HIS document 2

Loading Observations Data with the ODDataLoader (version 1.0)

A guide to using CUAHSI's ODDataLoader tool for loading observations data into an Observations Data Model compliant database

July 2007

Prepared by:

Ernest S. C. To and Tim Whiteaker
Center for Research in Water Resources
University of Texas at Austin

David Valentine
San Diego Supercomputer Center
University of California at San Diego

Distribution

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of the University of California, San Diego, San Diego Supercomputer Center nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

CUAHSI's ODDataLoader is documented at the following URL:
http://water.sdsc.edu/ODDataloader/

Disclaimers

Although much effort has been expended in the development and testing of the ODDataLoader, errors and inadequacies may still occur. Users must make the final evaluation as to the usefulness of ODDataLoader for his or her application.

WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

The ODDataLoader software and this documentation are based upon work supported by the National Science Foundation under Grant No. EAR-0413265. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Acknowledgements

The team of engineers, scientists and research assistants that contributed to this document includes:

Tim Whiteaker (editor), Research Associate, Center for Research in Water Resources, University of Texas, Austin, TX.
David Valentine, GIS Programmer, San Diego Supercomputer Center, University of California at San Diego, La Jolla, CA.
Ernest To, Doctoral Candidate, Center for Research in Water Resources, University of Texas, Austin, TX.

Warranty

This software is supplied "AS IS". The author(s) disclaim all warranties, expressed or implied, including, without limitation, the warranties of merchantability and of fitness for any purpose. The author(s) assume no liability for direct, indirect, incidental, special, exemplary, or consequential damages, which may result from the use of this software, even if advised of the possibility of such damage.

Funding

Funding for ODDataLoader and this document was provided by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) under NSF Grant No. EAR-0413265. In addition, much input and feedback has been received from the CUAHSI Hydrologic Information System development team. Their contribution is acknowledged here.

Technical Support

There is no formal ongoing support for the freely distributed ODDataLoader software. However, we are interested in feedback. If you find errors, have suggestions, or are interested in any later versions, please contact:

David Valentine
University of California, San Diego
San Diego Supercomputer Center, MC 0505
9500 Gilman Drive
La Jolla, CA 92093-0505
valentin@sdsc.edu

Table of Contents

1.0 Introduction

The Observations Data Model (ODM) defines a database structure for storing hydrologic observations time series data, along with metadata such as site information, variable description, methods used and more. This design is implemented with a SQL database, using either SQL Express or SQL Server 2005.

There are a number of ways for importing your observations data into an Observations Database (OD), which is a database that is ODM compliant. These tools range from the new SQL Server Integration Services (SSIS) to manual entry. However, for the database administrator new to SQL and/or new to ODM, the process of loading data may seem difficult.

To alleviate simple data loading, CUAHSI provides the ODDataLoader tool. While not as robust as SSIS, the ODDataLoader does not have as steep of a learning curve, and is useful when the data do not have quirks that could only be handled with more sophisticated data loading means.

This tutorial demonstrates the use of the ODDataLoader tool to import observations data into an ODM database. In this tutorial, we will use a comma delimited text file as the input file and an ODM SQL database as the output.

2.0 Computer and Skill Requirements

To complete this tutorial, your computer must meet the following requirements:

1. Windows 2000 or above
2. Internet connection
3. Microsoft .NET Framework version 2.0 (x86) (this is available at: http://www.microsoft.com/downloads/details.aspx?FamilyID=0856eacb-4362-4b0d-8edd-aab15c5e04f5&DisplayLang=en)
4. SQL Server:
    a. Microsoft SQL Server 2005, or
    b. SQL Server Express
        i. Microsoft SQL Server 2005 Express Edition (http://msdn.microsoft.com/vstudio/express/sql/download/)
        ii. SQL Server Management Studio Express 2005 (http://msdn.microsoft.com/vstudio/express/sql/download/)
5. Basic knowledge of the CUAHSI's ODM schema.

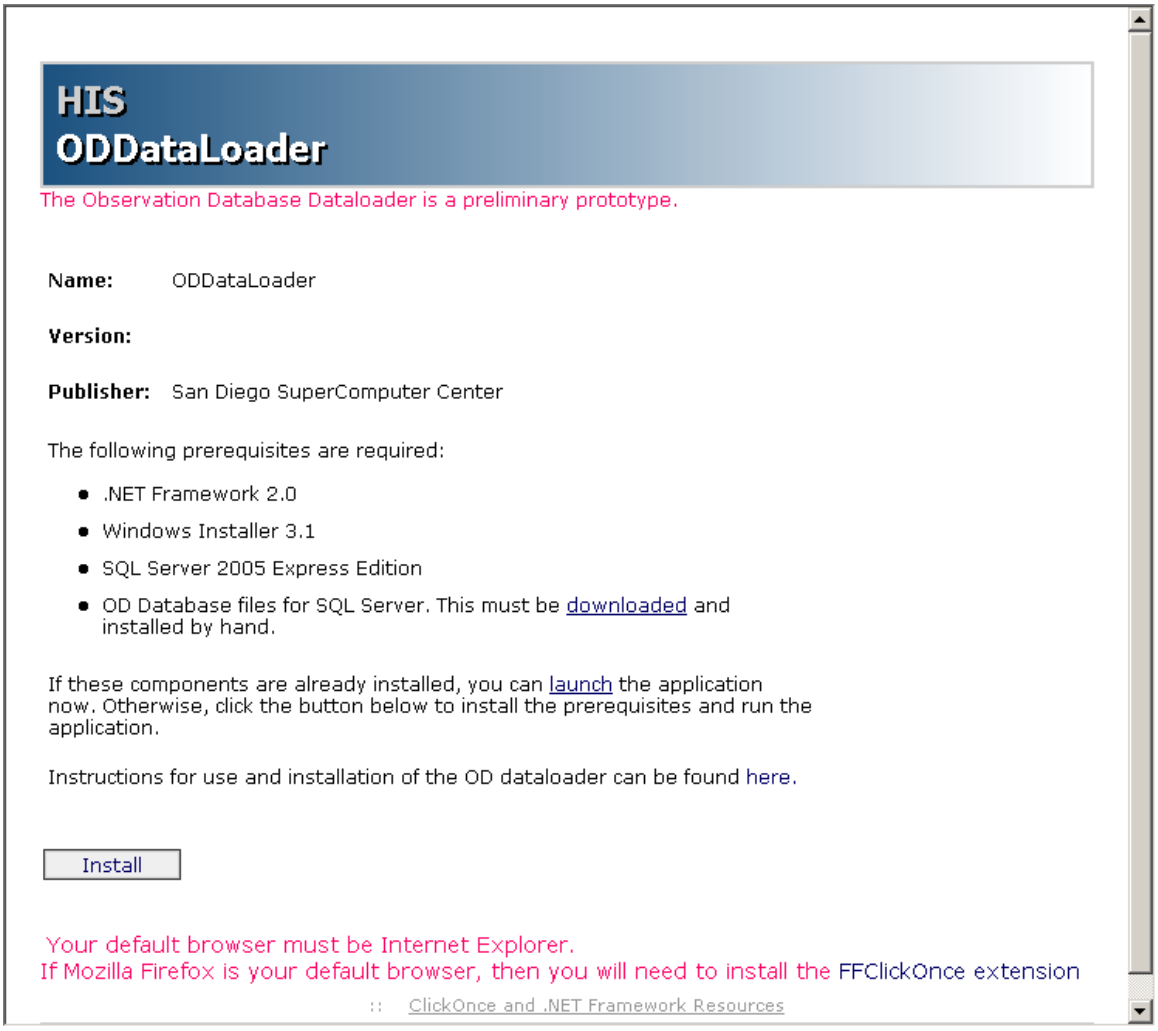
The .NET Framework and SQL Server Express can be downloaded for free at the links provided above. It is strongly recommended that the user install the .NET Framework (and SQL Server or SQL Express) first before downloading ODDataLoader.

NOTE: If you don't already have SQL Server installed, be aware that the installation process may take 30 minutes or longer. During installation, you will be prompted to use Mixed Authentication or Windows Authentication. If you wish to use the ODM Tools (another CUAHSI product) with your SQL databases, you must select Mixed Authentication. If unclear about the other installation options, please consult your database administrator.

3.0 Installation

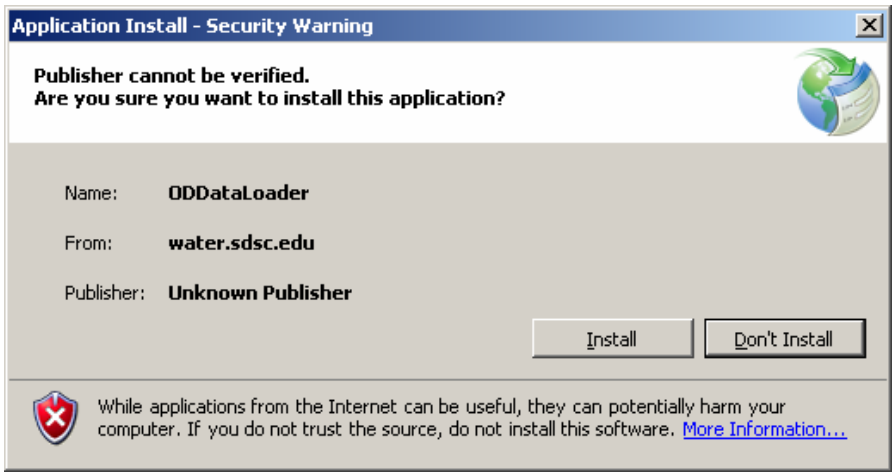
Please note that the ODDataloader website uses the ClickOnce installer to download, setup and open the application on the user's computer. In order to use the ClickOnce installer properly you either need to have Internet Explorer 6.0 (IE) or above as the default web browser, or if using Firefox, you must install an extension called "Firefox FFClickOnce". Instructions for installing "FireFFClickOnce" are on the ODDataLoader website.

1. Once you have all the prerequisites listed under Computer and Skill Requirements, go to the website for ODDataloader at http://water.sdsc.edu/ODDataloader/

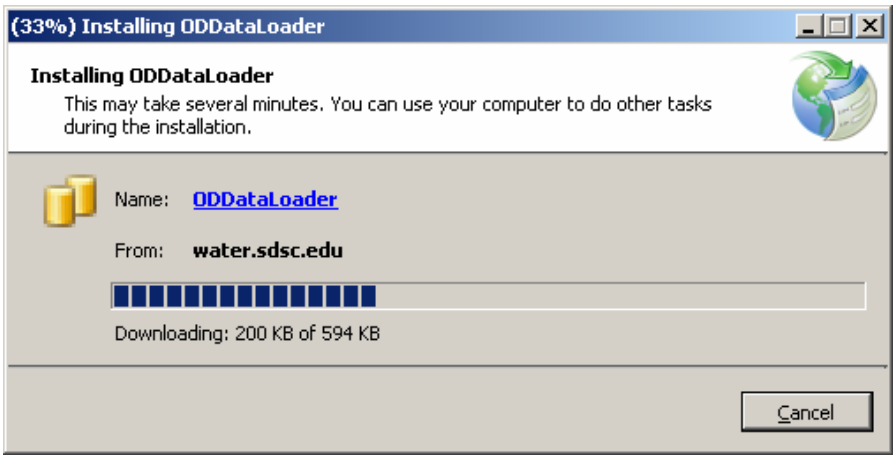


2. Click on the Install button, and run the Setup.exe file. If you are using Firefox, you will have to click the Run ClickOnce Application button.

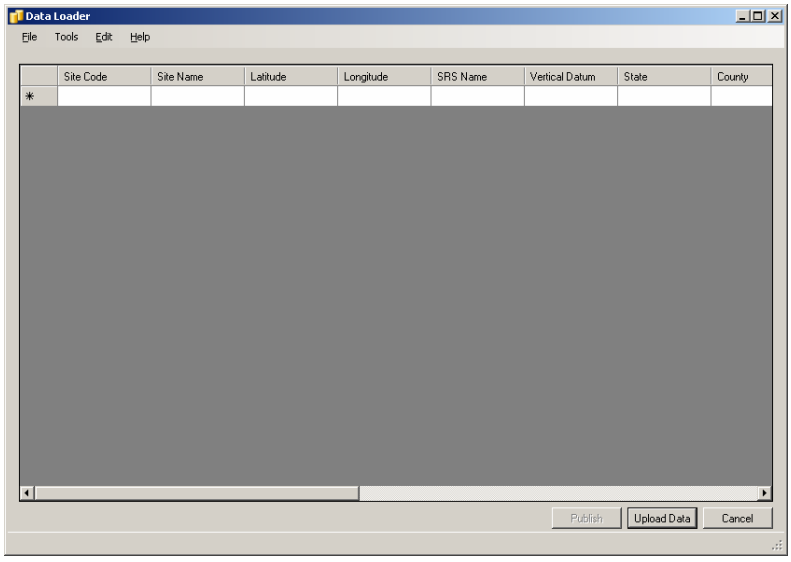
The following window will pop up asking whether you want to run the setup file for ODDataLoader.



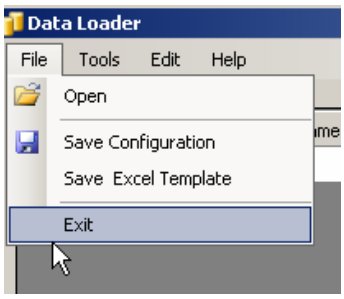
3. Click Install and the ODDataLoader setup files will be downloaded and installed onto your computer.



The program will open itself once installation is completed.



4. For now, exit the program clicking the File menu, and then clicking Exit.



In the next sections, you will set up the spreadsheet of input data, and the database that you will load data into.

4.0 Setting up Input Files

In this tutorial, you will use an example spreadsheet, formatted as a comma delimited text file, containing some aquatic biological data (courtesy of Dr. Tim Bonner, Texas State University and Eric Hersh, University of Texas at Austin) as the input file.

1. Download the spreadsheet from the following link, and save it to your local hard drive.

ftp://ftp.crwr.utexas.edu/pub/outgoing/CUAHSI/ODM_tutorial/BonnerData.csv

2. Open the spreadsheet with Excel. If you do not have Excel installed, you may use another spreadsheet program or simply a text editor such as Wordpad.

Peruse the spreadsheet.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	SeriesID	ValueID	DataValue	ValueAccu	LocalDateTim	UTCOffset	SiteCode	SiteName	Latitude	Longitude	SRSName	VariableCo	VariableNa	VariableUr	VariableUr
2			0		10/1/03 0:00	-5	1	1	30.09816	-98.5467	NAD83	200000	Mean Velocity	ters per secc	m/s
3			0.01		10/1/03 0:00	-5	2	2	30.10281	-98.5076	NAD83	200000	Mean Velocity	ters per secc	m/s
4			0		10/1/03 0:00	-5	3	3	30.08329	-98.3226	NAD83	200000	Mean Velocity	ters per secc	m/s
5			0		10/1/03 0:00	-5	4	4	29.98379	-98.0548	NAD83	200000	Mean Velocity	ters per secc	m/s
6			0.265891		10/1/03 0:00	-5	5	5	29.98379	-98.0548	NAD83	200000	Mean Velocity	ters per secc	m/s
7			0.598		10/1/03 0:00	-5	6	6	30.00611	-97.9776	NAD83	200000	Mean Velocity	ters per secc	m/s
8			0.046667		10/1/03 0:00	-5	7	7	29.93637	-97.8948	NAD83	200000	Mean Velocity	ters per secc	m/s
9			0.57		10/1/03 0:00	-5	8	8	29.87593	-97.9134	NAD83	200000	Mean Velocity	ters per secc	m/s
10			0		10/1/03 0:00	-5	9	9	30.02935	-98.2659	NAD83	200000	Mean Velocity	ters per secc	m/s
11			0.03		10/1/03 0:00	-5	10	10	30.00239	-98.0901	NAD83	200000	Mean Velocity	ters per secc	m/s
12			0.128539		1/1/04 0:00	-6	1	1	30.09816	-98.5467	NAD83	200000	Mean Velocity	ters per secc	m/s
13			0.01		1/1/04 0:00	-6	2	2	30.10281	-98.5076	NAD83	200000	Mean Velocity	ters per secc	m/s
14			0		1/1/04 0:00	-6	3	3	30.08329	-98.3226	NAD83	200000	Mean Velocity	ters per secc	m/s
15			0		1/1/04 0:00	-6	4	4	29.98379	-98.0548	NAD83	200000	Mean Velocity	ters per secc	m/s
16			0.037795		1/1/04 0:00	-6	5	5	29.98379	-98.0548	NAD83	200000	Mean Velocity	ters per secc	m/s
17			0.486325		1/1/04 0:00	-6	6	6	30.00611	-97.9776	NAD83	200000	Mean Velocity	ters per secc	m/s

The columns in this spreadsheet match those that the ODDataLoader will look for during the import process. The spreadsheet includes time series values and some associated metadata in the same row as each value, resulting in one large flat table. Notice that not all of the columns need to be filled out in this spreadsheet. The data loader will use default values in some cases if no values are provided. See your ODDataLoader documentation for more information about how the data loader responds to the content in the input spreadsheet.

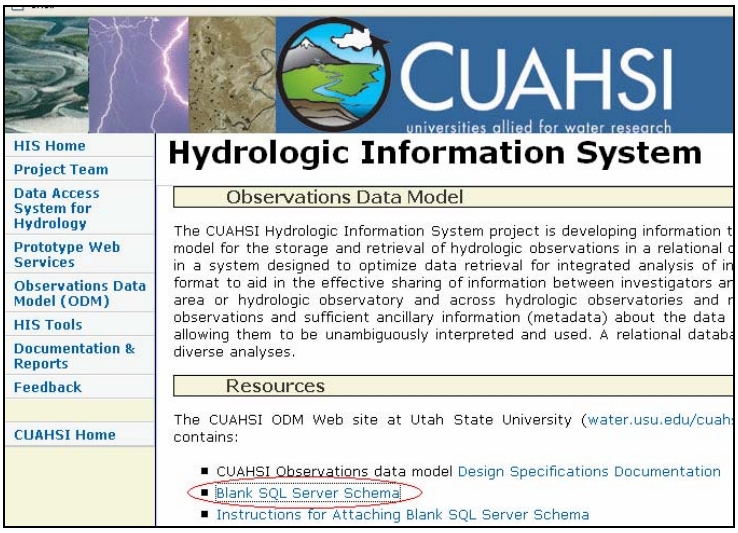
NOTE: To generate your own template spreadsheet for data import, start the ODDataLoader, then click the File menu, and then click Save Template File. You can choose from a variety of export formats.

3. Close the spreadsheet.

5.0 Setting up Output Files

The ODDataLoader cannot generate an ODM database for you. To use the ODDataLoader, you must already have an ODM compliant database. A blank ODM database is available for download from the ODM website. You will go there to download a copy of the database, and then attach it using SQL Server Management Studio.

1. Go to the website at http://www.cuahsi.org/his/odm.html.
2. Click the "Blank SQL Server Schema" link.



3. The database will come as a zip file. Save the zip file to your computer and open it.
4. Extract the OD.mdf and OD_log.ldf files from the zip file, and then close the zip file.

NOTE: You can extract the database to any location on disk. However, if you do so and you have connected to SQL Server using SQL Server authentication and not Windows Authentication, you will have to give SQL Server access to read and write to the folder where you extracted your database prior to attaching it. SQL Server already has access to its default data folder using either SQL Server or Windows authentication and so this is the easiest location in which to work. The default SQL Server data folder is located at e.g., C:\Program Files\Microsoft SQL Server\MSSQL.1\MSSQL\Data\.

Now that the database has been downloaded to your computer, the next step is to attach to the SQL database engine running on your computer. As this is a common task when working with HIS products, instructions have already been prepared and are available on the CUAHSI website.

5. Go to the website at http://www.cuahsi.org/his/odm.html.
6. Click the "Instructions for Attaching Blank SQL Server Schema" link.
7. Follow the instructions to attach the database that you just downloaded.

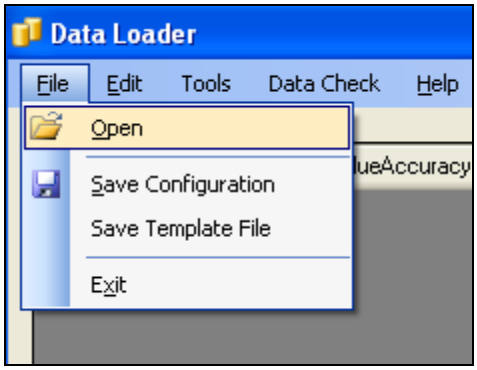
6.0 Loading Data with ODDataLoader

In this section, you will you load data into the OD database from values in a spreadsheet.

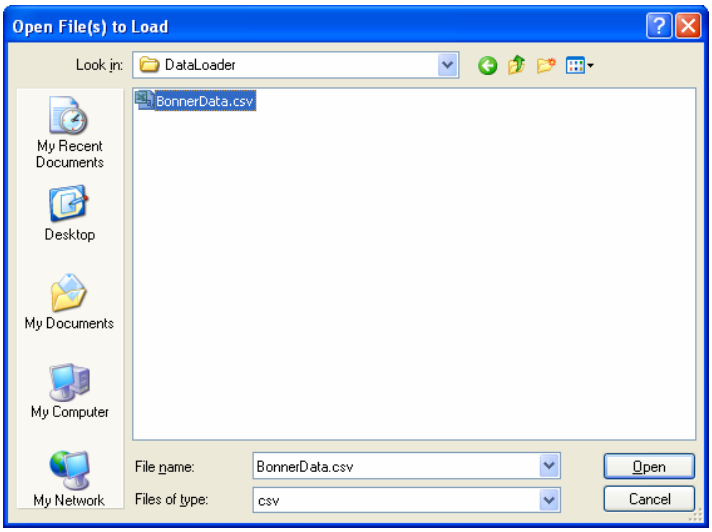
6.1 Loading Data

1. Start the ODDataLoader (Typically Start --- All Programs --- CUAHSI HIS --- ODDataLoader.

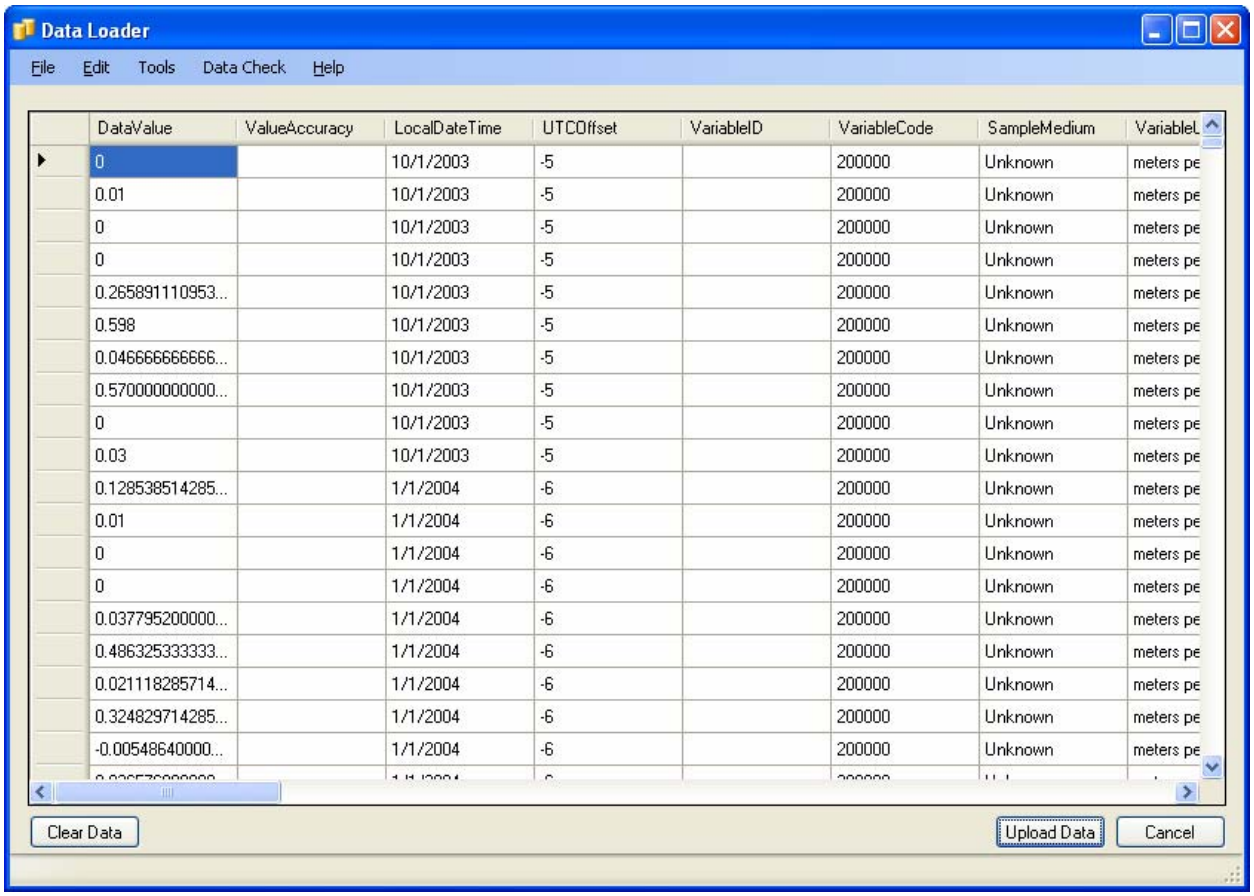
2. Click the File menu, and then click Open.



3. Choose ".csv" as the file type, then navigate to the spreadsheet with our data, and click Open.

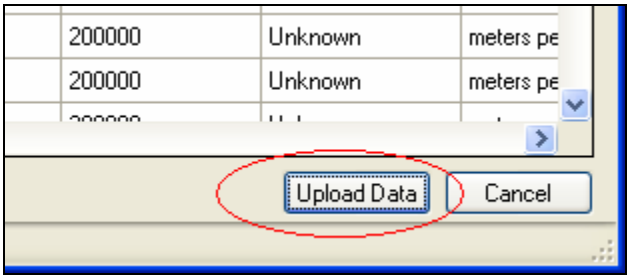


In a few moments, the ODDataLoader populates its on table with information read from the spreadsheet.



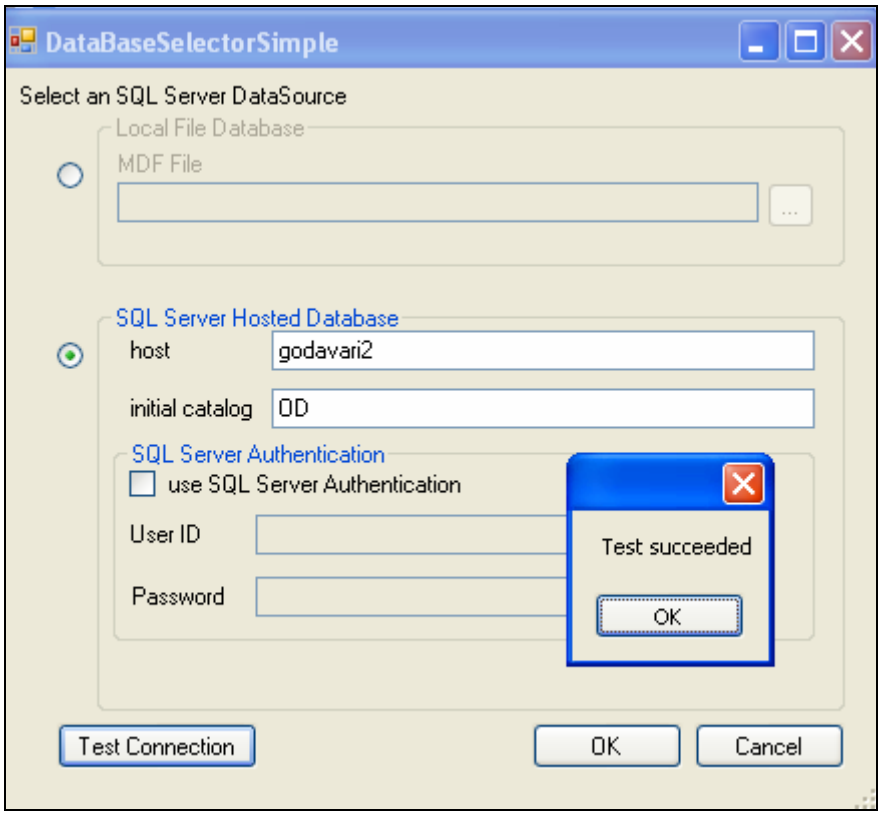
The ODDataLoader has also populated some default values, which were left blank in the spreadsheet. For example, notice that the values in the SampleMedium column now read "Unknown", instead of a blank as in the spreadsheet.

4. Click Upload Data.



The DataBaseSelectorSimple dialog will open. In this dialog, you specify to which database you want to upload the data.

5. Select SQL Server Hosted Database as the data source. Type in your host name (this should be the same name that you connected to earlier when using the Management Studio.)
6. For the initial catalog, type the name of the database, e.g. “OD”.
7. Click the Test Connection button to make sure a connection can be established. If the test is not successful, then make sure that you have entered the host and initial catalog correctly.



8. Click OK in the DataBaseSelectorSimple dialog to perform the upload.
9. The dialog will disappear when the tool finishes. When this happen, close the ODDataLoader.

6.2 Checking Results

1. Open the SQL Management Studio and connect to the database as you did before when attaching the database.
2. In the Object Explorer, click the plus sign next to Databases to expand that item. Then expand OD (the name of the database), and then expand Tables.
3. Right click on the DataValues table and click Open Table.

Notice that the table is populated with values from the spreadsheet.

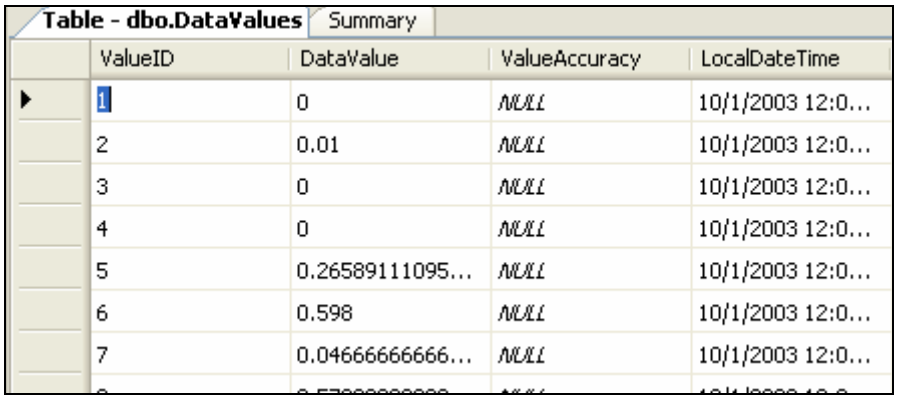


Table - dbo.DataValues | Summary

ValueID	DataValue	ValueAccuracy	LocalDateTime
1	0	NULL	10/1/2003 12:0...
2	0.01	NULL	10/1/2003 12:0...
3	0	NULL	10/1/2003 12:0...
4	0	NULL	10/1/2003 12:0...
5	0.26589111095...	NULL	10/1/2003 12:0...
6	0.598	NULL	10/1/2003 12:0...
7	0.04666666666...	NULL	10/1/2003 12:0...

4. Open the Variables table.

The ODDataLoader not only populated the DataValues table, but also the relevant metadata tables.

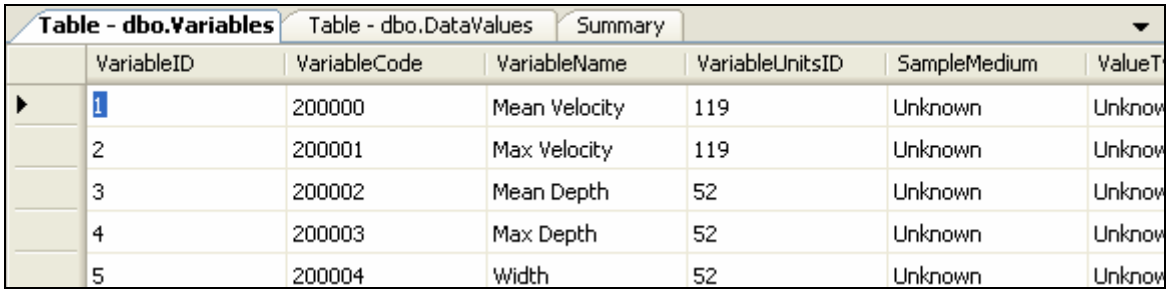


Table - dbo.Variables | Table - dbo.DataValues | Summary

VariableID	VariableCode	VariableName	VariableUnitsID	SampleMedium	ValueT
1	200000	Mean Velocity	119	Unknown	Unknov
2	200001	Max Velocity	119	Unknown	Unknov
3	200002	Mean Depth	52	Unknown	Unknov
4	200003	Max Depth	52	Unknown	Unknov
5	200004	Width	52	Unknown	Unknov

5. Open the SeriesCatalog table.

This table is one of the most important tables that the ODDataLoader populates. The ODDataLoader looks through the data to be uploaded, and determines the series that are present, where a series is defined by a particular variable being measured at a particular site. These series are then cataloged in the database, so that we have a very quick way of getting a sense of what is in this database.

NOTE: The way we uniquely define series in ODM is actually a bit more complex, but we'll that out of this discussion. See ODM documentation for more information on how series are defined.

6. Close SQL Management Studio.

Congratulations! You have completed this exercise, and now know how to load data into an ODM database using the ODDataLoader.