>
    <UnitsType>Temperature</UnitsType>
    <UnitsAbbreviation>degC</UnitsAbbreviation>
  </VariableUnits>
  <SampleMedium>Surface Water</SampleMedium>
  <ValueType>Field Observation</ValueType>
  <IsRegular>False</IsRegular>
```

Figure 15 Metadata exported from ODM using the ODM Tools

11. Right click the data series again, and this time click to **Export Single Data**.
12. **Save** the file to disk at a location of your choosing.
13. When the data export is complete, dismiss the message box and open the data file. There are your data again, this time with internal identifiers and other pieces that the ODM Data Loader filled in for you when loading data.

Tip
You can select additional items to be included in the data export by clicking **Tools | Options**.

14. **Close** the metadata file and exported data file when you are finished looking at them.

Now let's plot a graph of the data.

15. Right click the data series, and click **Plot**.

You are brought to the Visualize tab, where the ODM Tools plot a graph of the data. Information about the data series you are plotting is shown at the bottom of the application window (Figure 16).

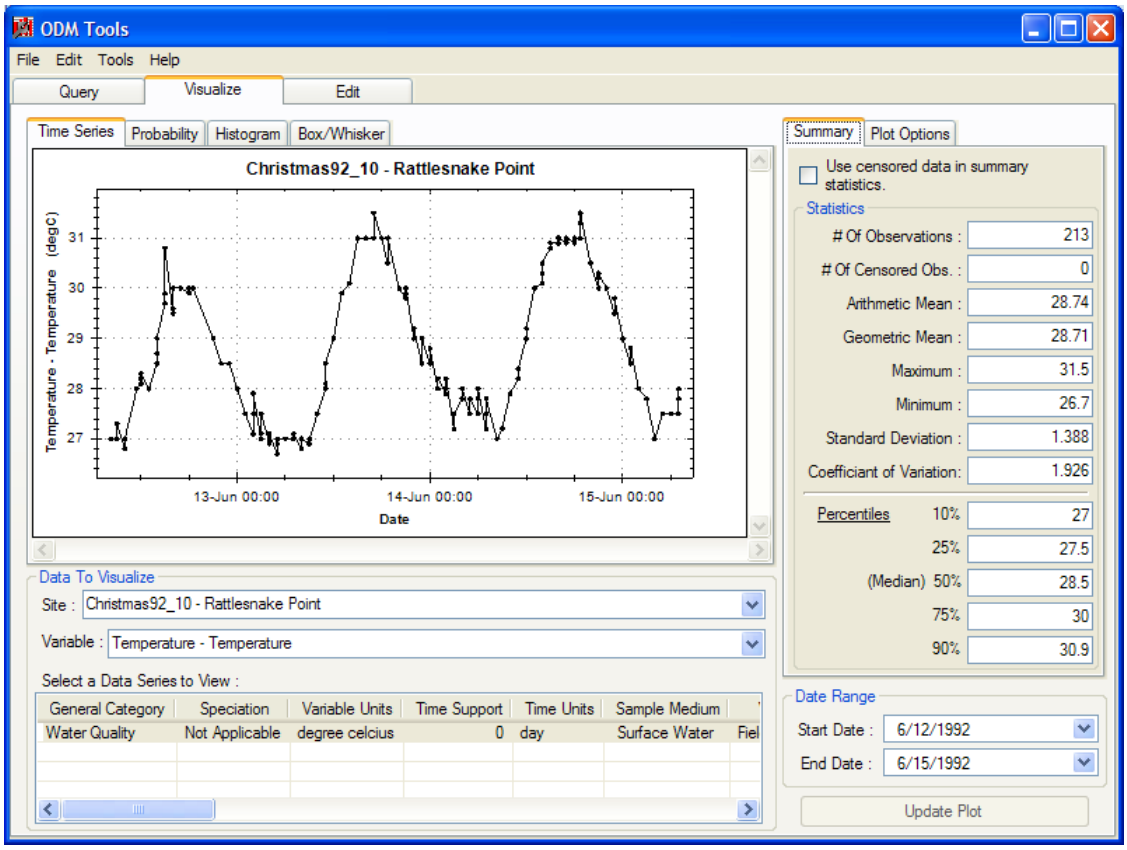

Figure 16 Visualize tab of the ODM Tools

16. Experiment with the charting capabilities. You can change plot options, view summary statistics, show a probability plot, show a histogram, and show a box/whisker plot. You can also interact with the plot by clicking on it. Left click and drag a box to zoom in, and right click to zoom out or copy and print the chart.

Finally, let's briefly take a look at the Edit tab. You won't be doing any editing for this workshop, but you can at least get a sense of what you can do with the ODM Tools.

17. Click the **Query tab** (Figure 17). Your previous results are still visible at the bottom.
18. **Right click** the Temperature time series at the bottom, and click **Edit** (Figure 17).

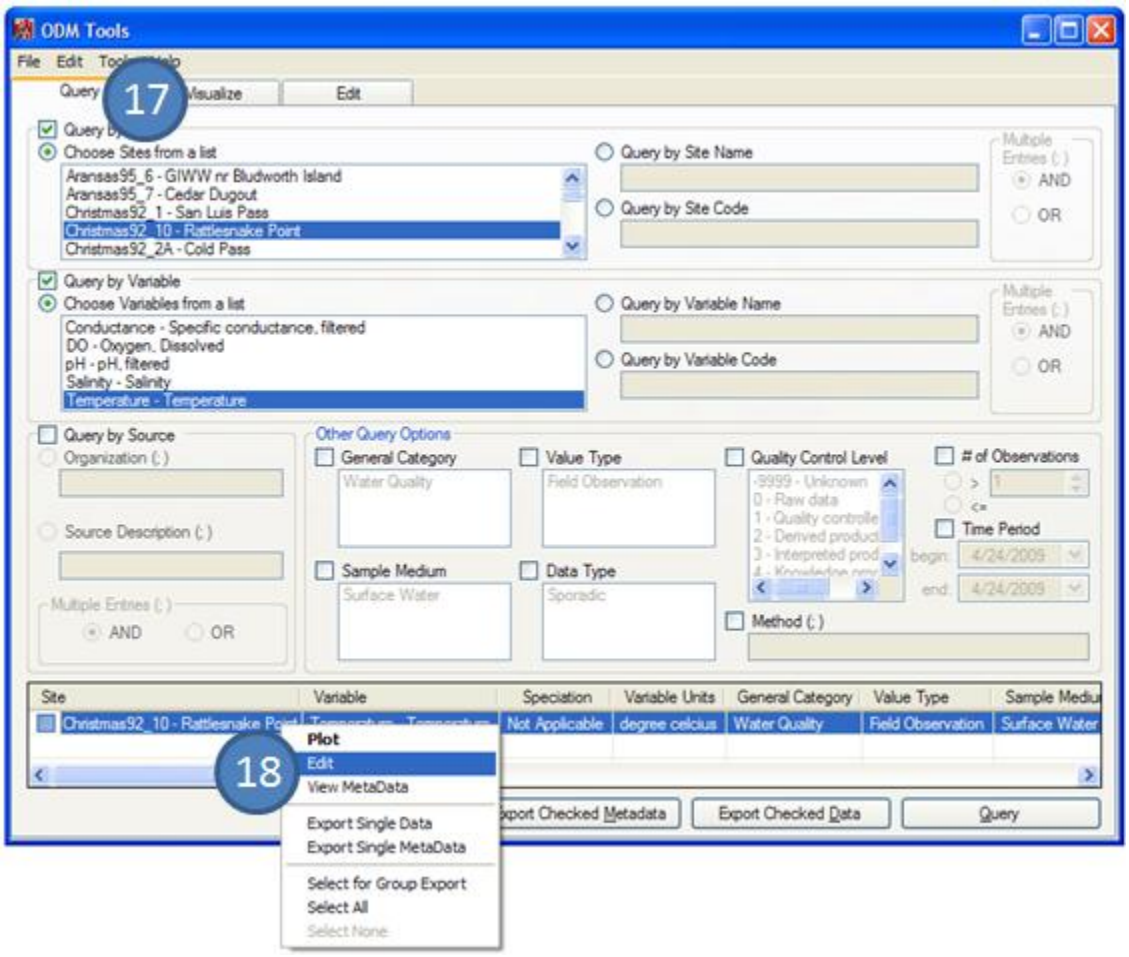

Figure 17 Selecting a series to edit in ODM Tools

This brings you to the Edit tab and pulls up the time series you selected. Look at the options on the right. You can change individual values or apply a data filter to perhaps look for outliers. Let's try that.

19. Click the option to set a **Value Change Threshold** (Figure 18). This option is useful for locating values that differ greatly from the other values around it, which could indicate a sensor malfunction, human error in observation, or some other anomaly.
20. Type a value for the change threshold that will capture some of your data, just to see what happens. For my data, I used a value of 1 (Figure 18). Press **ENTER** after you have typed in your value to confirm it. The Apply Filter button should now be enabled.
21. Click **Apply Filter**. Any values that match the filter will be highlighted in red in the plot (Figure 18). This shows us values that we may want to check for quality assurance.

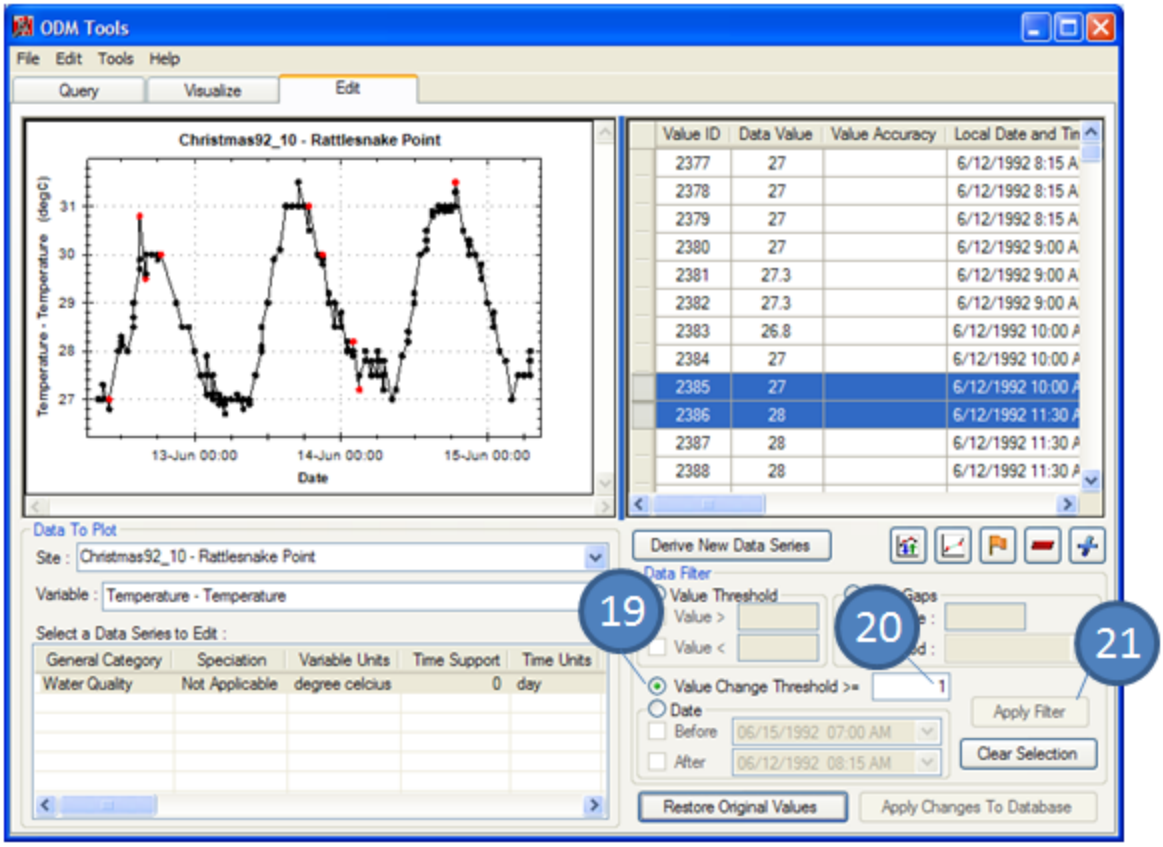

Figure 18 Applying a Data Filter with the ODM Tools

Lastly, let's take a look at options for deriving new data series.

22. Click **Derive New Data Series**.

In the window that opens, you'll see options for applying an algebraic equation, using a daily aggregate function, creating attributes to describe the output, and more (Figure 19). These functions are useful for turning your raw information into knowledge products, like when gage height data for rivers are converted to streamflow.

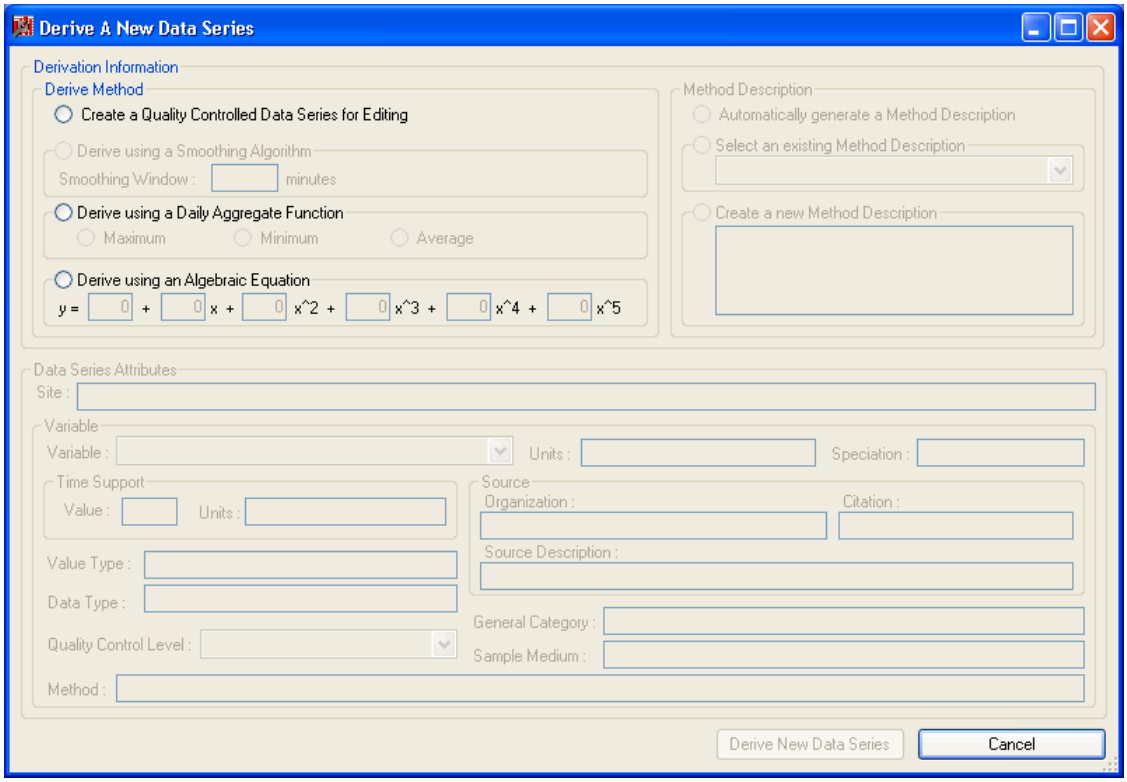

Figure 19 Window to derive a new data series in ODM Tools

In your case, because the quality control level is unknown, you'd first have to create a quality controlled data series for editing before the other options would be enabled. However, editing data series is beyond the scope of this introductory workshop, so we'll leave the ODM Tools for now.

23. Click **Cancel** to close the Derive A New Data Series window.
24. **Close** the ODM Tools.

Congratulations! You are now an expert in loading, querying, and visualizing data in an ODM database. For some scientists, this may be all that is desired: a means of storing and working with hydrologic observations data. However, there is often merit in sharing data with a larger community. That's where the WaterOneFlow web services and HIS Central come in.

Now that you have prepared an ODM database, you will publish the data with a WaterOneFlow web service to make it accessible online.

Publishing an ODM Database with WaterOneFlow

As far as your local setup goes, you are now ready for action with your ODM database. But what if you want to share the data with others online? That's where WaterOneFlow web services come in. WaterOneFlow defines a standard set of queries and a standard output format for accessing data, regardless of whether the data are accessed internally from an ODM database, some other database, or even through another website. Additionally, WaterOneFlow provides a layer of security over your database which makes it less susceptible to hackers than exposing the database itself with public access.

For those who have their own database format for storing data, they must write their own WaterOneFlow web service to publish their data in HIS. However, you're in luck! HIS includes a free WaterOneFlow web service specifically designed to work with an ODM database. This means you don't have to write a single line of programming code for your service. You just have to set it up on your computer and tell it to talk to your ODM database.

In this portion of the workshop, you'll download and install a WaterOneFlow web service to work with your ODM database. The main steps are:

1. Install the WaterOneFlow web service on your computer.
2. Tell your ODM database to allow the web service to talk to it.
3. Tell your web service which ODM database to talk to.
4. Check the result.

Note
Detailed instructions on how to install and configure an ODM WaterOneFlow web service are available at http://his.cuahsi.org/wofws.html. These instructions include advanced configuration options that are outside the scope of this workshop. If you have questions about advanced configuration or security settings for your web service, please refer to instructions linked above, ask the workshop instructor, or ask the HIS team by visiting our contact page at http://his.cuahsi.org/contactus.html.

Installing a WaterOneFlow Web Service

To install a web service, you'll download a copy from the HIS website and put it in a folder on your computer. Then you'll tell the folder that it can be accessed from the web and give it some permissions to allow the web service to run. Finally, you'll tell the web service to allow anonymous access, so that people don't need to know your login and password in order to connect.

To install a WaterOneFlow web service:

1. Create a folder for the web service.
    a. In Windows Explorer, navigate to the **HIS Training** folder.
    b. Right click in the HIS Training folder, and then click **New | Folder**.
    c. Type "**DataService**" (without quotes) as the name, and press ENTER.

Tip
You can name the folder whatever you want. We're just using the name DataService for the workshop.

2. Download the web service into the folder.
    a. In a web browser, navigate to http://his.cuahsi.org.
    b. Follow the links to download an ODM WaterOneFlow web service.
        i. Under Data Access in the middle of the page, click **WaterML Web Services**.
        ii. Click the link to **Download ODM WaterOneFlow Web Services**.
    c. Unzip the contents of the downloaded file into the **HIS Training\DataService** folder.
3. Set security permissions on the folder to allow the web service to run (Figure 20). These permissions are associated with three user names on your computer related to web applications. You'll tell the folder to allow those user names to work with the folder.
    a. In Windows Explorer, right click the **DataService** folder and click **Properties**.
    b. In the DataService Properties, click the **Security** tab, and then click **Add**.

Figure 20 Accessing the Security tab of the web service folder to add permissions

    c. In the **Select User or Groups** dialog, click **Advanced**.
    d. In the next dialog that opens, click **Find Now**. This lets you pick from the accounts on your machine.
    e. Select **ASPNET**, **Everyone**, and the **IWAM** account, and click **OK** (Figure 21).

Note
The IWAM account will have some other letters trailing off of it, so that it will appear as something like "IWAM_COELAB-A823KS6A". Just look for the first few letters, "IWAM". Also note that for Windows 2003 Server, you will add the IIS_WPG account instead of IWAM.

Tip
Hold down the CTRL key as you select the user names in order to select more than one at the same time.

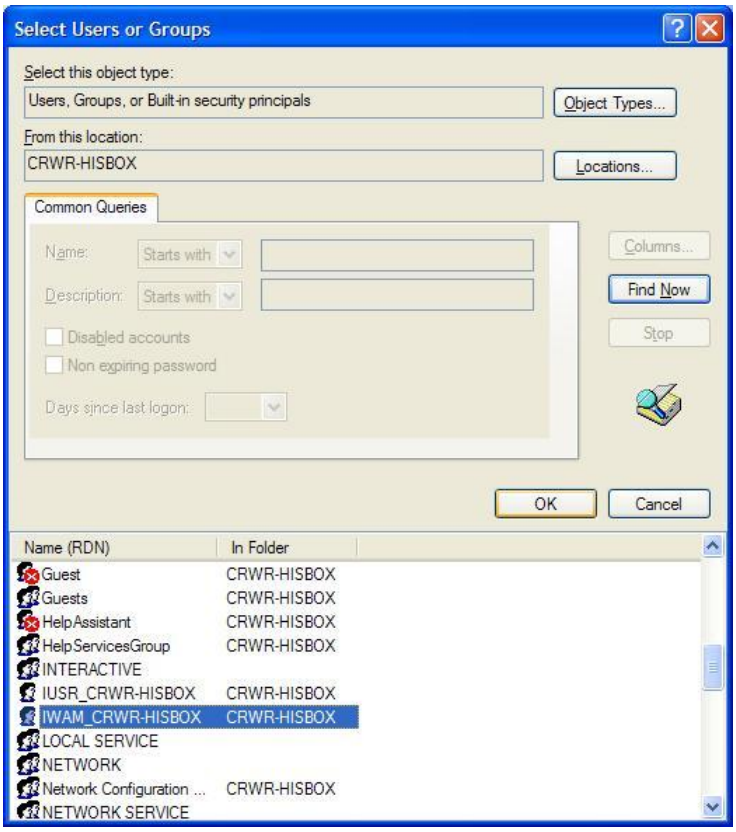

Figure 21 Selecting users for enabling the web service to run

f. Verify that the user names you selected appear in the Select Users or Groups dialog, and click **OK** (Figure 22).

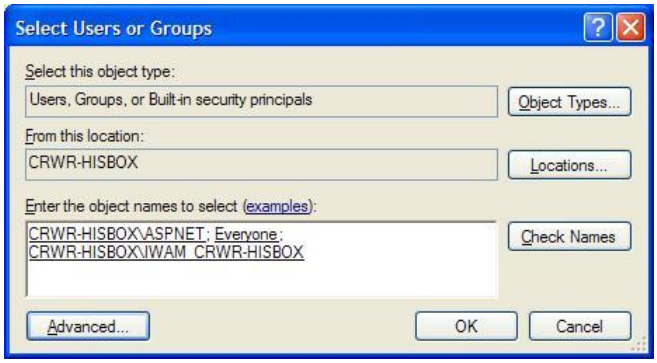

Figure 22 Confirming users to add to the web service folder

g. In the DataService Properties dialog, select user names and check the appropriate boxes to assign these permissions:
    i. Launch IIS Process Account – Modify (Figure 23)
    ii. ASPNET – Full Control

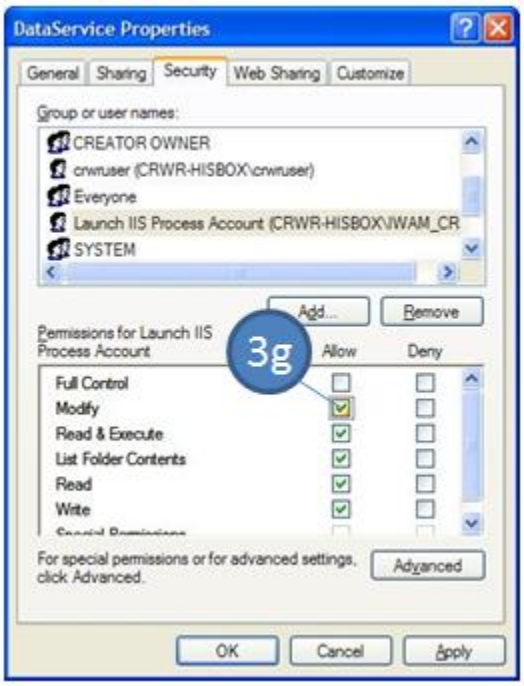

Figure 23 Assigning permissions to the web folder users

Now that security settings are in place, you'll tell the folder that it can be accessed from the web.

4. Share the folder on the web (Figure 24).
    a. In the DataService Properties dialog, click the **Web Sharing** tab.
    b. Click the option to **Share this folder**.
    c. In the Edit Alias dialog that opens, click to allow **Write access**. Click **Yes** when prompted, "Are you sure…."
    d. Click **OK** on the Edit Alias dialog.
    e. Click **OK** on the DataService Properties dialog.

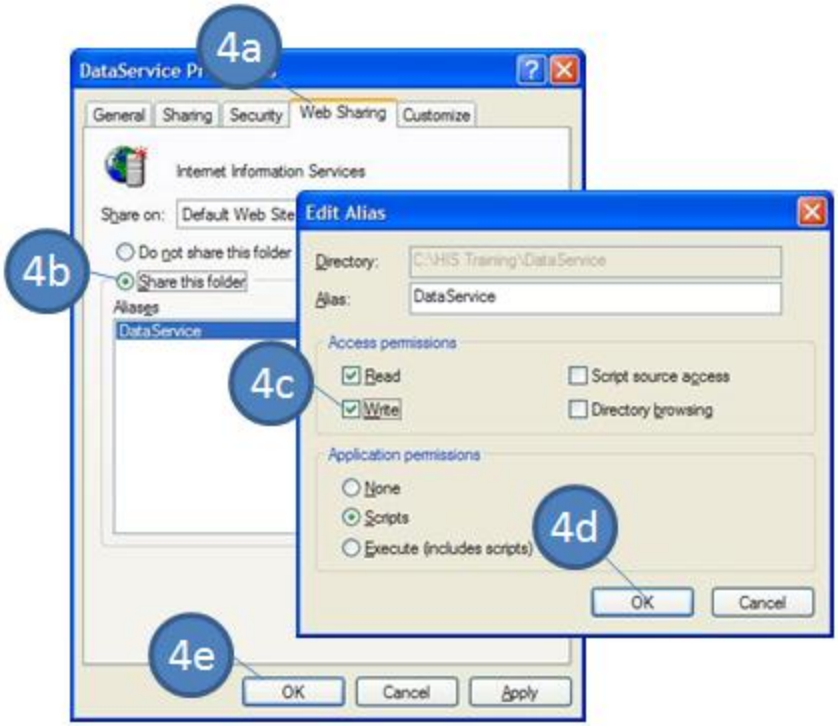

Figure 24 Sharing the web service folder on the web

Next you will tell the web application to allow anonymous access. Otherwise, users would have to know a valid username and password on your computer in order to connect. You'll assign this setting using Internet Information Services, which is one of the administrative tools on your computer.

5. Allow anonymous access to the web service.
    a. Open **Control Panel**. (Start | Settings | Control Panel)
    b. Double click **Administrative Tools**.
    c. Double click **Internet Information Services**.
    d. In the Internet Information Services window, click the plus sign to expand the items until you see **DataService** under **Web Sites | Default Web Site** (Figure 25).
    e. Right click **DataService** and click **Properties** (Figure 25).

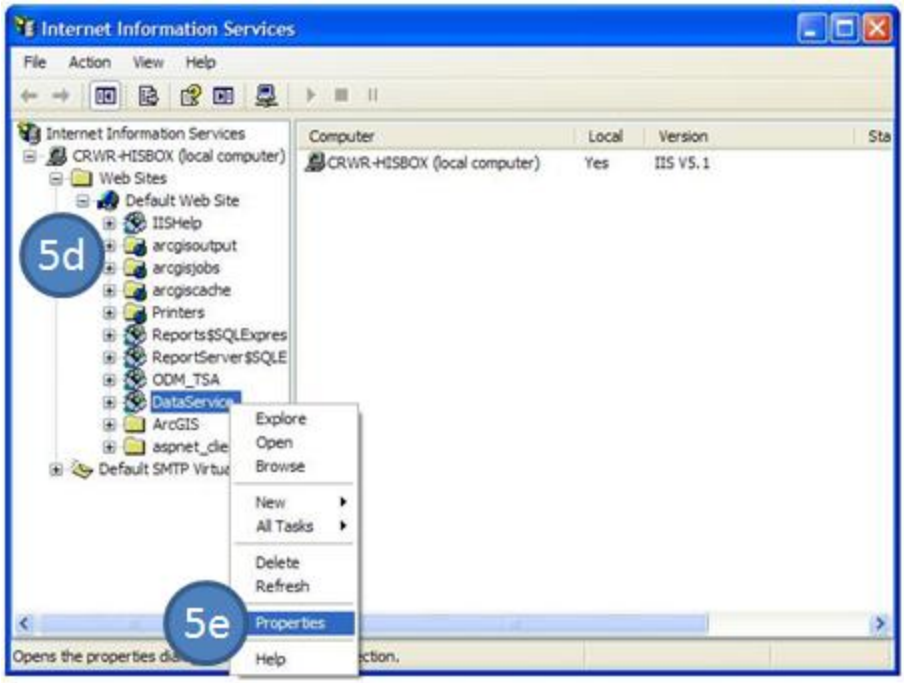

Figure 25 Accessing IIS Properties of the DataService Web Service

f. In the DataService Properties dialog that opens, click the **Directory Security** tab, and then click **Edit**.

g. In the Authentication Methods dialog, click the check box at the top to **allow anonymous access**, and then click **OK** (Figure 26).

Figure 26 Allowing anonymous access to the web service

h. Click **OK** to close the DataService Properties dialog. You may also now **close IIS**.

Your web service is now online. Next you will tell your ODM database to allow the web service to access it, and then you will let your web service know about that ODM database. Think of it as introducing the database and web service to each other.

Setting up a Web Account for your ODM Database

Earlier when you accessed your ODM database with the ODM Data Loader and ODM Tools, you used the SQL Server administrator account, "sa", to connect. This is fine for local applications, but when it comes to web applications, it's a good idea to restrict what the web application can do. Therefore, instead of using an administrator account, you'll set up a special **webclient** account specifically for allowing WaterOneFlow web services to access ODM databases.

To set up a webclient account in SQL Server:

1. If it is not already open, **open** and connect to **SQL Server Management Studio**.
2. In the Object Explorer on the left, expand **Security | Logins** (Figure 27).

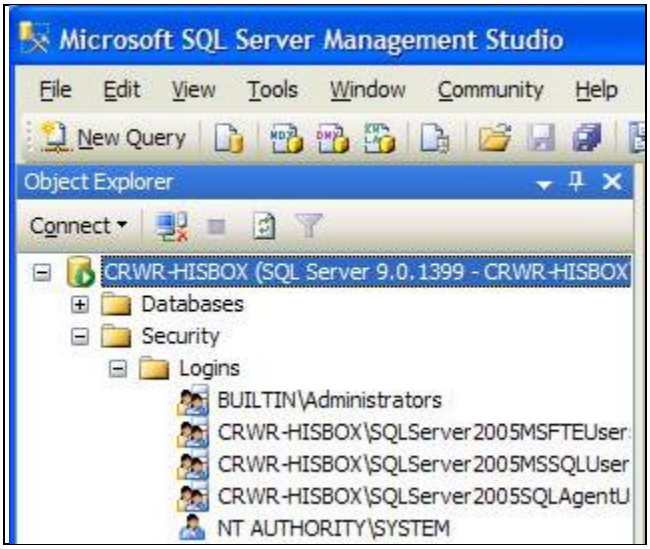

Figure 27 Accessing logins for SQL Server

3. Right click **Logins** and click **New Login**.
4. Type "**webclient**" (without quotes) as the name (Figure 28).
5. Select the **SQL Server authentication** option (Figure 28).
6. Type "**webclient**" (without quotes) as the password (Figure 28). Normally you'd want a more secure password, but for the purposes of this workshop, we're keeping things simple and easy to remember.
7. **Uncheck** the option to **Enforce password policy** (Figure 28). Again, we're choosing to keep things simple for the workshop. You may want to enforce the SQL Server password policy for your own installation.

Figure 28 Creating a webclient SQL Server login

Note
Suppose you've already published one ODM database with WaterOneFlow, and have now created a second database that you want to publish. Since the webclient account was created when you published the first database, you do need to repeat those steps. You can start from the next step with User Mapping.

With the webclient login created, you will now add the BayWQ database to the list of databases that the login can access.

8. In the **Select a page** pane on the left, click **User Mapping** (Figure 29).
9. Place a check next to **BayWQ** and click **OK** (Figure 29).

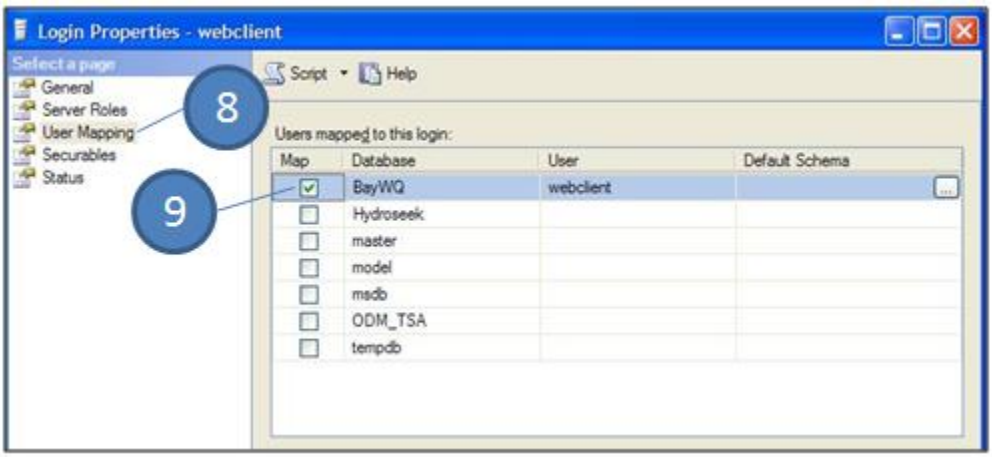

Figure 29 Allow the webclient to access the database

Looking good. You're almost finished. The last thing to do is allow the webclient account to perform Select operations on the database. The web service needs this in order to properly query the database. You allow this permission on the properties page for the database itself.

10. In the Object Explorer on the left, expand Databases until you see your BayWQ database.
11. Right click **BayWQ** and click **Properties**.

12. In the Select a page pane on the left, click **Permissions** (Figure 30). You should see the webclient account in the list of users for the database.
13. In the list of permissions, scroll down and place a check in the **Grant** column for **Select** (Figure 30).
14. Click **OK** to close the dialog (Figure 30). You may also **close SQL Server Management Studio**.

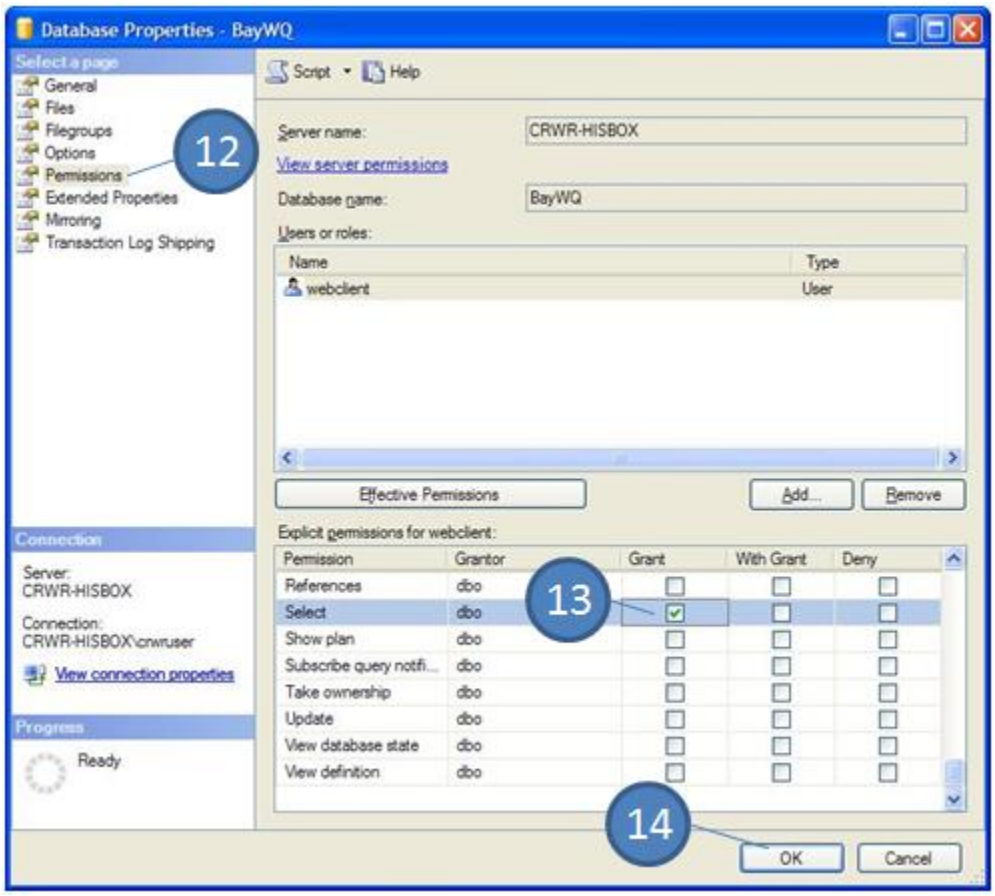

Figure 30 Setting permissions for webclient on the database

Great job! The database now has a special login for working with the web service. The web service will use this login when connecting to the database. Next, you will use this login information to configure the web service so that it knows which ODM database to work with.

CONFIGURING THE WEB SERVICE

Although the web service is up and running, it currently does not know which ODM database it should be pulling its information from. In this section, you will configure the web service to access your BayWQ database.

To configure the WaterOneFlow web service to access an ODM database:

1. In a web browser, navigate to http://localhost/DataService.

Tip
The keyword "localhost" tells your browser to look on your own computer for the web page that you are attempting to access.

The web service displays this web page when accessed with a browser (Figure 31). You can see descriptions of the WaterOneFlow methods, which are used to query the web service for data. There's also a link to configure the web service, which will only function if the link is clicked from the computer hosting the service. In other words, outside users will not be able to change your configuration settings.

Figure 31 WaterOneFlow includes a web page that is displayed when the service is accessed with a web browser

2. Click the **Configuration** link.

On the configuration page, you tell the web service how to connect to your ODM database, and also what prefixes to use for the service Network and Vocabulary. The term **Network** refers to the observations network within which sites in your database participate. The term **Vocabulary** refers to the context within which the variable descriptions from your database apply. For example, if you described your variables using terminology from the USGS National Water Information System, then you would use "NWIS" as the Vocabulary. For this workshop, we'll just use the term "TCWR" for the Network and Vocabulary, since the data come from the Texas Center for Water Resources.

3. Enter "**TCWR**" (without quotes) as the **Network** (Figure 32).
4. Enter "**TCWR**" (without quotes) as the **Vocabulary** (Figure 32).

Next you will enter a connection string to the database. The proper **connection string** depends upon your installation of SQL Server, the name of your database, and the login information for the account that will be used to access the database. There is extensive discussion of connection strings in the document on **Installing & Configuring your own ODM WaterOneFlow Web Services** at http://his.cuahsi.org/wofws.html. If you've followed along with the steps in this workshop and are using a workshop computer or one configured just like it, you will use the connection string below. For other scenarios, please see the documentation linked above.

5. For the connection string, enter "**Data Source = .\SQLExpress; Initial Catalog = BayWQ; User ID = webclient; Password = webclient; Persist Security Info = True**" (without quotes) (Figure 32).
6. Click **Ok** to confirm the settings (Figure 32).
7. Click **Test Database Connection** to see if the settings are correct (Figure 32).

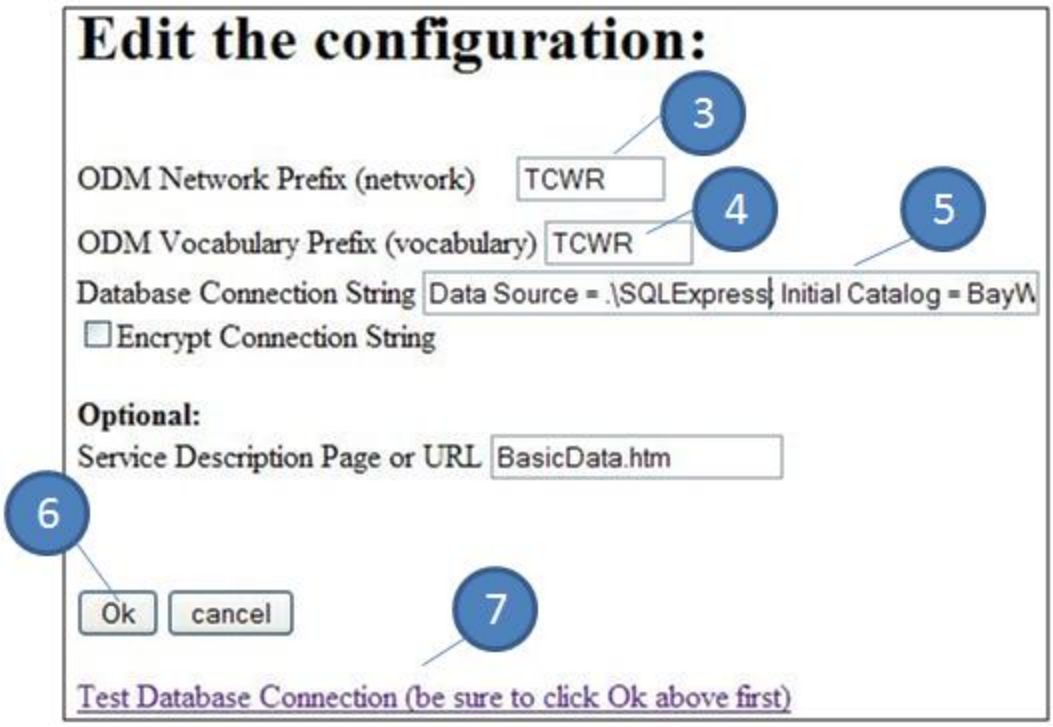

Figure 32 Configuring the web service

Note
When you are setting up a real WaterOneFlow service at your organization, you should click the check box to **Encrypt Connection String** in order to protect the webclient login and password information.

If the settings are correct, you should see a web page that shows some of the data from your database (Figure 33).

This should display up to 10 sites

SiteCode	SiteName
Aransas95_6	GIWW nr Bludworth Island
Aransas95_7	Cedar Dugout
Christmas92_1	San Luis Pass

Figure 33 If configuration is successful, information from your database is shown in the web browser

Excellent! The web service should now be up and running. Let's test the service using HydroExcel.

TESTING THE WEB SERVICE WITH HYDROEXCEL

At this point in the data publication process, you should be able to give someone the URL to your web service, and they should then be able to query data from it using any software that communicates with web services. CUAHSI HIS includes free software called HydroExcel that lets you access WaterOneFlow web services from within a Microsoft Excel spreadsheet. In this portion of the workshop, you'll use HydroExcel to extract data from your web service.

To test the web service with HydroExcel:

1. Download HydroExcel.
    a. In a web browser, navigate to http://his.cuahsi.org.
    b. Follow the links to download HydroExcel.
        i. Under Data Access in the middle of the page, click **HydroExcel**.
        ii. Click the link to download the **1.1.1 version for Microsoft Office 2007**. (The workshop computers have Microsoft Office 2007 installed.)
    c. **Open** the file when it has finished downloading.

> **Note**
> HydroExcel requires the free HydroObjects software to be installed. This software was installed on the workshop computers prior to the workshop. If you are working from a different computer, you can find the installation file at http://his.cuahsi.org/hydroobjects.html.

2. HydroExcel uses macros to communicate with web services. Enable the macros (Figure 34):
    a. Click the **Options** button.
    b. Click **Enable the content**.
    c. Click **OK**.

Figure 34 Enable macros in order to use HydroExcel

The worksheets in HydroExcel call methods from a WaterOneFlow web service to query data and write the result into the spreadsheet. For an in depth tutorial on HydroExcel, see the software manual at

http://his.cuahsi.org/hydroexcel.html. For this workshop, we'll just do a quick test to download sites, variables, and a time series.

3. Activate the **Data Source** worksheet. (Click "Data Source" at the bottom of HydroExcel.)

On the Data Source worksheet, you tell HydroExcel which web service you want to work with by inputting the URL address of the service next to the box that says "WSDL Location". Some URLs are already listed in the spreadsheet, but you will have to locate your own URL that points to the web service you just created. Since you're still using the same computer on which the web service is installed, you can just use the localhost URL. But you'll need to provide the URL of the actual service description to HydroExcel, which is linked from that initial page you navigated to earlier.

4. In a web browser, navigate to http://localhost/DataService.
5. Under the heading "Service Description" click the link for **Service Description**. This takes you to the WSDL for your web service.

Tip

The service WSDL (Web Services Description Language) is where your web service defines what it can do and how programs can interact with it. It is designed for programs to read, so don't worry if you can't make sense of it. When a program accesses your web service, it will read the WSDL and know exactly how to send requests to it, and will also know what format of output to expect back.

6. From the address bar of your web browser, **copy the URL** of the WSDL and paste it into the cell next to the cell that says "**WSDL Location**" on the Data Source worksheet of HydroExcel (Figure 35).

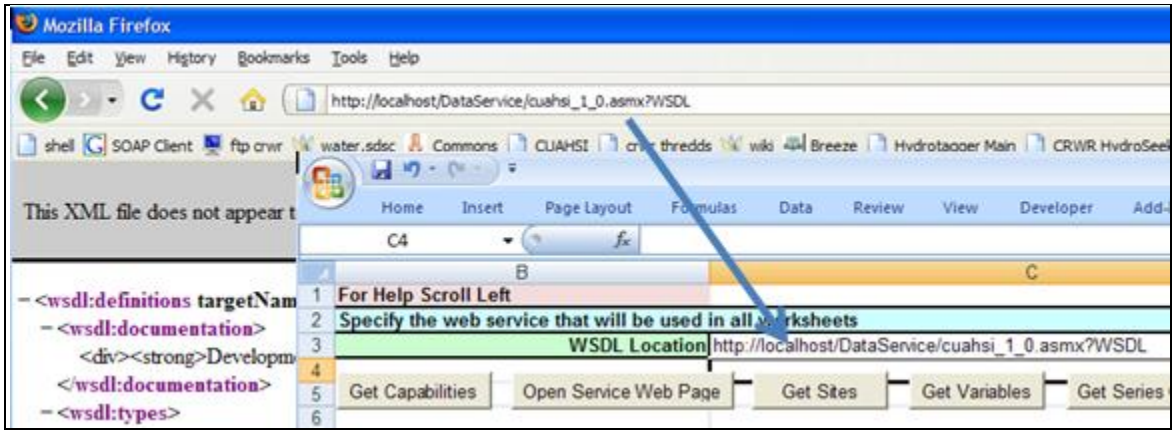

Figure 35 Use the Data Source worksheet to tell HydroExcel which web service to work with

7. Activate the **Series Catalog** worksheet.
8. Change the option to **Create and open KML file after download** to **TRUE** (Figure 36).
9. Click **Get Series Catalog** (Figure 36).

After a moment, your spreadsheet is updated with information about the sites and variables measured at the sites. Also, Google Earth opens to show the site locations.

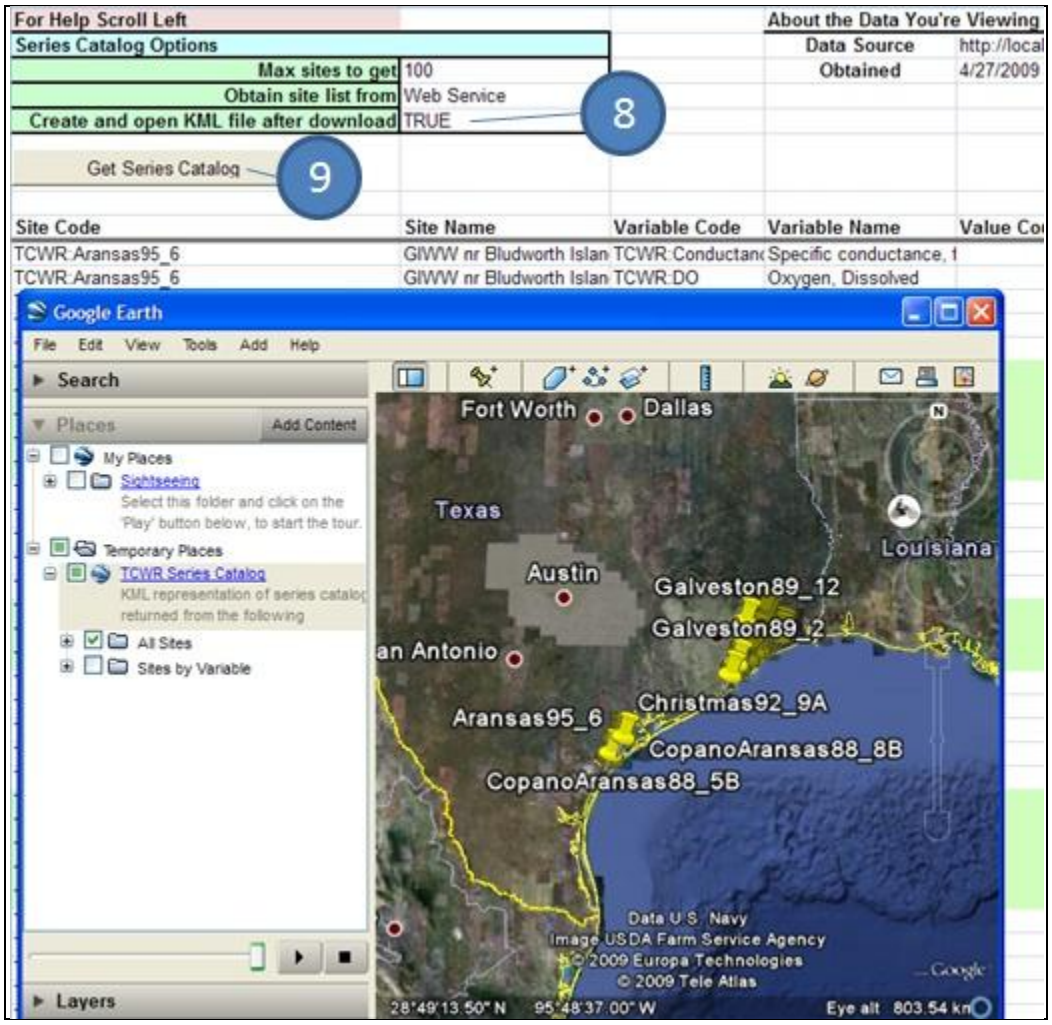

Figure 36 HydroExcel can show site locations in whatever KML viewer you have installed

10. Browse around Google Earth to see your site locations. Click on placemarks for sites to see information associated with the sites.
11. Switch back to **HydroExcel**.
12. Dismiss the message box indicating that the download is complete.

Take a look at the information in the Series Catalog. You not only get information about the location of sites in your database, but also the variables measured at those sites. Notice that the start date, end date, and number of records of time series observations are included with each variable. You'll use this information to download a time series for one of the sites.

13. Locate a site and variable for which you'd like to download time series data. For my screenshots, I'm going to get temperature data at Rattlesnake Point, so you may want to do the same.
14. In the Series Catalog worksheet, right click anywhere on the row for the site and variable that you want.
15. In the context menu that opens, point to HydroExcel, and then click to download the time series.

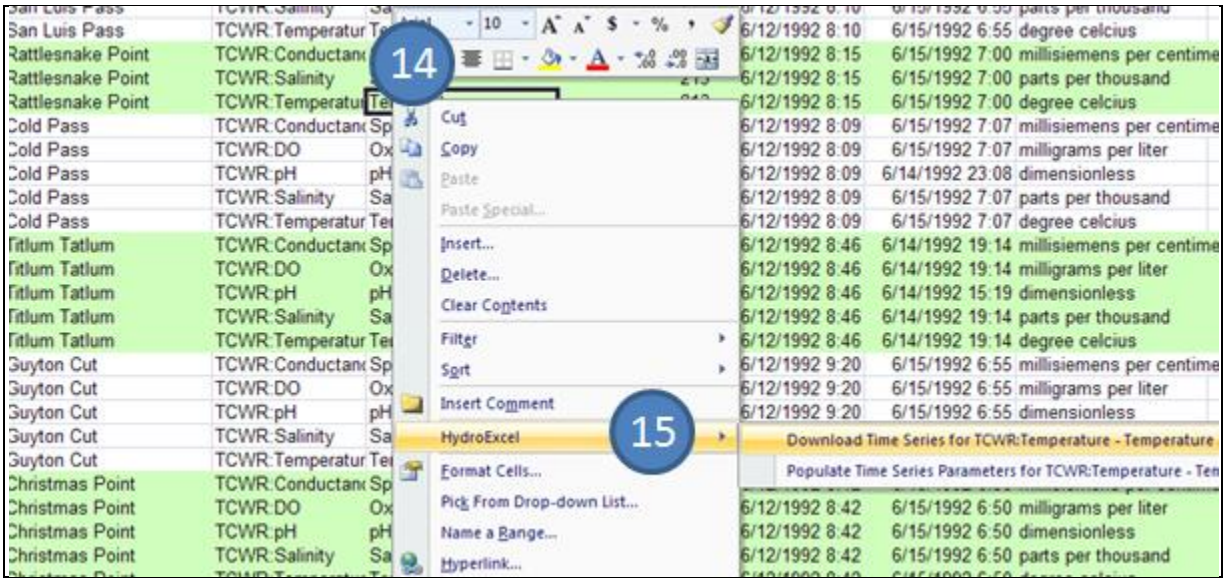

Figure 37 Right click menus are a handy way of make queries for data in HydroExcel

You are brought to the Time Series worksheet, where HydroExcel filled in the parameters to make the request for the data, called the web service, and populated the result in the spreadsheet.

16. Dismiss the message box indicating the download is complete.

Notice there are several measurements taken at a given datetime, but at different depths below the water surface (Figure 38).

Get Values Options	☑ Ignore NoData Value
Site Code/Location	TCWR:Christmas92_10
Variable Code	TCWR:Temperature
Start Date	6/12/1992 8:15
End Date	6/15/1992 7:00

Get Values

Variable Name	Temperature
Units	degree celcius
SampleMedium	Surface Water
Site Name	Rattlesnake Poin
Latitude	29.02833333
Longitude	-95.21
Data Source	http://localhost/D
Obtained	4/27/2009 17:02
Ignore NoDataValue	TRUE

DateTime	Value	Qualifier	Offset	Offset Units
6/12/1992 8:15	27		1.4	international foot
6/12/1992 8:15	27		3.5	international foot
6/12/1992 8:15	27		5.6	international foot
6/12/1992 9:00	27		1.4	international foot
6/12/1992 9:00	27.3		3.5	international foot
6/12/1992 9:00	27.3		5	international foot

Figure 38 Time series returned from the web service

17. Activate the **Statistics and Charts** worksheet to see a quick plot of the data (Figure 39). The plot shows max, min, and average temperature for each datetime across all depths for the given site.

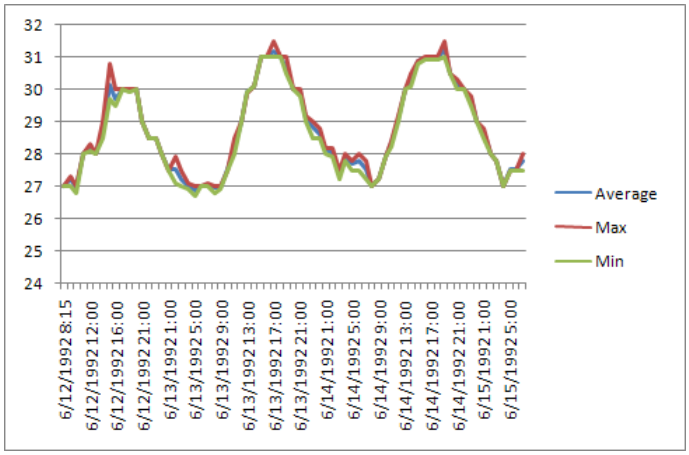

Figure 39 A plot of temperature reveals the diurnal variation over a three-day period

Kudos to you! You've come a long way, and now have successfully made your data available online using WaterOneFlow. Do you feel the magic in this moment? I certainly do. The next step would be to register your service with Texas HIS at http://waterdatafortexas.org so that others can discover it. This may be facilitated by the Texas HIS Viewer, which is still in development at the time of this writing. Contact the Texas Water Development Board or Tim Whiteaker (twhit@mail.utexas.edu) for more information about registration with Texas HIS.

Congratulations! You have now completed the HIS data publication process. You started from raw data, loading the data into an ODM database, wrapped web services around the database, and verified that the service was working by using HydroExcel.

If you'd like to know more about the data publication process, please visit the HIS website at http://his.cuahsi.org and also the Texas HIS Viewer at http://waterdatafortexas.org. You can check for new releases of helpful HIS tools, and even partner with the Texas HIS team in developing the next generation of HIS software and methods to further hydrologic information science.

This concludes the workshop.