FINAL REPORT

DOE ER64982

Establishment of the Northeast Coastal Watershed Geospatial Data Network (NECWGDN)

TEAM MEMBERS:
Dr. Zong-Guo Xia (Principal Investigator)
Dr. Robyn Hannigan (Lead Scientist)
Dr. Wei Ding (Technical Lead)
Dr. Bob Bowen (Senior Personnel)
Dr. Meng Zhou (Senior Personnel)
Dr. Mingshun Jiang (Senior Personnel)
Dr. Cascade Sorte (Post-Doctoral Investigator)
Ms. Marin Kress (PhD Student, School for the Environment)
Mr. Dai Wei (PhD student, Comp Sci)
Ms. Pradnya Khutafale (MS Student, Comp Sci, graduated May 2011)
MS. Siyi Liu (MS Student, Comp Sci, graduated May 2011)
Dr. Rick Hooper (Consultant, CUAHSI)
Ms. Yoori Choi (Consultant, CUAHSI)

Contents

SUMMARY OF ACTIVITIES:

The goals of NECWGDN were to establish integrated geospatial databases that interfaced with existing open-source (water.html) environmental data server technologies (e.g., HydroDesktop) and included ecological and human data to enable evaluation, prediction, and adaptation in coastal environments to climate- and human-induced threats to the coastal marine resources within the Gulf of Maine. We have completed the development and testing of a "test bed" architecture that is compatible with HydroDesktop and have identified key metadata structures that will enable seamless integration and delivery of environmental, ecological, and human data as well as models to predict threats to end-users. Uniquely this database integrates point as well as model data and so offers capacities to end-users that are unique among databases. Future efforts will focus on the development of integrated environmental-human dimension models that can serve, in near real time, visualizations of threats to coastal resources and habitats.

Our efforts centered on (1) integration of data architecture with CUAHSI Water.HTML based Hydrodesktop architecture, (2) establishment of servers and database architecture, (3) identification of metadata for inclusion of modeled data into databases, (4) identification of key ecosystem indicator data for association with environmental data, (5) identification of human-dimension data and human health risk indicators for integration into the database, and (6) internal testing of database architecture.

For development of the test bed we identified the Gulf of Maine (Figure 1) as the geographic boundary. This would be expanded in future years but narrow definition of geographic interest enabled us to focus efforts on coastal watersheds of this region and acquire appropriate hydro/water data from various agencies/organizations like MWRA, USGS, EPA, CUAHSI and other regional watershed associations.

Figure 1. Geographic boundary of test bed (A - Cape Cod; B - Stellwagen Bank).

We had, on hand, hydro data received from MWRA (1998 to 2005) and river data from USGS. This data was used to create the two water models namely:

- HydroDynamic model
- Water Quality model

These models are used to run on the input data (meteorological data, others) and the model output was compared with observational data and conclusions are drawn from it. The models produce 3-dimensional physical (T, S, currents, tides) and water quality (nutrients, chlorophyll, and productivity) data. The model output data is currently saved in .netcdf format files. All of this data is currently saved on School for the Environment Unix server and some PCs.

Observational data was acquired from the Massachusetts Water Resources Authority (MWRA) in zip format and was not provided in a pre-defined or fixed format. The other source of observational data (hydrologic data) were acquired from USGS from coastal watersheds within the Gulf of Maine.

DATABASE DEVELOPMENT

Given that our goal is to develop a geospatially references database that supports distribution and modeling of hydrological, coastal oceanographic, ecological and human data we needed to create a new database structure with unique capacities. Using metadata and data provided by the data team (see below) we developed a database that integrated metadata from disparate sources (e.g., MWRA, USGS, model output). We also ensured that the data structure was compatible with HydroDesktop and could integrate putput generated by HydroDynamic models. We worked collaboratively within the team and with computer scientists associated with the CUAHSI Informatics group that manages HydroDesktop. This included writing a program that would upload observational data into our database and HydroDesktop and do so in the background of other data service activities. This was particularly done to accommodate the upload of model output data for which there is no current metadata framework within water.html.

For uploading of model data into ours and other databases we wrote a batch program that would look at the model data for inconsistencies in output format when compared with the target database and would report errors to the user. Uploaded model data would be uploadable in 3 formats.

(1) ASCII format - batch program would upload the model data output created by the two water models - HyrdoDynamic model and Water Quality model. The model data output resides in two separate tables in the new database.

b) netCDF - then the program will read those netCDF files and upload model data to new database tables.

c) ArcGIS 9.2 - Supports netcdf format. So the other option is to save these netCDF files in ArcGIS(table format). The users can query and retrieve the desired model data from ArcGIS server.

Regardless of format for uploading, once the model data is uploaded to database or stored on

ArcGIS, the user can query and select data(columns etc.) from the database/ArcGIS. The data retrieved from query is provided in downloadable format.

We designed the model output-upload program to have a user interface similar to the one provided on USGS site http://waterdata.usgs.gov/nwis/gw. This interface enables user to select parameters and then run query on uploaded data. Finally geo-database schema was setup on an ArcGIS server to use data from the new database server. Once fully tested the system will be made available as a part of UMB network. We are also considering a standalone web data service , rather than direct integration with HydroDesktop which could also be provided for faculty members or students at UMass Boston which will serve hydrologic information along with maps.

We wrote additional scripts to support upload new, non-server based, observational and model data. Observational data must be in excel format and is uploaded using a specially written java script with no special settings required to run those programs(except Java class path etc). Model output data in .netcdf files were integrated with Arc Map 10 using a specially written Python /NumPy script that enabled direct upload to the SQL Server.

In Appendix A is provided information about the development phase of the server.

DESIGN OF A CUAHSI-HYDRODESKTOP INTEGRATED SERVER

A number of options were explored including a UMB data server or a CUAHSI-based server. After considering various implementation models, the data team decided to follow the ICEWATER[1] project (Utah University and several other universities) model. In our model also, all the hydro data (observational and model output data) will reside on the database of UMass Boston Server (called as Hydro Server). The UMass Boston Hydro-Server will provide a web interface and web services to access the data stored on the Hydro-Server and also host ArcGIS which will provide geospatial data capabilities. (Note: At this point, the geo-spatial database requirements are unclear, and the report will be revised when requirements are clarified).

This UMass Boston Hydro-Server will be made part of CUAHSI network. All the requests for UMass Boston -MWRA data made by CUAHSI users (users accessing CUAHSI website) will be redirected to UMass Boston Hydro-Server. The UMass Boston Hydro-Server (Figure 2) will provide the requested data using its web- services to users accessing CUAHSI website.

[1] ICEWATER project has also collaborated with CUAHSI for sharing hydro data to research communities. But in ICEWATER model, instead of hydro data residing on CUAHSI Hydro Server, data resides on University's server (Utah and few other university's servers). Please visit http://icewater.usu.edu/ for more information. ICEWATER project has provided web-interface to retrieve this data. ICEWATER has registered its web services with CUAHSI (WaterOneFlow web services). So when CUAHSI users wants to retrieved ICEWATER data, a request is send from CUAHSI Server to ICEWATER Hydro Server and data is provided by ICEWATER HydroServer to CUAHSI users through the web services .

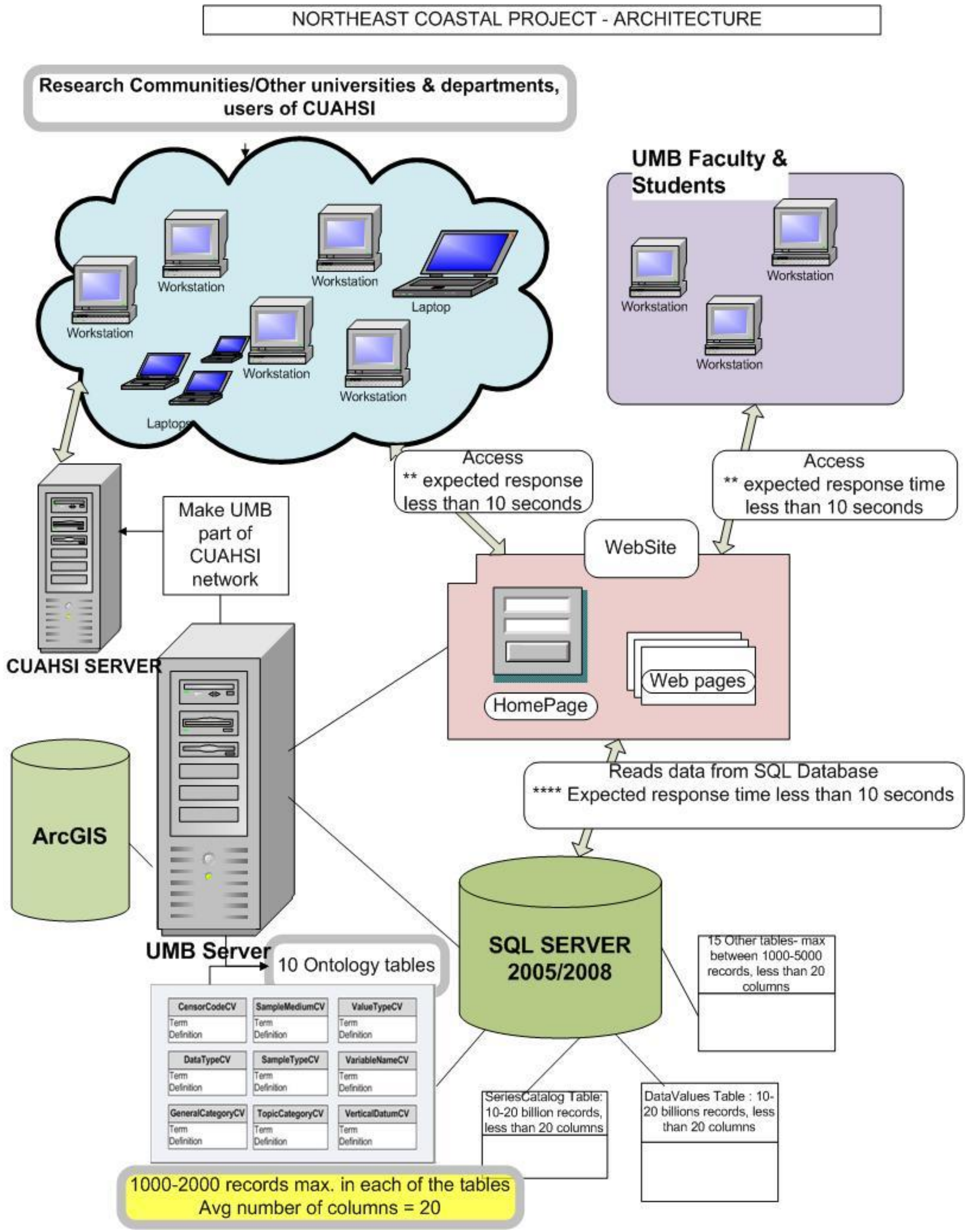


Figure 2. Hydro Server Architecture.

IMPLEMENTATION OF ARCGIS GEO-DATABASE

For integration of geospatial data we obtained a copy of MassGIS data (Oracle-based ArcSDE enterprise database; courtesy of Michael Trust). We also integrated an Open Source approach to serving data over the Web. MassGIS has an application called OLIVER which allows anyone to view nearly all MassGIS data over the Web, free of charge, with the capability to download data to one's local PC. OLIVER is accessible within a standard Web browser.

ECOLOGICAL-HUMAN-ENVIRONMENTAL DATA ASSESSMENT AND METADATA COMPILATION

Coastal ecosystems are responsive to both land-based changes to watersheds as well as marine physical and chemical processes. Therefore, coastal ecosystem data must include watershed data and marine data. In collaboration with the Massachusetts Water Resources Authority, Massachusetts Department of Environmental Protection, and Gulf of Maine Ocean Observatory System we began evaluation of critical coastal ecosystem data with a focus on climate-ecosystem interactions as we hope to develop models that use the NECWGDN to predict impacts on coastal ecosystems. Through this lens we identified key physical and chemical data that are present in existing data sets across the land-sea interface. We also established a moving window time frame for all data such that NECWGDN would include 30 years of data. We integrated the following data into the test bed database from a variety of federal, local, and published sources: bacterial count, tide levels, river discharge, water and air temperature, salinity, nutrients, fish counts, contaminants, dissolved oxygen, and dissolved carbon. These data were present for the requisite 30 -year window in data from coastal watersheds and coastal-marine environments within the Gulf of Maine and were geospatially attributable with appropriate metadata.

In order to develop an integrated data framework that would offer capacity for data distribution, storage and modeling we developed a boundary framework to help focus indicator identification and data acquisition. The approach provided deliberate focus beyond the general geospatial boundaries of the watershed and ensured a cohesive structure in which to characterize test bed progress. The Driver-Pressure-State-Impact-Exposure-Effect-Action (DPSEEA) framework (Figure 3) has been developed by the World Health Organization for integrated assessments of health problems and policy responses to them. We have identified five major categories of marine-sourced risk that may pose a hazard to persons interacting with the marine environment or its products: **human enteric bacteria, human enteric viruses, indigenous environmental bacteria, natural toxins, and anthropogenic pollutants**. These categories have long been recognized as representing the complex relationship of environmental information within the social- environmental-human health-management system.

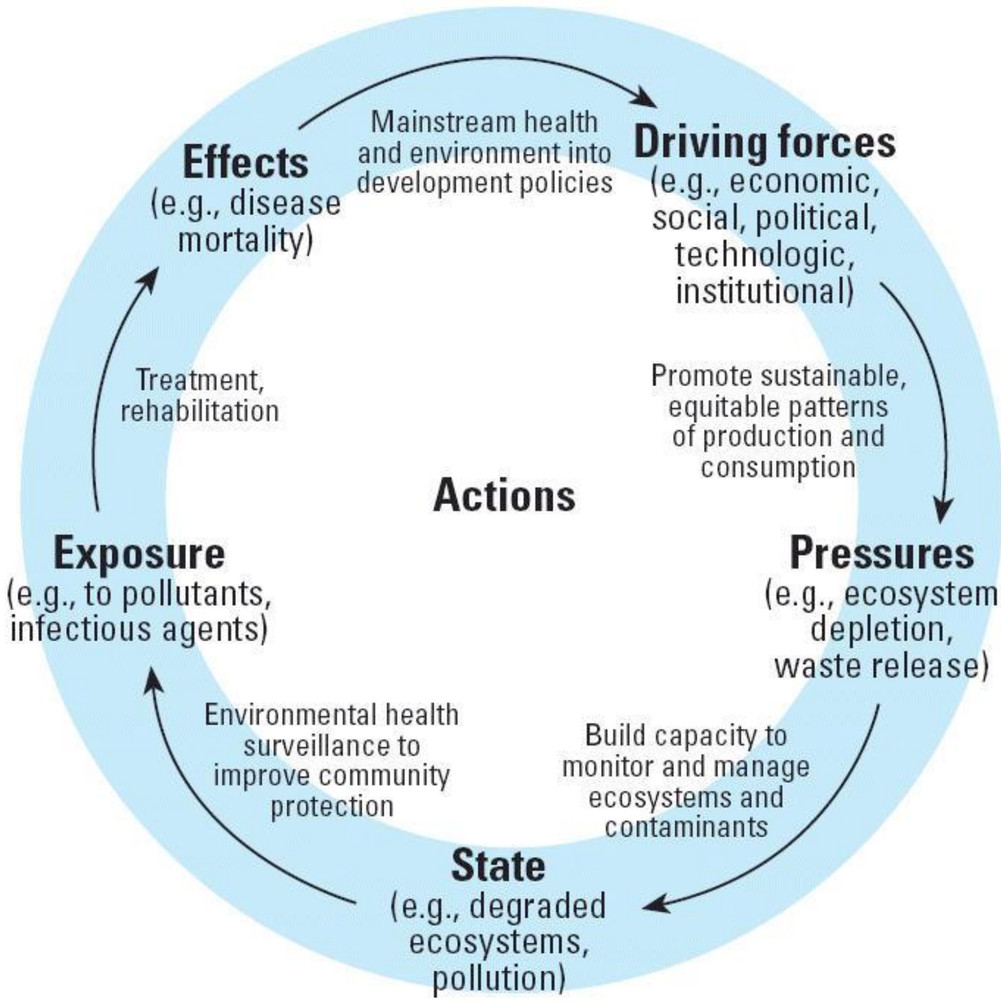


Figure 3: An example of the DPSEEA framework

We have identified a representative pathogen/toxin from each of the risk areas: Fecal coliform bacteria, Hepatitis A Virus, *Vibrio parahaemolyticus*, Domoic Acid, and therapeutic antibiotics. We reveal a list of potential environmental (and early, draft socio-economic) 'determinants' to assess the increase or decrease of risk probability for each category. **Information on these determinants, in absence of critical epidemiological data, can yield insights into changing risk probabilities and human security.** These determinants were ranked using a analytical approach related to previous work by Bowen but specifically refine for the present project. We believe this methodology demonstrates an important, and adaptable, application of a data-integrating framework that relates to current environment and human health concerns.

Summaries of Each 5 Human Health Risks

Indigenous bacteria: ***Vibrio parahaemolyticus*** *(V.p)* : Cosmopolitan bacteria found in estuarine and marine waters , it is a leading cause of human gastroenteritis associated with seafood consumption in the US and an important seafood-borne pathogen worldwide.[1] Symptoms of *V.p*. infection include gastroenteritis and vomiting, although rare, severe cases can be fatal. Many localities exhibit a season cycle of *V.p.* risk due to the link between outbreaks and warmer seawater temperatures.[2] Recently, outbreaks of infections usually associated with seafood harvested from warm waters were reported from more northerly areas of the USA that had not previously reported outbreaks. including Alaska.[1]

Enteric virus: Hepatitis A Virus (HAV): HAV replicates in human liver cells, causing an inflammation that can interfere with liver functioning. Symptoms of infection range from gastrointestinal problems to jaundice. However, many infected children are asymptomatic. This contributes to the spread of HAV via the fecal-oral route through direct contact or consumption of contaminated food or water.[3] Although a vaccine exists, in Massachusetts the vaccination rate for young children is below 50%.[4] Although there is no known animal reservoir, HAV can survive in an infectious form outside the body for weeks, and in groundwater for at least 3 months.[5] There are concerns that shellfish harvest areas meeting other approval criteria could be contaminated with HAV.[6]

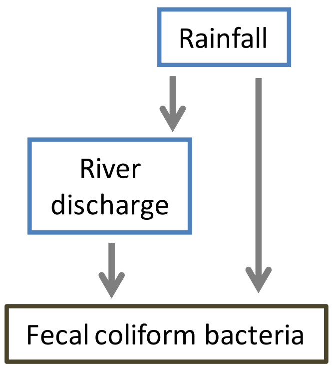
Enteric bacteria: Fecal coliforms: Fecal coliforms are a group of bacteria that live in the intestinal tract of humans and other mammals. These bacteria are shed by the millions in fecal material, and if ingested or inhaled can cause mild to severe illness.[7] However, within this study it is used because of the level and nature of accepted monitoring protocols which use fecal coliforms as indicators of all pathogenic bacteria. Every year, bathing in coastal waters polluted with fecal contamination is estimated to cause more than 120 million cases of gastrointestinal illness and 50 million cases of respiratory disease around the world.[8] Additionally, beaches may harbor fecal coliforms, and 'sand contact activities' have been positively associated with enteric illness.[10]

Natural toxin: Domoic Acid (DA): Produced by diatoms of the *Pseudo-nitzschia* genus, DA has been found in waters both temperate and tropical. An excitotoxin that causes amnesiac shellfish poisoning; symptoms of DA poisoning include nerve damage, memory loss, and GI upset. DA has been linked to intoxication and mortality events in sea birds and sea lions.[11] Blooms of *Pseudo-nitzschia* have been linked to anthropogenic runoff rich in phosphates, nitrates, silicates,[12] and copper.[13] Although DA biotransfers it does not bioaccumulate, so evolving risk factors for human exposure depend on the changing availability and consumption of potentially contaminated seafood.

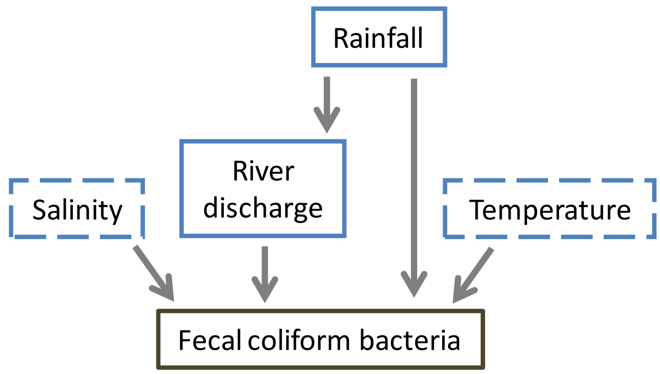
Anthropogenic Pollutant: Therapeutic antibiotics: There are about 250 different antibiotics registered for use in human and veterinary medicine, but usage patterns vary widely between countries. On a global level, the ß-lactam antibiotics, which includes penicillins and related subgroups, make up the largest share of human use antibiotics, accounting for ~50–70% of total antibiotic use.[14] An increase in resistant marine-borne pathogens is one outcome of concern from excessive antibiotic release. An area of higher uncertainty is the effect of antiobiotic release on the marine food web. For example, Cyanobacteria, which account for more than 70% of the total phytoplankton mass are considered sensitive to antimicrobial agents.[15]

We focused, initially, on the identification and ranking of ~30 environmental indicator indentified in the literature as most directly contributing to an ability to assess changes in human health. Data collection is underway as is a refinement of a protocol to better establish socio-economic indicators.

Human Health is one focus of application for the database and serves as a lens through which we focus development of the NECWGDN as a research enabling tool. In addition to human data we also focused on identification and acquisition of data that enables evaluation and prediction impacts of climate change (e.g., sea level rise, ocean freshening) on coastal ecosystem. We established a model that would be run, within NECWGDN, to evaluate the relative influence of primary environmental indicators on bacteria (Figure 4). These models also incorporated model output data from Dr. Meng Zhou and Dr. Mingshun Jiang's long-term outfall predictive model.



Hypothesized paths between primary environmental indicators of fecal coliform bacteria counts.



Hypothesized paths between primary (solid boxes) and secondary (broken boxes) environmental indicators of fecal coliform bacteria counts.

Figure 4. Model of interactions between a "model" pathogen (fecal coliform) and environmental indicators. Using data incorporated into NECWGDN it is possible to evaluate, at a specific location, the potential for risk of high fecal coliform using real-time observatory data.

FUTURE PLANS

Two defining decisions were made early in Year One of the project. One was to establish a geographic boundary condition (the Massachusetts Bays) while the other was to establish a boundary condition on the kind of indicator that would form the focus of Year One effort – i.e., those indicators that contributed most directly to an understanding of human health.

Over the coming year we will:

1. Expand the geographic boundary condition of the project either north into Maine or south (e.g., to the south reach of Long Island Sound), but retain the human health focus.

2. We ranked environmental indicators according to the degree to which they contributed to an understanding of the various selected attributes of human health. This ranking was conducted using a version of the indicator ranking protocol developed by the Boston Indicator Group (a School for the Environment Research Cluster). We will expand efforts to determine the availability of these environmental indicators.

3. A full view of the association between environmental conditions and human health need incorporate the relationship of socio economic functions as well. Using the same general approach and ranking protocol will be used to identify socio economic indicators most relevant to understanding the influence on and management options for question relating to the coastal environment and human health.

4. We will continue and refine sourcing best available data on the most important (ranked) environmental and socio economic indicators. This effort will, of necessity, include both traditional and non-traditional data sourcing. The complex and critical question of how human health and environmental condition are linked is one that requires information originally collected for myriad reasons (e.g., human population density (census data), wastewater outfall location (wastewater management), bathing beach attendance (recreational value and municipal hiring practices), percent impermeable surface . . .) not specifically originally directed at human health mitigation. This is also true for most of the complex, interdependent questions faced by environmental community. Human health provides, therefore, a useful and important illustration of management question that require information from a broad and diverse set of attributes originally collected for other purposes.

5. With data sourcing complete the next questions related to data quality, spatial compatibility and overall comparability. Data collected for and serving value to one analytical purpose may well far short if applied to other informational needs. This quality assessment will be an essential component of our future work.

6. Collaborate with Computer Science to ensure that the data management system represents the most consistent and user-accessible data portal available.

7. Future efforts will also include, to the degree possible, the building of visualization tools to better convey the nuance and trends of the data to the broadest possible audience.

Appendix B contains specific information regarding indicator approaches used.

APPENDIX A - DEVELOPER MANUAL

Uploading Observations data

Installing OD Blank Schema on SQL Server

This is one time activity that needs to be done only if you have installed a fresh SQL Server version and the OD database is not available either in \db Backup folder or there is some problem copying the OD database from \db Backup folder to SQL Server. Otherwise, go to section 5.2 .The instructions below are downloaded from http://his.cuahsi.org/documents/GettingStartedWithODM.pdf

Getting Started with ODM

Instructions for Attaching the ODM Blank Schema Database to an Instance of Microsoft SQL Server 2005

Jeffery S. Horsburgh1

5-3-2008

Introduction

This document describes how to attach the ODM blank SQL Server schema database to your instance of SQL Server so that you can get started with using ODM. In order to do so, you must be running a version of Microsoft SQL Server 2005 (i.e., Express, Standard, or Enterprise). If you do not already have an instance of Microsoft SQL Server running, you can download and install Microsoft SQL Server 2005 Express from Microsoft for free. It is recommended that you download and install both SQL Server 2005 Express and SQL Server Management Studio Express. You can get both of these products in a single installation (download and install the SQL Server 2005 Express Edition with Advanced Services SP2) as well as installation instructions at the following Microsoft URL:

http://www.microsoft.com/sql/editions/express/default.mspx

When you install SQL Server 2005 Express, it is recommended that you enable mixed mode authentication (both SQL Server authentication and Windows authentication). This will allow you to work with the ODM Tools OMD Data Loader, and ODM Streaming Data Loader applications that have been developed for ODM. These applications rely on SQL Server authentication to connect to ODM databases, and SQL Server authentication is only enabled when you choose the mixed mode authentication during installation. When you enable

Attaching the Blank ODM Schema Databases to SQL Server

The following are the steps required to attach the blank ODM schema database to an instance of Microsoft SQL Server. These steps were written using SQL Server Management Studio Express; however, the steps are similar regardless of which version of the Microsoft SQL Server Management Studio you are using.

1. Extract the blank schema database and its log file from the zip file to a location on your hard drive using WinZip or some other equivalent software. It is suggested that you extract your database to the default SQL Server data folder, which is located at the following location on disk: C:\Program Files\Microsoft SQL Server\MSSQL.1\MSSQL\Data\

NOTE: You can extract your blank schema databases 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.

2. Open the Microsoft SQL Server Management Studio Express from the Start Menu by clicking on Start --- All Programs --- Microsoft SQL Server 2005 --- SQL Server Management Studio Express.

NOTE: The path to your SQL Server Management Studio shortcut in the Start menu may be different depending on which version of SQL Server you have installed and where you chose to put the shortcut in the Start Menu.

3. It is assumed that you are connecting to your local instance of SQL Server. You should see your computer's name followed by "\SQLEXPRESS" in the "Server Name" drop down. In the following figure, the computer's name is "NOOKTEST." Change the Authentication dropdown to "SQL Server Authentication," enter "sa" for your login, and then enter your

4. Click on the "Connect" button. This will connect the Management Studio to your local SQL Server instance. Your Management Studio window should look similar to the following.

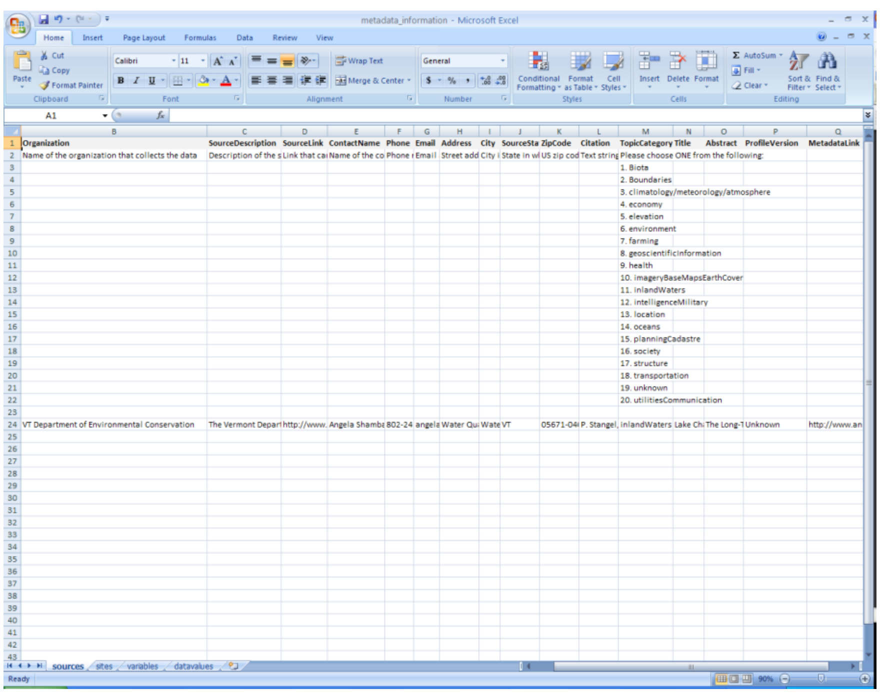
5. Right click on the "Databases" item under your server in the Object Explorer

Preparing and uploading meta-information

Before uploading observations data, meta-information needs to be created in SQL Server. Please read the document provided on the link http://his.cuahsi.org/documents/ODM1.1DesignSpecifications.pdf , before heading any further sections. This document explains the ODM schema,table structures etc.

If you have just installed SQL Server, and OD database does not exists in the Server (and not able to configure OD.bak in SQL server as OD database then follow instructions in section 5.1 before reading this section or any further sections.

To prepare and upload meta-information, we need to prepare four .csv sheets as shown below



Sources

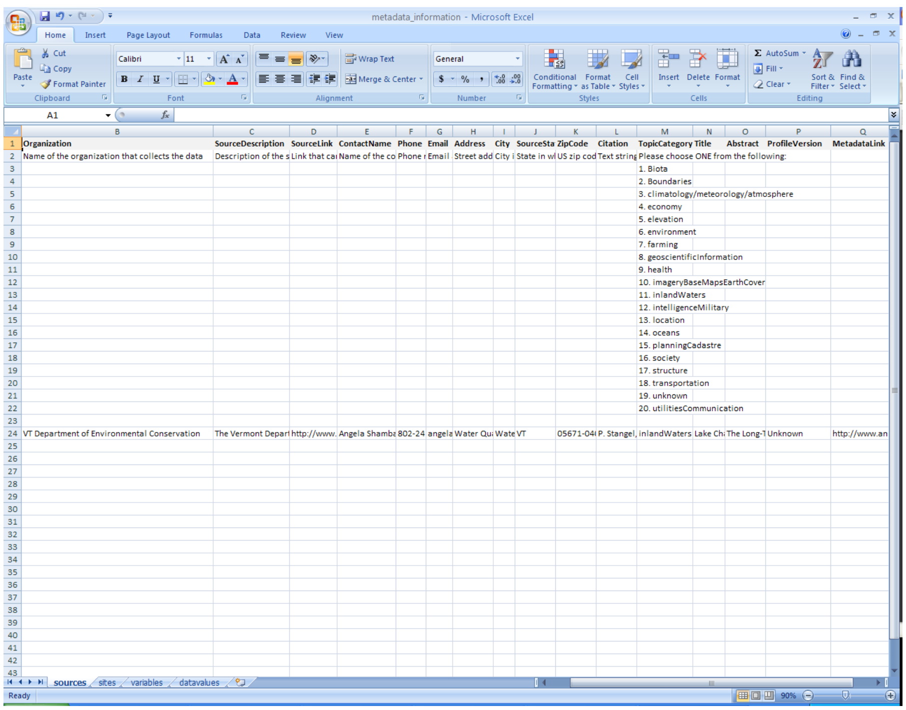
Sites

Variables

Datavalues

Let’s look into each file in detail.

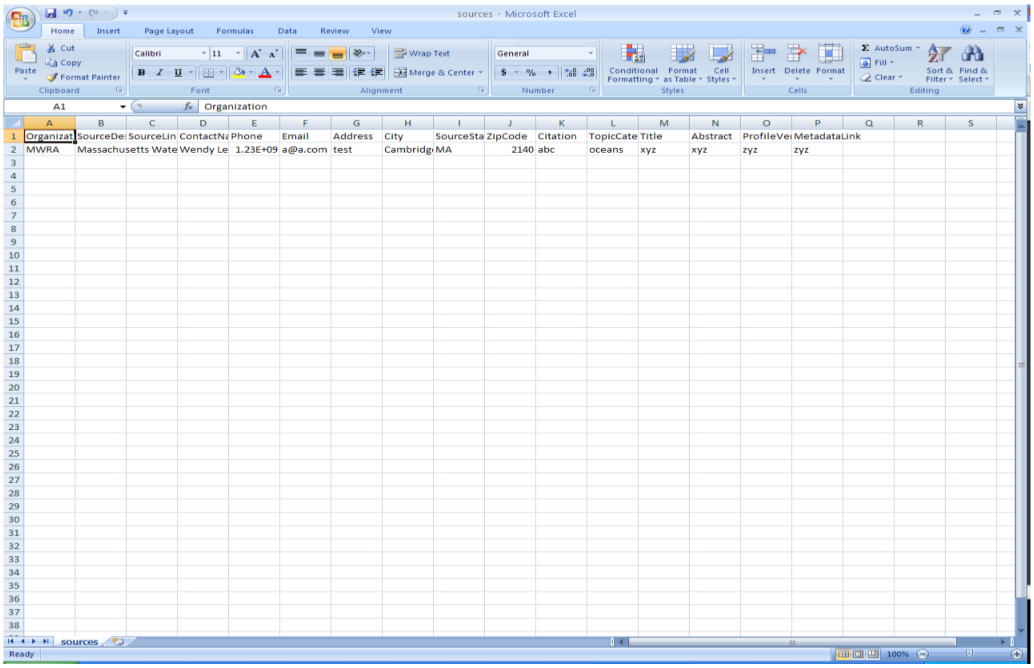
Preparing and uploading Sources meta-information



As shown in figure above, you need to fill the information about the source from where the data is obtained for example – MWRA, EPA, USGS or any other agency. Currently, for phase I of the project we are uploading only MWRA data. If, in future any more data sources are available then those also needs to be filled in here.

The information like contact person, address of the agency etc. can be requested from domain expert(Mingshun)

For more details please read page 13 on KDL Server(Pradnya's profile) in \Templates FROM Yoori Choi\Irvine-HIS-Workshop_final.pdf

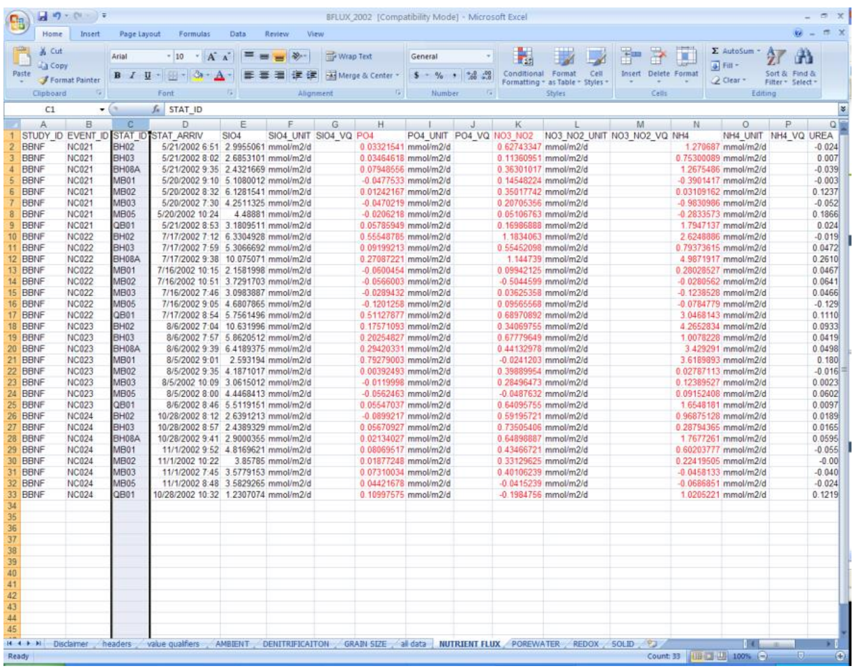


The sources.csv file is located at F:\Pradnya\EEOS\META INFORMATION\sources.csv

Preparing and uploading Sites meta-information

For creating sites information, I went through all files present in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS\CD dATA\MWRA 1998-2005.

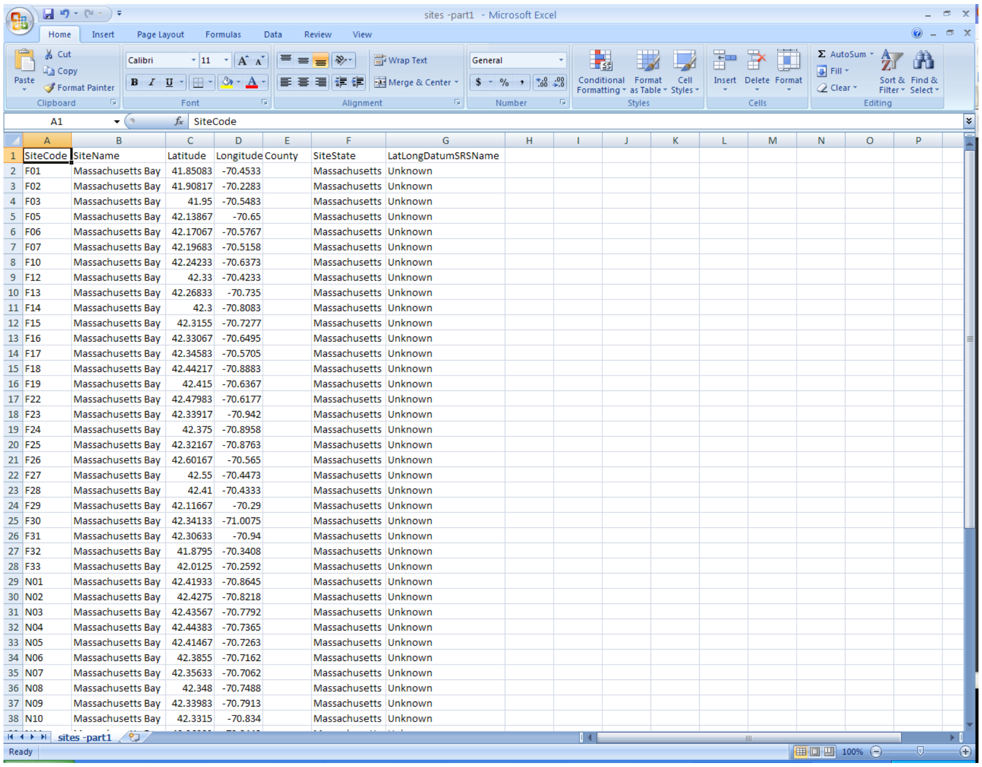
In each file, performed unique function on the column which provided site information (see sheets header section to see which column stores the site column name)



Then copy that site column and its corresponding latitude, longitude information in sites.csv file.

Repeat steps 1 and 2 for every file in CD Data folder.

Provide appropriate site name, site state, county and other required information after consulting with domain expert.



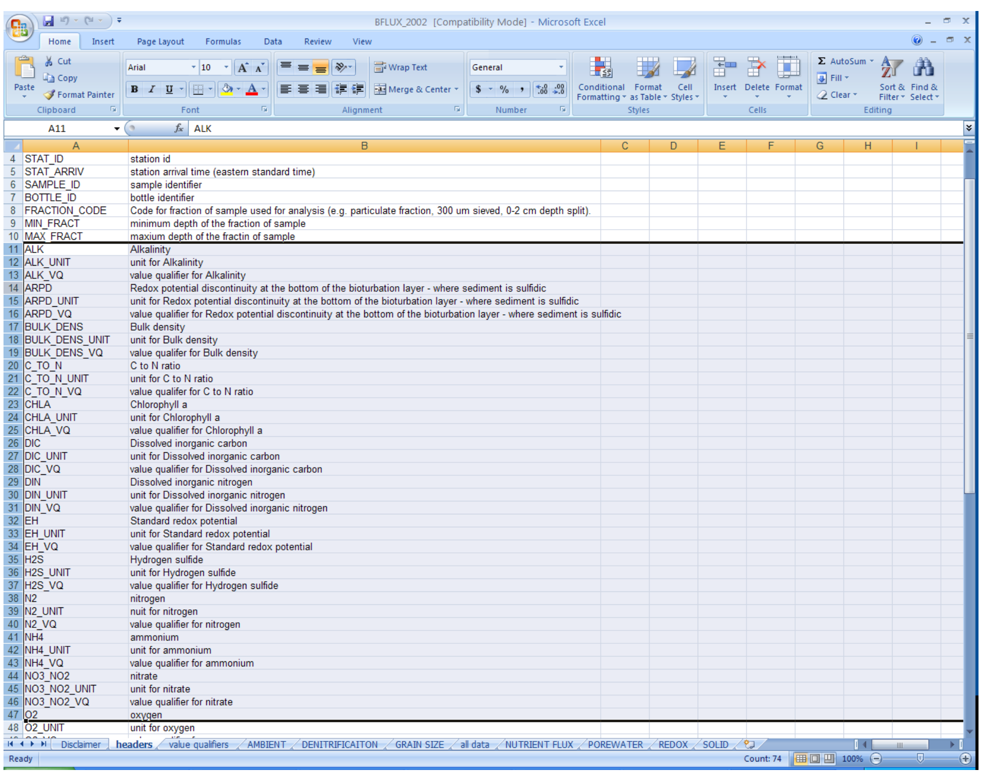
SiteCode	SiteName	Latitude	Longitude	County	SiteState	LatLongDatumSRSName
F01	Massachusetts Bay	41.85083	-70.4533		Massachusetts	Unknown
F02	Massachusetts Bay	41.90817	-70.2283		Massachusetts	Unknown
F03	Massachusetts Bay	41.95	-70.5483		Massachusetts	Unknown
F05	Massachusetts Bay	42.13867	-70.65		Massachusetts	Unknown
F06	Massachusetts Bay	42.17067	-70.5767		Massachusetts	Unknown
F07	Massachusetts Bay	42.19683	-70.5158		Massachusetts	Unknown
F10	Massachusetts Bay	42.24233	-70.6373		Massachusetts	Unknown
F12	Massachusetts Bay	42.33	-70.4233		Massachusetts	Unknown
F13	Massachusetts Bay	42.26833	-70.735		Massachusetts	Unknown
F14	Massachusetts Bay	42.3	-70.8083		Massachusetts	Unknown
F15	Massachusetts Bay	42.3155	-70.7277		Massachusetts	Unknown
F16	Massachusetts Bay	42.33067	-70.6495		Massachusetts	Unknown
F17	Massachusetts Bay	42.34583	-70.5705		Massachusetts	Unknown
F18	Massachusetts Bay	42.44217	-70.8883		Massachusetts	Unknown
F19	Massachusetts Bay	42.415	-70.6367		Massachusetts	Unknown
F22	Massachusetts Bay	42.47983	-70.6177		Massachusetts	Unknown
F23	Massachusetts Bay	42.33917	-70.942		Massachusetts	Unknown
F24	Massachusetts Bay	42.375	-70.8958		Massachusetts	Unknown
F25	Massachusetts Bay	42.32167	-70.8763		Massachusetts	Unknown
F26	Massachusetts Bay	42.60167	-70.565		Massachusetts	Unknown
F27	Massachusetts Bay	42.55	-70.4473		Massachusetts	Unknown
F28	Massachusetts Bay	42.41	-70.4333		Massachusetts	Unknown
F29	Massachusetts Bay	42.11667	-70.29		Massachusetts	Unknown
F30	Massachusetts Bay	42.34133	-71.0075		Massachusetts	Unknown
F31	Massachusetts Bay	42.30633	-70.94		Massachusetts	Unknown
F32	Massachusetts Bay	41.8795	-70.3408		Massachusetts	Unknown
F33	Massachusetts Bay	42.0125	-70.2592		Massachusetts	Unknown
N01	Massachusetts Bay	42.41933	-70.8645		Massachusetts	Unknown
N02	Massachusetts Bay	42.4275	-70.8218		Massachusetts	Unknown
N03	Massachusetts Bay	42.43567	-70.7792		Massachusetts	Unknown
N04	Massachusetts Bay	42.44383	-70.7365		Massachusetts	Unknown
N05	Massachusetts Bay	42.41467	-70.7263		Massachusetts	Unknown
N06	Massachusetts Bay	42.3855	-70.7162		Massachusetts	Unknown
N07	Massachusetts Bay	42.35633	-70.7062		Massachusetts	Unknown
N08	Massachusetts Bay	42.348	-70.7488		Massachusetts	Unknown
N09	Massachusetts Bay	42.33983	-70.7913		Massachusetts	Unknown
N10	Massachusetts Bay	42.3315	-70.834		Massachusetts	Unknown

Use ODM Data loader tool (ODM_DL1.1.3) to upload sites.csv file (See Section 8.2 use sites.csv in place of variables.csv)

Preparing and uploading Variables meta-information

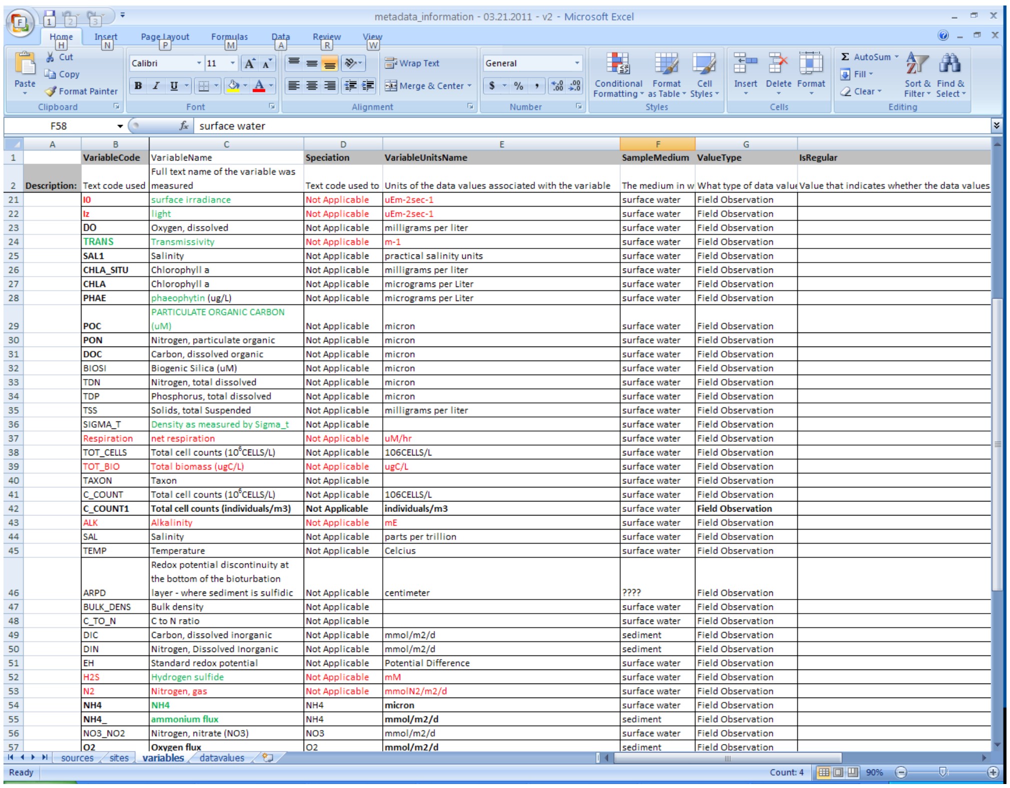
To prepare variable information, go through every file, and every sheet in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \CD FOLDER (ALL OBSERVATIONS DATA)

Mostly, the headers sheet will give all variables that are measured (leave out, sites, event_id, SAMPLE_ID etc. field. If you have doubt consult with domain expert. Also, see the units used. For some variables like oxygen, it has two different units. So create two separate variables for oxygen.



	A	B
4	STAT_ID	station id
5	STAT_ARRIV	station arrival time (eastern standard time)
6	SAMPLE_ID	sample identifier
7	BOTTLE_ID	bottle identifier
8	FRACTION_CODE	Code for fraction of sample used for analysis (e.g. particulate fraction, 300 um sieved, 0-2 cm depth split).
9	MIN_FRACT	minimum depth of the fraction of sample
10	MAX_FRACT	maxium depth of the fractin of sample
11	ALK	Alkalinity
12	ALK_UNIT	unit for Alkalinity
13	ALK_VQ	value qualifier for Alkalinity
14	ARPD	Redox potential discontinuity at the bottom of the bioturbation layer - where sediment is sulfidic
15	ARPD_UNIT	unit for Redox potential discontinuity at the bottom of the bioturbation layer - where sediment is sulfidic
16	ARPD_VQ	value qualifier for Redox potential discontinuity at the bottom of the bioturbation layer - where sediment is sulfidic
17	BULK_DENS	Bulk density
18	BULK_DENS_UNIT	unit for Bulk density
19	BULK_DENS_VQ	value qualifer for Bulk density
20	C_TO_N	C to N ratio
21	C_TO_N_UNIT	unit for C to N ratio
22	C_TO_N_VQ	value qualifer for C to N ratio
23	CHLA	Chlorophyll a
24	CHLA_UNIT	unit for Chlorophyll a
25	CHLA_VQ	value qualifier for Chlorophyll a
26	DIC	Dissolved inorganic carbon
27	DIC_UNIT	unit for Dissolved inorganic carbon
28	DIC_VQ	value qualifier for Dissolved inorganic carbon
29	DIN	Dissolved inorganic nitrogen
30	DIN_UNIT	unit for Dissolved inorganic nitrogen
31	DIN_VQ	value qualifier for Dissolved inorganic nitrogen
32	EH	Standard redox potential
33	EH_UNIT	unit for Standard redox potential
34	EH_VQ	value qualifier for Standard redox potential
35	H2S	Hydrogen sulfide
36	H2S_UNIT	unit for Hydrogen sulfide
37	H2S_VQ	value qualifier for Hydrogen sulfide
38	N2	nitrogen
39	N2_UNIT	nuit for nitrogen
40	N2_VQ	value qualifier for nitrogen
41	NH4	ammonium
42	NH4_UNIT	unit for ammonium
43	NH4_VQ	value qualifier for ammonium
44	NO3_NO2	nitrate
45	NO3_NO2_UNIT	unit for nitrate
46	NO3_NO2_VQ	value qualifier for nitrate
47	O2	oxygen
48	O2_UNIT	unit for oxygen

Prepare a sheet, containing all variables found in observations data. As shown below



	A	B	C	D	E	F	G	
1		VariableCode	VariableName	Speciation	VariableUnitsName	SampleMedium	ValueType	IsRegular
2	Description:	Text code used	Full text name of the variable was measured	Text code used to	Units of the data values associated with the variable	The medium in w	What type of data valu	Value that indicates whether the data values
21		I0	surface irradiance	Not Applicable	uEm-2sec-1	surface water	Field Observation	
22		Iz	light	Not Applicable	uEm-2sec-1	surface water	Field Observation	
23		DO	Oxygen, dissolved	Not Applicable	milligrams per liter	surface water	Field Observation	
24		TRANS	Transmissivity	Not Applicable	m-1	surface water	Field Observation	
25		SAL1	Salinity	Not Applicable	practical salinity units	surface water	Field Observation	
26		CHLA_SITU	Chlorophyll a	Not Applicable	milligrams per liter	surface water	Field Observation	
27		CHLA	Chlorophyll a	Not Applicable	micrograms per Liter	surface water	Field Observation	
28		PHAE	phaeophytin (ug/L)	Not Applicable	micrograms per Liter	surface water	Field Observation	
29		POC	PARTICULATE ORGANIC CARBON (uM)	Not Applicable	micron	surface water	Field Observation	
30		PON	Nitrogen, particulate organic	Not Applicable	micron	surface water	Field Observation	
31		DOC	Carbon, dissolved organic	Not Applicable	micron	surface water	Field Observation	
32		BIOSI	Biogenic Silica (uM)	Not Applicable	micron	surface water	Field Observation	
33		TDN	Nitrogen, total dissolved	Not Applicable	micron	surface water	Field Observation	
34		TDP	Phosphorus, total dissolved	Not Applicable	micron	surface water	Field Observation	
35		TSS	Solids, total Suspended	Not Applicable	milligrams per liter	surface water	Field Observation	
36		SIGMA_T	Density as measured by Sigma_t	Not Applicable		surface water	Field Observation	
37		Respiration	net respiration	Not Applicable	uM/hr	surface water	Field Observation	
38		TOT_CELLS	Total cell counts (10^6CELLS/L)	Not Applicable	106CELLS/L	surface water	Field Observation	
39		TOT_BIO	Total biomass (ugC/L)	Not Applicable	ugC/L	surface water	Field Observation	
40		TAXON	Taxon	Not Applicable		surface water	Field Observation	
41		C_COUNT	Total cell counts (10^6CELLS/L)	Not Applicable	106CELLS/L	surface water	Field Observation	
42		C_COUNT1	Total cell counts (individuals/m3)	Not Applicable	individuals/m3	surface water	Field Observation	
43		ALK	Alkalinity	Not Applicable	mE	surface water	Field Observation	
44		SAL	Salinity	Not Applicable	parts per trillion	surface water	Field Observation	
45		TEMP	Temperature	Not Applicable	Celcius	surface water	Field Observation	
46		ARPD	Redox potential discontinuity at the bottom of the bioturbation layer - where sediment is sulfidic	Not Applicable	centimeter	????	Field Observation	
47		BULK_DENS	Bulk density	Not Applicable		surface water	Field Observation	
48		C_TO_N	C to N ratio	Not Applicable		surface water	Field Observation	
49		DIC	Carbon, dissolved inorganic	Not Applicable	mmol/m2/d	sediment	Field Observation	
50		DIN	Nitrogen, Dissolved Inorganic	Not Applicable	mmol/m2/d	sediment	Field Observation	
51		EH	Standard redox potential	Not Applicable	Potential Difference	surface water	Field Observation	
52		H2S	Hydrogen sulfide	Not Applicable	mM	surface water	Field Observation	
53		N2	Nitrogen, gas	Not Applicable	mmolN2/m2/d	surface water	Field Observation	
54		NH4	NH4	NH4	micron	surface water	Field Observation	
55		NH4_	ammonium flux	NH4	mmol/m2/d	sediment	Field Observation	
56		NO3_NO2	Nitrogen, nitrate (NO3)	NO3	mmol/m2/d	surface water	Field Observation	
57		O2	Oxygen flux	O2	mmol/m2/d	sediment	Field Observation	

The above sheet is located at \ META INFORMATION\ metadata_information - 03.21.2011 - v2 , (or also can refer v3 or v4 version of it) see the sites sheet

Fill in the variable code (anything that is easy to read for developer)

Fill Speciation, if you know or leave blank for domain expert to fill it.

Send this sheet to domain expert and confirm what all variables need to be created, and request him to specify values for sample medium etc all remaining fields (except VariableName, VariablesUnitsName).

After you confirmed variables that needs to be created from domain expert and you have information about the remaining fields go to next step. Follow steps below to fill VariablesName and VariableUnitsName field.

To fill VariableName, follow steps below

Start SQL Server. Open OD Database -> Tables -> on VariablesCV table

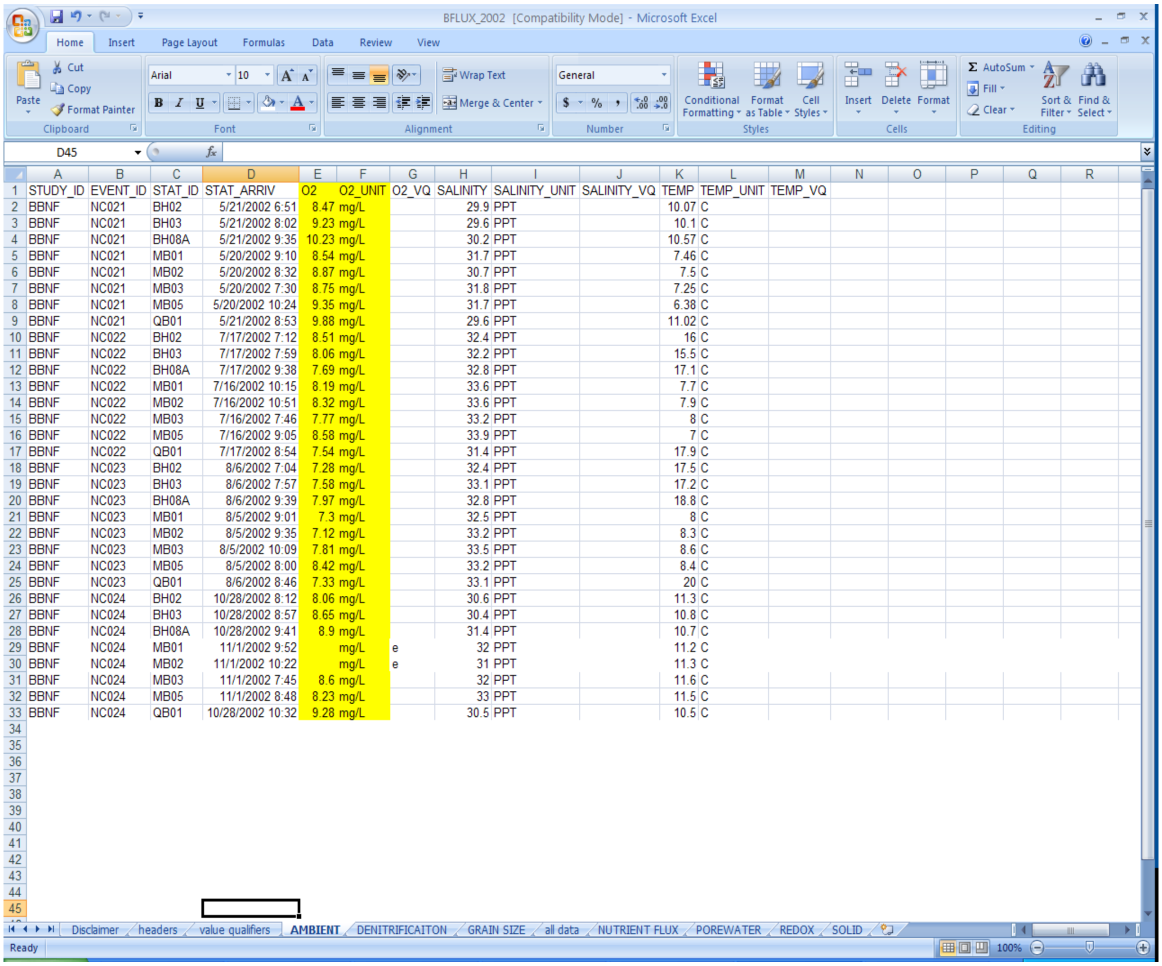
Run similar query to look for variable in VariablesCV

Or you can also search for the variable on http://his.cuahsi.org/mastercvreg/edit_cv11.aspx?tbl=VariableNameCV&id=821577965

If you find the variable in controlled vocabulary then copy paste it and go to step 8

Else send request to CUAHSI to create such variable in controlled vocabulary. Follow section 8.6 and then go to next step.

To fill, the VariablesUnitName field, see the unit used in the observations data sheet(CD Folder and refer to the file and sheet name(or header sheet) for which you are creating variable) Lets say we have Oxygen variable and its unit is mg/L i.e milligrams per Litre (if you do not know units's long form ask domain expert)



STUDY_ID	EVENT_ID	STAT_ID	STAT_ARRIV	O2	O2_UNIT	O2_VQ	SALINITY	SALINITY_UNIT	SALINITY_VQ	TEMP	TEMP_UNIT	TEMP_VQ
BBNF	NC021	BH02	5/21/2002 6:51	8.47	mg/L		29.9	PPT		10.07	C	
BBNF	NC021	BH03	5/21/2002 8:02	9.23	mg/L		29.6	PPT		10.1	C	
BBNF	NC021	BH08A	5/21/2002 9:35	10.23	mg/L		30.2	PPT		10.57	C	
BBNF	NC021	MB01	5/20/2002 9:10	8.54	mg/L		31.7	PPT		7.46	C	
BBNF	NC021	MB02	5/20/2002 8:32	8.87	mg/L		30.7	PPT		7.5	C	
BBNF	NC021	MB03	5/20/2002 7:30	8.75	mg/L		31.8	PPT		7.25	C	
BBNF	NC021	MB05	5/20/2002 10:24	9.35	mg/L		31.7	PPT		6.38	C	
BBNF	NC021	QB01	5/21/2002 8:53	9.88	mg/L		29.6	PPT		11.02	C	
BBNF	NC022	BH02	7/17/2002 7:12	8.51	mg/L		32.4	PPT		16	C	
BBNF	NC022	BH03	7/17/2002 7:59	8.06	mg/L		32.2	PPT		15.5	C	
BBNF	NC022	BH08A	7/17/2002 9:38	7.69	mg/L		32.8	PPT		17.1	C	
BBNF	NC022	MB01	7/16/2002 10:15	8.19	mg/L		33.6	PPT		7.7	C	
BBNF	NC022	MB02	7/16/2002 10:51	8.32	mg/L		33.6	PPT		7.9	C	
BBNF	NC022	MB03	7/16/2002 7:46	7.77	mg/L		33.2	PPT		8	C	
BBNF	NC022	MB05	7/16/2002 9:05	8.58	mg/L		33.9	PPT		7	C	
BBNF	NC022	QB01	7/17/2002 8:54	7.54	mg/L		31.4	PPT		17.9	C	
BBNF	NC023	BH02	8/6/2002 7:04	7.28	mg/L		32.4	PPT		17.5	C	
BBNF	NC023	BH03	8/6/2002 7:57	7.58	mg/L		33.1	PPT		17.2	C	
BBNF	NC023	BH08A	8/6/2002 9:39	7.97	mg/L		32.8	PPT		18.8	C	
BBNF	NC023	MB01	8/5/2002 9:01	7.3	mg/L		32.5	PPT		8	C	
BBNF	NC023	MB02	8/5/2002 9:35	7.12	mg/L		33.2	PPT		8.3	C	
BBNF	NC023	MB03	8/5/2002 10:09	7.81	mg/L		33.5	PPT		8.6	C	
BBNF	NC023	MB05	8/5/2002 8:00	8.42	mg/L		33.2	PPT		8.4	C	
BBNF	NC023	QB01	8/6/2002 8:46	7.33	mg/L		33.1	PPT		20	C	
BBNF	NC024	BH02	10/28/2002 8:12	8.06	mg/L		30.6	PPT		11.3	C	
BBNF	NC024	BH03	10/28/2002 8:57	8.65	mg/L		30.4	PPT		10.8	C	
BBNF	NC024	BH08A	10/28/2002 9:41	8.9	mg/L		31.4	PPT		10.7	C	
BBNF	NC024	MB01	11/1/2002 9:52		mg/L	e	32	PPT		11.2	C	
BBNF	NC024	MB02	11/1/2002 10:22		mg/L	e	31	PPT		11.3	C	
BBNF	NC024	MB03	11/1/2002 7:45	8.6	mg/L		32	PPT		11.6	C	
BBNF	NC024	MB05	11/1/2002 8:48	8.23	mg/L		33	PPT		11.5	C	
BBNF	NC024	QB01	10/28/2002 10:32	9.28	mg/L		30.5	PPT		10.5	C	

Then start SQL Server, and open OD database->tables->Units table

And run query something like

```
Select * from Units where UnitsName like
'milligram%'
```

OR

Search for it on http://his.cuahsi.org/mastercvreg/edit_cv11.aspx?tbl=Units&id=789577851

a. If you find the unit that you are looking for, copy the unit and paste it in VariableUnitsName field (and go to step 11)

b. If you did not find the unit that means it does not exist in CUAHSI Controlled Vocabulary.So you need to request it to CUAHSI to create such unit .Please go to the link
http://his.cuahsi.org/mastercvreg/cv11.aspx

c. Follow steps in section 8.6 and then proceed from here

11) Repeat steps from 3 to 6 for all the variables.

12) Follow up with CUAHSI for units or variable names creation in controlled vocabulary.

13) Follow section 8.2 to upload variables.csv to SQL Server

Preparing and uploading Observations data - DataValues meta-information

Observations data is available from 1998 to 2005 which was provided by Mingshun on CD (which is now handed over to Prof. Ding).Although data is available in excel, it's not well formatted. Also, at the beginning of the project not all variables were created in controlled vocabulary. So some sections of the code was commented and those programs were re-run to get datavalues.csv for other variables(which were commented initially)

ODM tool 1.1.3 was used to upload observations data. To upload data it needs to be in format specified by CUAHSI. See KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \ Templates FROM Yoori Choi\ metadata_information.csv .The datavalues sheet must contain the observations data.

Before heading in this section, make sure that you have completed section 5.

For phase I of the project, domain expert has provided MWRA observations data which is located in CD dATA\MWRA 1998-2005. As per the files provided I have categorized them into following

Bflux_YYYY

Hydro_fluorlight_YYYY

Hydro_nutrients_YYYY

Hydro_productivity_YYYY

Phyto_sp_biomass_YYYY

where YYYY = year

Uploading BFlux Files – DataValues MetaInformation

To upload MWRA observations data, a java program needs to be run which is located in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Java Programs FOR Observations Data folder.

Bflux files has many sheets. Domain expert asked to load

a) ALL_DATA sheet

b) Ambient sheet

c) PoreWater sheet

d) REDOX sheet

e) SOLID sheet

f) NUTRIENT_FLUX sheet

a) Uploading ALL_DATA sheet

Open program , EEOS.Util.BLUX.FileReader_BFLUX_ALLDATA.java .Change the path highlighted in blue.

```
private void readSheet(Sheet curSheet) { //
column row format


            try {

                  // Create new Excel sheet for
output

                  WorkbookSettings ws = new
WorkbookSettings();
```

```
public static void main(String[] args) {

            try {

                  FileReader_BFLUX_ALLDATA
xlReader = new FileReader_BFLUX_ALLDATA();

                  // ===========CHANGE FILE PATH,
FILE NAME ==================

                  xlReader
```

If required, uncomment lines in orange

```
//===== UNCOMMENT LINE BELOW IF TIME COLUMN PROVIDED ===

            //          Cell TIME = curSheet.getCell(5, i);


                        // trim white spaces and create date format string

                        String recordDate = MONTH.getContents().trim() + "/"
```

Make sure the input file is in the current directory.

Make sure ALL_DATA sheet is first sheet of the input file.

And now run the program, and the DataValues.csv will be created.

By sampling few records, make sure that correct data is generated. And then upload the DataValues.csv. For uploading, see section 8.3

c) Uploading PoreWater Sheet

Open program located in in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Java Programs FOR Observations Data\EEOS\Util\BFLUX\ FileReader_BFLUX_Porewater.java

Make changes in path for input and output files.

Make sure input file is in current directory.

PoreWater Sheet should be first sheet in the excel file.

Run the program, and upload Datavalues.csv. For uploading refer section 8.3.

d) Uploading REDOX sheet

Open program located in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Java Programs FOR Observations Data\EEOS\Util\BFLUX\ FileReader_BFLUX_REDOX.java

Make changes in path for input and output files.

Make sure input file is in current directory.

REDOX Sheet should be first sheet in the excel file.

Run the program, and upload Datavalues.csv. For uploading refer section 8.3

e) Uploading SOLID Sheet

Open program located in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Java Programs FOR Observations Data\EEOS\Util\BFLUX\ FileReader_BFLUX_SOLID.java

Make changes in path for input and output files.

Make sure input file is in current directory.

SOLID Sheet should be first sheet in the excel file.

Run the program, and upload Datavalues.csv. For uploading refer section 8.3

f**) Uploading BUTRIENT_FLUX sheet**

Open program located in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Java Programs FOR Observations Data\EEOS\Util\BFLUX\ FileReader_BFLUX_NUTRIENTFLUX.java

Make changes in path for input and output files.

Make sure input file is in current directory.

NUTRIENT FLUX Sheet should be first sheet in the excel file.

Run the program, and upload Datavalues.csv. For uploading refer section 8.3

Uploading Hydro_fluorlight – DataValues MetaInformation

Open program located in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Java Programs FOR Observations Data\EEOS\Util\HydroFluro\ FileReader_HydroFluroLight.java

Make changes in path for input and output files.

Make sure input file is in current directory.

Data Sheet should be first sheet in the excel file.

Run the program, and upload Datavalues.csv. For uploading refer section 8.3

Uploading Hydro_nutrients – DataValues MetaInformation

Open program located in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Java Programs FOR Observations Data\EEOS\Util\nutrients\ FileReader_HydroNutrients.java

Make changes in path for input and output files.

Make sure input file is in current directory.

Data Sheet should be first sheet in the excel file.

Run the program, and upload Datavalues.csv. For uploading refer section 8.3

Uploading Hydro_productivity –DataValues MetaInformation

Open program located in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Java Programs FOR Observations Data\EEOS\Util\hydroProductivity\ FileReader_HydroProductivity.java

Make changes in path for input and output files.

Make sure input file is in current directory.

Data Sheet should be first sheet in the excel file.

Run the program, and upload Datavalues.csv. For uploading refer section 8.3

Uploading Phyto_sp_biomass – DataValues MetaInformation

Open program located in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Java Programs FOR Observations Data\EEOS\Util\ PhytoBioMass \ FileReader_PyhtoBioMass.java

Make changes in path for input and output files.

Make sure input file is in current directory.

Data Sheet should be first sheet in the excel file.

Run the program, and upload Datavalues.csv. For uploading refer section 8.3

Uploading Model Data –Meta-information

Model data provided by domain expert is available in netcdf files. This format can be read using python programs. Please see section 9, to see which model data files have already been uploaded. Before you start uploading model data we need to verify if meta-information required for the model data is already existing in OD database.

I have already uploaded sites,variables,sources meta-information to the OD database. If you are using same machine as I did, with OD database available on SQL Server 2008, then skip this section and section 7.1 and 7.2

Preparing Sites Meta-information for Model Data

I have already uploaded sites meta-information to the OD database which is available on LD Lab machine. If you are using same machine with OD database available on SQL Server 2008, then skip this section. In case, a new machine has been assigned, then check Section 4 to upload all required software for the project.

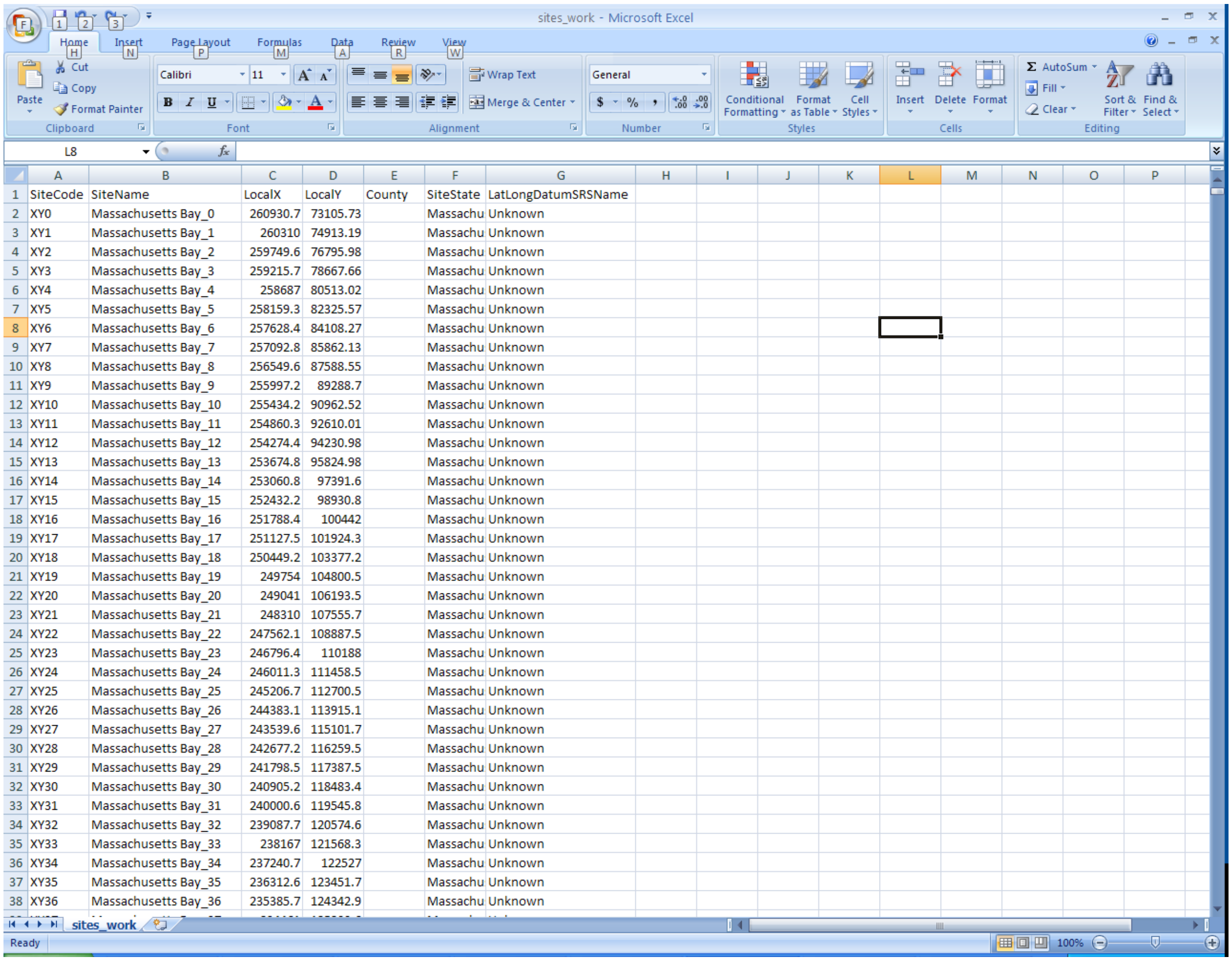
Once the SQL Server is installed and then copy the latest backup of OD database which available in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \db Backup folder. If for some reason, OD database is not configured back to SQL Server, then follow steps below to upload sites information to OD.

Make sure that ecom2005.01.cdf is in same folder as the Generate_Model_Sites.py program

Sites information is common to both HydroModel as well as Dynamic model.

Run Script Python Scripts\Generate_Model_Sites. The program reads from HydroModel Jan 2005 file to create sites information.

As shown below, sites.csv is created, but it does not have complete information. The netcdf file has x and y co-ordinates values, but sites table on OD database requires latitude and longitude information as well. So we have to use ArcMap to generate that information.



SiteCode	SiteName	LocalX	LocalY	County	SiteState	LatLongDatumSRSName
XY0	Massachusetts Bay_0	260930.7	73105.73		Massachu	Unknown
XY1	Massachusetts Bay_1	260310	74913.19		Massachu	Unknown
XY2	Massachusetts Bay_2	259749.6	76795.98		Massachu	Unknown
XY3	Massachusetts Bay_3	259215.7	78667.66		Massachu	Unknown
XY4	Massachusetts Bay_4	258687	80513.02		Massachu	Unknown
XY5	Massachusetts Bay_5	258159.3	82325.57		Massachu	Unknown
XY6	Massachusetts Bay_6	257628.4	84108.27		Massachu	Unknown
XY7	Massachusetts Bay_7	257092.8	85862.13		Massachu	Unknown
XY8	Massachusetts Bay_8	256549.6	87588.55		Massachu	Unknown
XY9	Massachusetts Bay_9	255997.2	89288.7		Massachu	Unknown
XY10	Massachusetts Bay_10	255434.2	90962.52		Massachu	Unknown
XY11	Massachusetts Bay_11	254860.3	92610.01		Massachu	Unknown
XY12	Massachusetts Bay_12	254274.4	94230.98		Massachu	Unknown
XY13	Massachusetts Bay_13	253674.8	95824.98		Massachu	Unknown
XY14	Massachusetts Bay_14	253060.8	97391.6		Massachu	Unknown
XY15	Massachusetts Bay_15	252432.2	98930.8		Massachu	Unknown
XY16	Massachusetts Bay_16	251788.4	100442		Massachu	Unknown
XY17	Massachusetts Bay_17	251127.5	101924.3		Massachu	Unknown
XY18	Massachusetts Bay_18	250449.2	103377.2		Massachu	Unknown
XY19	Massachusetts Bay_19	249754	104800.5		Massachu	Unknown
XY20	Massachusetts Bay_20	249041	106193.5		Massachu	Unknown
XY21	Massachusetts Bay_21	248310	107555.7		Massachu	Unknown
XY22	Massachusetts Bay_22	247562.1	108887.5		Massachu	Unknown
XY23	Massachusetts Bay_23	246796.4	110188		Massachu	Unknown
XY24	Massachusetts Bay_24	246011.3	111458.5		Massachu	Unknown
XY25	Massachusetts Bay_25	245206.7	112700.5		Massachu	Unknown
XY26	Massachusetts Bay_26	244383.1	113915.1		Massachu	Unknown
XY27	Massachusetts Bay_27	243539.6	115101.7		Massachu	Unknown
XY28	Massachusetts Bay_28	242677.2	116259.5		Massachu	Unknown
XY29	Massachusetts Bay_29	241798.5	117387.5		Massachu	Unknown
XY30	Massachusetts Bay_30	240905.2	118483.4		Massachu	Unknown
XY31	Massachusetts Bay_31	240000.6	119545.8		Massachu	Unknown
XY32	Massachusetts Bay_32	239087.7	120574.6		Massachu	Unknown
XY33	Massachusetts Bay_33	238167	121568.3		Massachu	Unknown
XY34	Massachusetts Bay_34	237240.7	122527		Massachu	Unknown
XY35	Massachusetts Bay_35	236312.6	123451.7		Massachu	Unknown
XY36	Massachusetts Bay_36	235385.7	124342.9		Massachu	Unknown

Insert 2 columns Latitude and Longitude in the sites.csv generated in step 3 And follow instructions provided by Timothy from CUAHSI

To generate the Latitude and Longitude columns. You will need to convert the local coordinates into latitude and longitude, and store the result in the Latitude and Longitude columns of your sites.csv file. I don't know what tools are available for doing this with netCDF files, but if I were doing it, I would probably take this approach since I have ArcGIS version 10.

1. In a new ArcMap document, using the Multidimension Tools toolbox in ArcToolbox, create a netCDF feature layer of the site locations by running Make NetCDF Feature Layer. Make sure you include x and y in the list of variables to include when running the tool.

2. Hopefully, ArcMap recognized the coordinate system used by the netCDF file. You can check by viewing the layer properties for the point layer that was created, and seeing if a coordinate system has been assigned to the layer (in the Source tab of the Layer Properties dialog). Let's assume a coordinate system was assigned.

3. Right-click the layer and click Data | Export to export the layer to a new feature class called MySites. This is so we can edit it.

4. Right-click the data frame name and click Properties. The data frame is probably named "Layers."

5. In the Coordinate System tab, select the Predefined Geographic Coordinate System for the World called "WGS 84" and click OK. This helps us calculate coordinates in latitude and longitude.

6. Open the attribute table for MySites. You should already see fields populated for x and y. Now you will add fields for Latitude and Longitude and calculate them.

7. Add Latitude and Longitude fields. Use a field type of Double.

8. Right-click the Latitude field and click Calculate Geometry.

9. In the dialog, choose "Y Coordinate of Point" and "Use coordinate system of data frame" and click OK. You should now have the Latitude field fully populated.

10. Repeat similarly for Longitude.

11. In the table window, choose to Export the table to a text file called sites.csv.

12. Continue formatting the text file to load into an ODM database.

And t

1. In ArcToolbox, run the Data Management Tools | Projections and Transformations | Define Projection tool to set the coordinate system of MySites to your Massachusetts State Plane system.

2. In the tool dialog, to set the coordinate system, click the button to the right of the second text box to open the Spatial Reference Properties dialog.

3. In the Spatial Reference Properties dialog, click Select.

4. Select the MA state plane system and click Add. It's at Projected Coordinate Systems | State Plane | NAD 1927 (US Feet) | NAD 1927 StatePlane Massachusetts FIPS 2001.prj. That's the closest one to State Plane, NAD 27, Massachusetts Mainland-2001, Meters, which you listed in your Word document. The coordinate system doesn't exactly match what you specified in your Word document, so now we'll modify it a bit.

5. When you clicked Add, it brought you back to the Spatial Reference Properties dialog. In this dialog, click Modify.

6. Change the Linear Unit to Meter by selecting it from the drop down list. Do not type "Meter." Select Meter from the list.

7. In the Geographic Coordinate System group at the bottom, click Select. This is to change the GCS from NAD27 to GRS80.

8. Select GRS 80 and click Add. It's at Geographic Coordinate Systems | Spheroid-based | GRS 1980.prj.

9. Click OK to close the Projected Coordinate System Properties dialog.

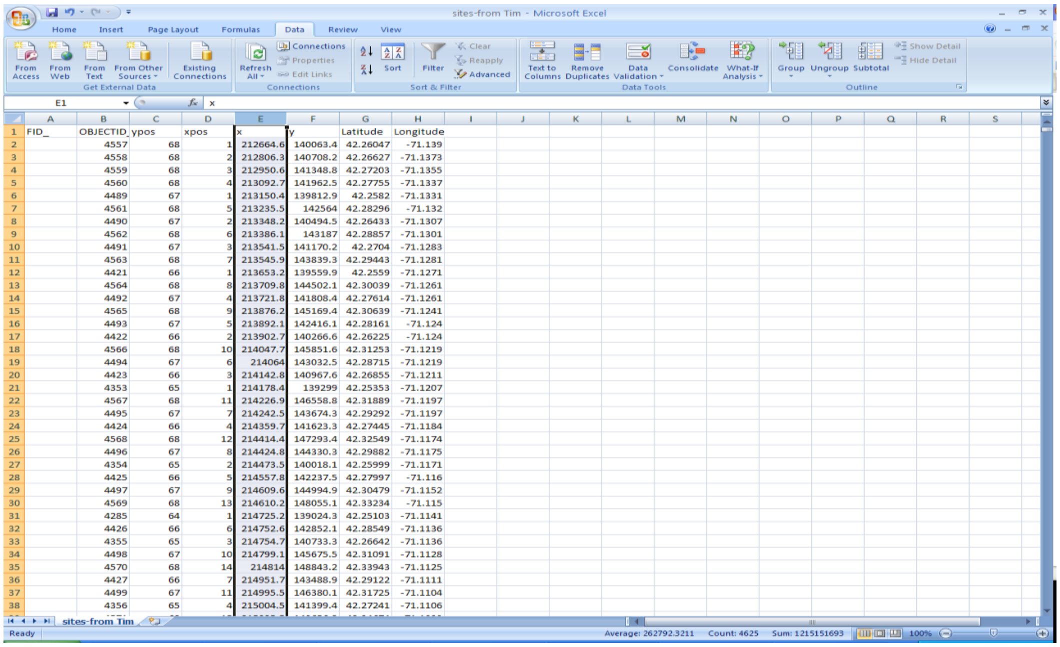
10. In the Spatial Reference Properties dialog, look over the details to make sure things look good.

11. Click OK to close the Spatial Reference Properties dialog.

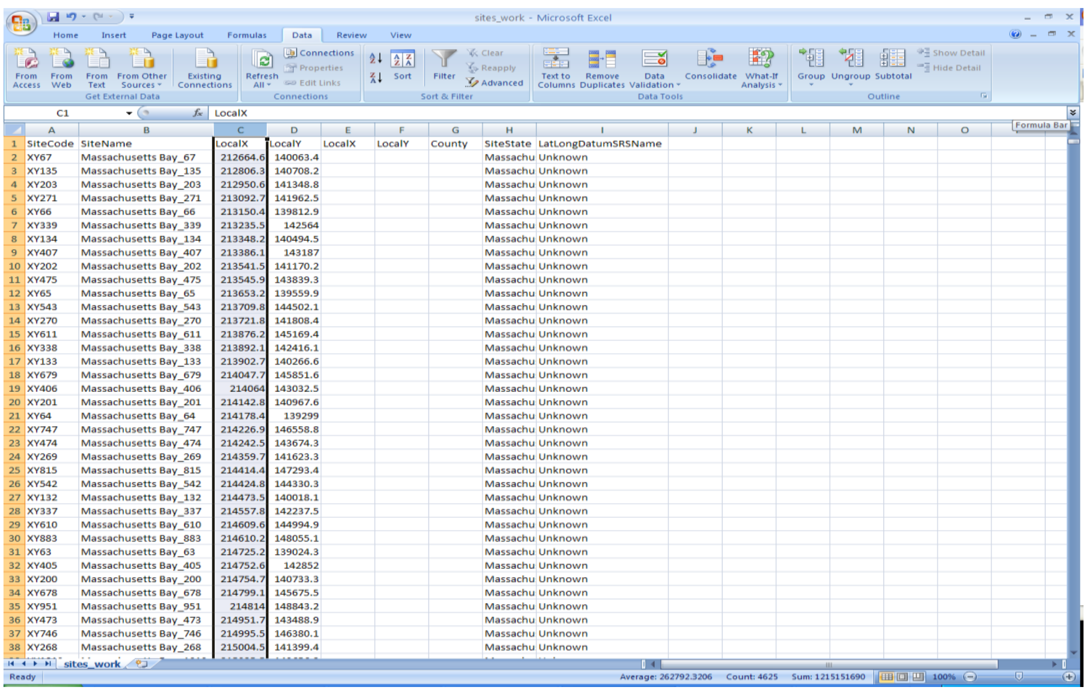
12. With the coordinate system set and MySites chosen, click OK to execute the Define Projection tool.

13. Now carry on with step 4 from my original instructions.

No

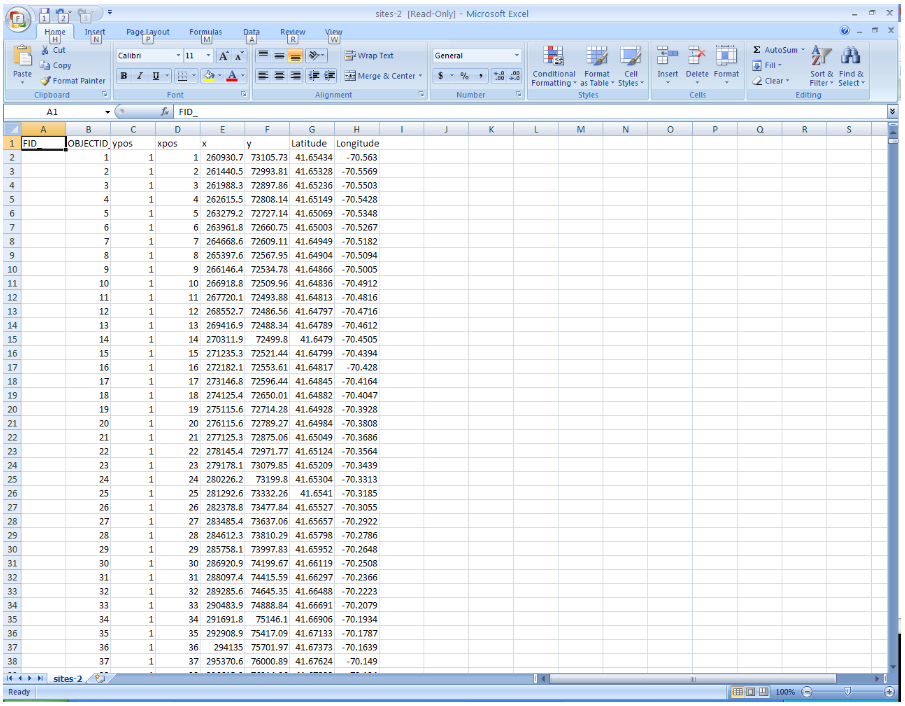


FID_	OBJECTID_	ypos	xpos	x	y	Latitude	Longitude
	4557	68	1	212664.6	140063.4	42.26047	-71.139
	4558	68	2	212806.3	140708.2	42.26627	-71.1373
	4559	68	3	212950.6	141348.8	42.27203	-71.1355
	4560	68	4	213092.7	141962.5	42.27755	-71.1337
	4489	67	1	213150.4	139812.9	42.2582	-71.1331
	4561	68	5	213235.5	142564	42.28296	-71.132
	4490	67	2	213348.2	140494.5	42.26433	-71.1307
	4562	68	6	213386.1	143187	42.28857	-71.1301
	4491	67	3	213541.5	141170.2	42.2704	-71.1283
	4563	68	7	213545.9	143839.3	42.29443	-71.1281
	4421	66	1	213653.2	139559.9	42.2559	-71.1271
	4564	68	8	213709.8	144502.1	42.30039	-71.1261
	4492	67	4	213721.8	141808.4	42.27614	-71.1261
	4565	68	9	213876.2	145169.4	42.30639	-71.1241
	4493	67	5	213892.1	142416.1	42.28161	-71.124
	4422	66	2	213902.7	140266.6	42.26225	-71.124
	4566	68	10	214047.7	145851.6	42.31253	-71.1219
	4494	67	6	214064	143032.5	42.28715	-71.1219
	4423	66	3	214142.8	140967.6	42.26855	-71.1211
	4353	65	1	214178.4	139299	42.25353	-71.1207
	4567	68	11	214226.9	146558.8	42.31889	-71.1197
	4495	67	7	214242.5	143674.3	42.29292	-71.1197
	4424	66	4	214359.7	141623.3	42.27445	-71.1184
	4568	68	12	214414.4	147293.4	42.32549	-71.1174
	4496	67	8	214424.8	144330.3	42.29882	-71.1175
	4354	65	2	214473.5	140018.1	42.25999	-71.1171
	4425	66	5	214557.8	142237.5	42.27997	-71.116
	4497	67	9	214609.6	144994.9	42.30479	-71.1152
	4569	68	13	214610.2	148055.1	42.33234	-71.115
	4285	64	1	214725.2	139024.3	42.25103	-71.1141
	4426	66	6	214752.6	142852.1	42.28549	-71.1136
	4355	65	3	214754.7	140733.3	42.26642	-71.1136
	4498	67	10	214799.1	145675.5	42.31091	-71.1128
	4570	68	14	214814	148843.2	42.33943	-71.1125
	4427	66	7	214951.7	143488.9	42.29122	-71.1111
	4499	67	11	214995.5	146380.1	42.31725	-71.1104
	4356	65	4	215004.5	141399.4	42.27241	-71.1106

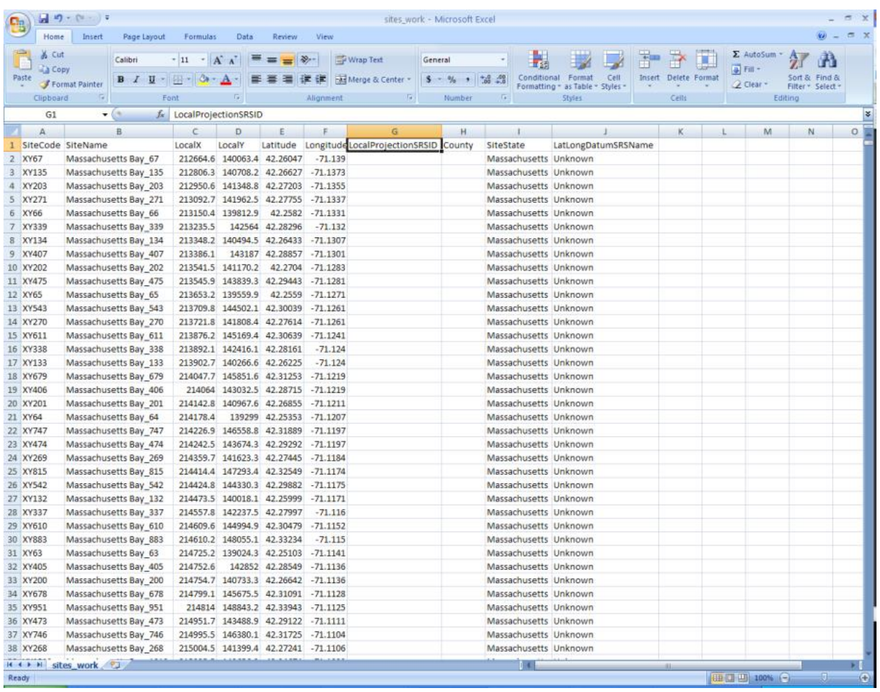


SiteCode	SiteName	LocalX	LocalY	LocalX	LocalY	County	SiteState	LatLongDatumSRSName
XY67	Massachusetts Bay_67	212664.6	140063.4				Massachu	Unknown
XY135	Massachusetts Bay_135	212806.3	140708.2				Massachu	Unknown
XY203	Massachusetts Bay_203	212950.6	141348.8				Massachu	Unknown
XY271	Massachusetts Bay_271	213092.7	141962.5				Massachu	Unknown
XY66	Massachusetts Bay_66	213150.4	139812.9				Massachu	Unknown
XY339	Massachusetts Bay_339	213235.5	142564				Massachu	Unknown
XY134	Massachusetts Bay_134	213348.2	140494.5				Massachu	Unknown
XY407	Massachusetts Bay_407	213386.1	143187				Massachu	Unknown
XY202	Massachusetts Bay_202	213541.5	141170.2				Massachu	Unknown
XY475	Massachusetts Bay_475	213545.9	143839.3				Massachu	Unknown
XY65	Massachusetts Bay_65	213653.2	139559.9				Massachu	Unknown
XY543	Massachusetts Bay_543	213709.8	144502.1				Massachu	Unknown
XY270	Massachusetts Bay_270	213721.8	141808.4				Massachu	Unknown
XY611	Massachusetts Bay_611	213876.2	145169.4				Massachu	Unknown
XY338	Massachusetts Bay_338	213892.1	142416.1				Massachu	Unknown
XY133	Massachusetts Bay_133	213902.7	140266.6				Massachu	Unknown
XY679	Massachusetts Bay_679	214047.7	145851.6				Massachu	Unknown
XY406	Massachusetts Bay_406	214064	143032.5				Massachu	Unknown
XY201	Massachusetts Bay_201	214142.8	140967.6				Massachu	Unknown
XY64	Massachusetts Bay_64	214178.4	139299				Massachu	Unknown
XY747	Massachusetts Bay_747	214226.9	146558.8				Massachu	Unknown
XY474	Massachusetts Bay_474	214242.5	143674.3				Massachu	Unknown
XY269	Massachusetts Bay_269	214359.7	141623.3				Massachu	Unknown
XY815	Massachusetts Bay_815	214414.4	147293.4				Massachu	Unknown
XY542	Massachusetts Bay_542	214424.8	144330.3				Massachu	Unknown
XY132	Massachusetts Bay_132	214473.5	140018.1				Massachu	Unknown
XY337	Massachusetts Bay_337	214557.8	142237.5				Massachu	Unknown
XY610	Massachusetts Bay_610	214609.6	144994.9				Massachu	Unknown
XY883	Massachusetts Bay_883	214610.2	148055.1				Massachu	Unknown
XY63	Massachusetts Bay_63	214725.2	139024.3				Massachu	Unknown
XY405	Massachusetts Bay_405	214752.6	142852				Massachu	Unknown
XY200	Massachusetts Bay_200	214754.7	140733.3				Massachu	Unknown
XY678	Massachusetts Bay_678	214799.1	145675.5				Massachu	Unknown
XY951	Massachusetts Bay_951	214814	148843.2				Massachu	Unknown
XY473	Massachusetts Bay_473	214951.7	143488.9				Massachu	Unknown
XY746	Massachusetts Bay_746	214995.5	146380.1				Massachu	Unknown
XY268	Massachusetts Bay_268	215004.5	141399.4				Massachu	Unknown

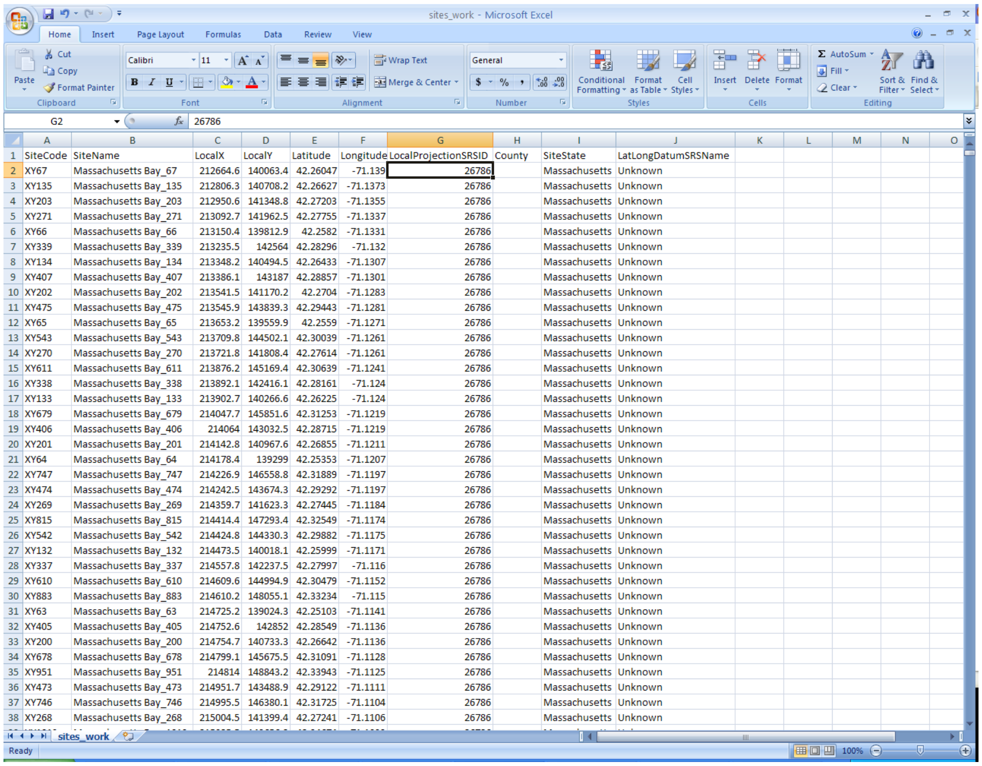
Now copy paste those Latitude and Longitude columns to original sites.csv file



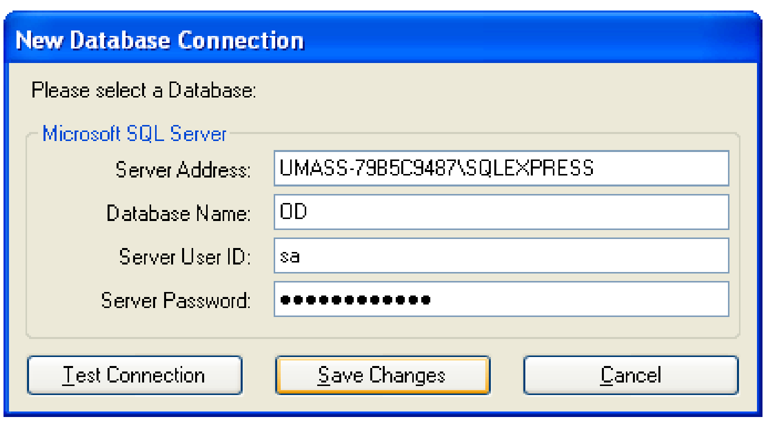
Now insert LocalProjectSRSID column



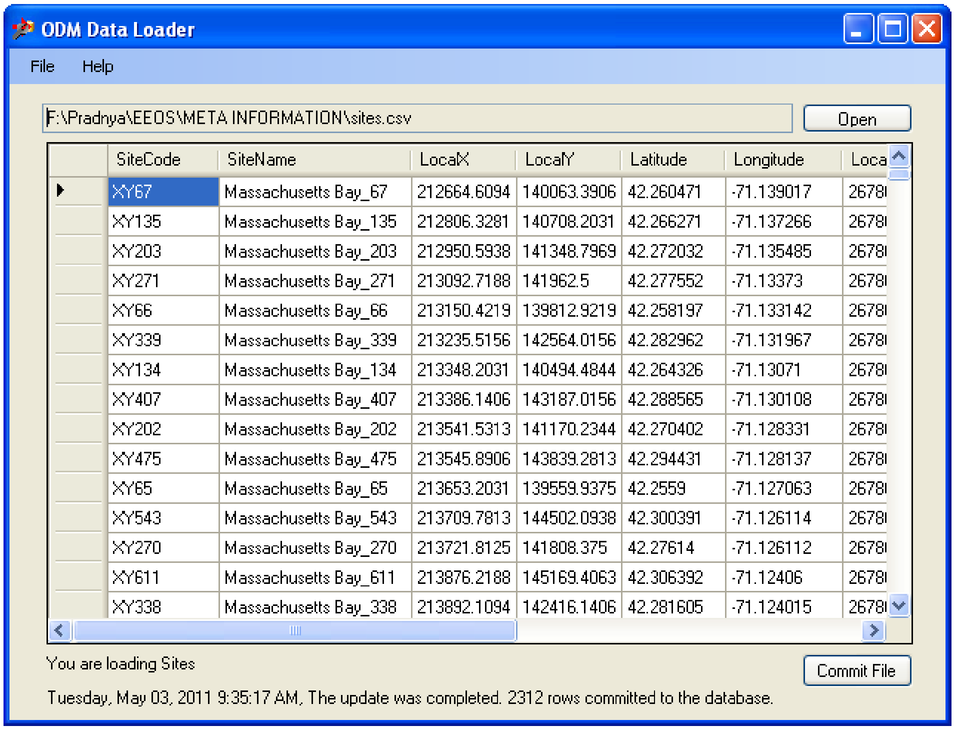
Execute select query on Spatial references table in OD database and you will find the local project ID for NAD27-Massachusetts-Mainland as shown below



Now open ODM_DL1.1.3 and specify username and password



Open sites.csv file and click Commit File.



SiteCode	SiteName	LocalX	LocalY	Latitude	Longitude	Loca
XY67	Massachusetts Bay_67	212664.6094	140063.3906	42.260471	-71.139017	2678
XY135	Massachusetts Bay_135	212806.3281	140708.2031	42.266271	-71.137266	2678
XY203	Massachusetts Bay_203	212950.5938	141348.7969	42.272032	-71.135485	2678
XY271	Massachusetts Bay_271	213092.7188	141962.5	42.277552	-71.13373	2678
XY66	Massachusetts Bay_66	213150.4219	139812.9219	42.258197	-71.133142	2678
XY339	Massachusetts Bay_339	213235.5156	142564.0156	42.282962	-71.131967	2678
XY134	Massachusetts Bay_134	213348.2031	140494.4844	42.264326	-71.13071	2678
XY407	Massachusetts Bay_407	213386.1406	143187.0156	42.288565	-71.130108	2678
XY202	Massachusetts Bay_202	213541.5313	141170.2344	42.270402	-71.128331	2678
XY475	Massachusetts Bay_475	213545.8906	143839.2813	42.294431	-71.128137	2678
XY65	Massachusetts Bay_65	213653.2031	139559.9375	42.2559	-71.127063	2678
XY543	Massachusetts Bay_543	213709.7813	144502.0938	42.300391	-71.126114	2678
XY270	Massachusetts Bay_270	213721.8125	141808.375	42.27614	-71.126112	2678
XY611	Massachusetts Bay_611	213876.2188	145169.4063	42.306392	-71.12406	2678
XY338	Massachusetts Bay_338	213892.1094	142416.1406	42.281605	-71.124015	2678

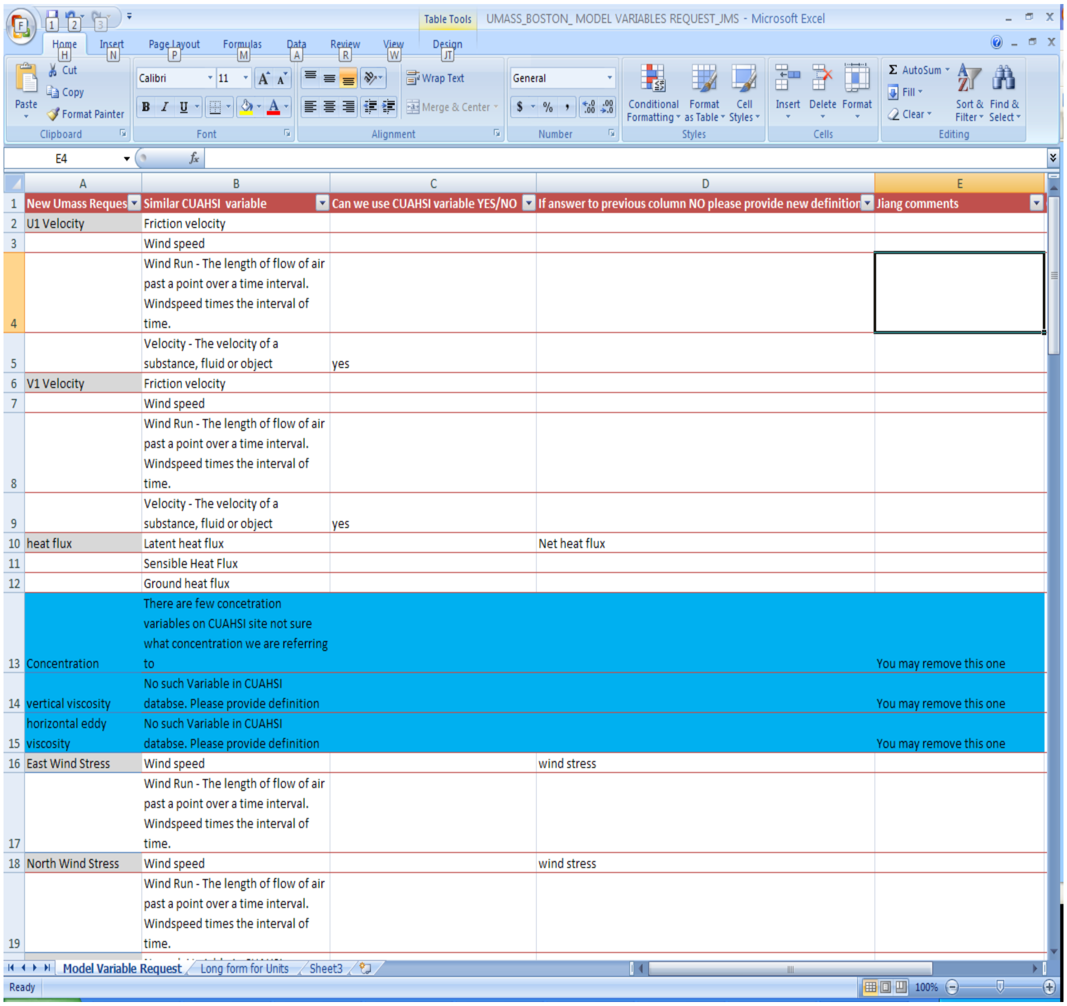
The records will be written to the Sites Table in OD database.

Preparing and Uploading Model Data Variables Information

Before uploading variables information, one must decide which variables do not exist in Variables table and VariablesNamesCV table in OD database and needs to be created.

I have already uploaded sites,variables,sources meta-information to the OD database. If you are using same machine as I did, with OD database available on SQL Server 2008, then skip this section 7.1 and 7.2

Refer section 8.4 for viewing the netcdf file for Hydrodynamic/WaterQuality model .And list them as shown below, and send it to domain expert to short-list the variables whose data needs to be uploaded. This file is located in \ folder

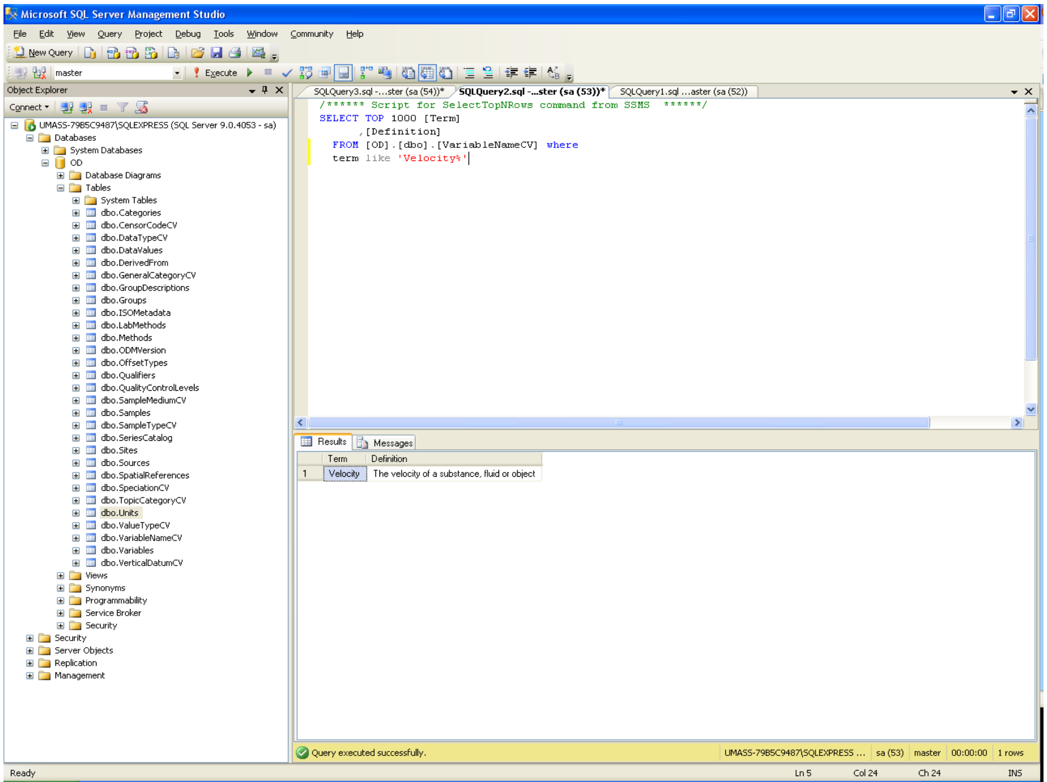


New Umass Request	Similar CUAHSI variable	Can we use CUAHSI variable YES/NO	If answer to previous column NO please provide new definition	Jiang comments
U1 Velocity	Friction velocity			
	Wind speed			
	Wind Run - The length of flow of air past a point over a time interval. Windspeed times the interval of time.			
	Velocity - The velocity of a substance, fluid or object	yes		
V1 Velocity	Friction velocity			
	Wind speed			
	Wind Run - The length of flow of air past a point over a time interval. Windspeed times the interval of time.			
	Velocity - The velocity of a substance, fluid or object	yes		
heat flux	Latent heat flux		Net heat flux	
	Sensible Heat Flux			
	Ground heat flux			
Concentration	There are few concetration variables on CUAHSI site not sure what concentration we are referring to			You may remove this one
vertical viscosity	No such Variable in CUAHSI databse. Please provide definition			You may remove this one
horizontal eddy viscosity	No such Variable in CUAHSI databse. Please provide definition			You may remove this one
East Wind Stress	Wind speed		wind stress	
	Wind Run - The length of flow of air past a point over a time interval. Windspeed times the interval of time.			
North Wind Stress	Wind speed		wind stress	
	Wind Run - The length of flow of air past a point over a time interval. Windspeed times the interval of time.			

Open the template for meta-information.csv located in \Templates FROM Yoori Choi .Use the variables.csv sheet.

Assign a variable code for the variable that need to be created.

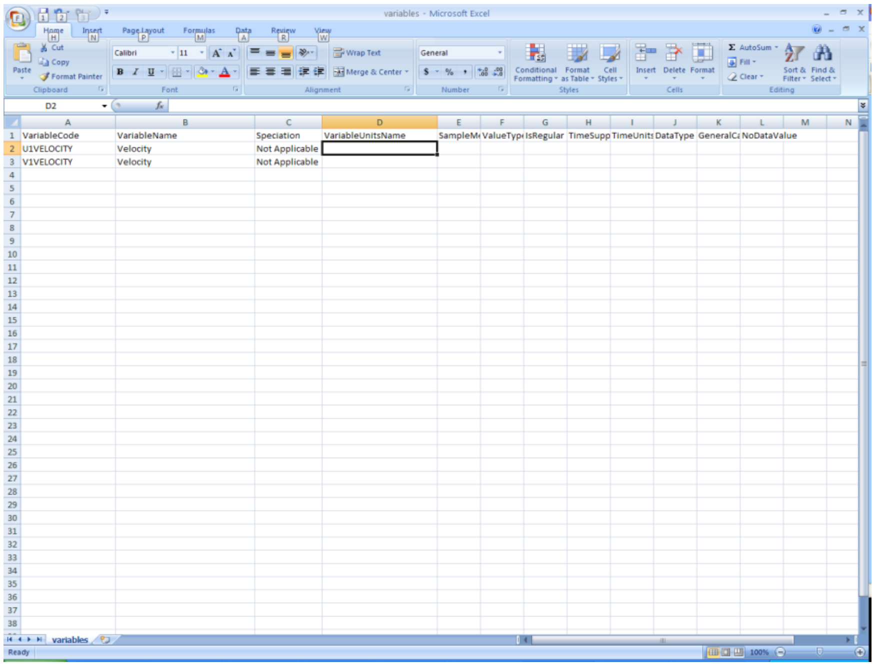
On getting confirmation as shown in image above, check to see if those variables are available in VariablesNamesCV as shown below. The example is to check if any variable like velocity exists.



If it does, then copy the term field to VariableName field in variables.csv go to step 9

Else request the variable creation in Controlled Vocabulary see Section 8.6

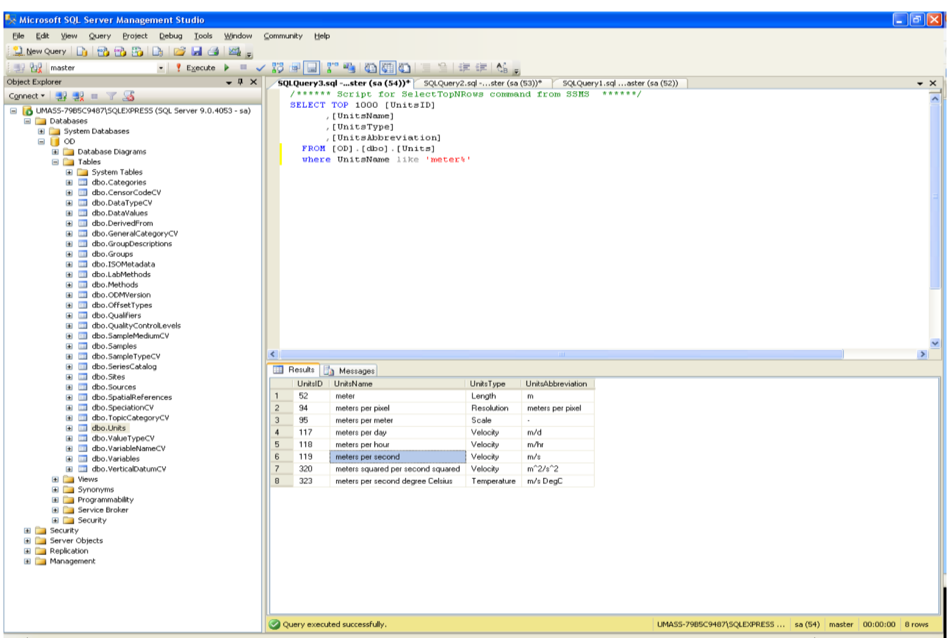
After updating local UMass Boston Server go to Step 4.



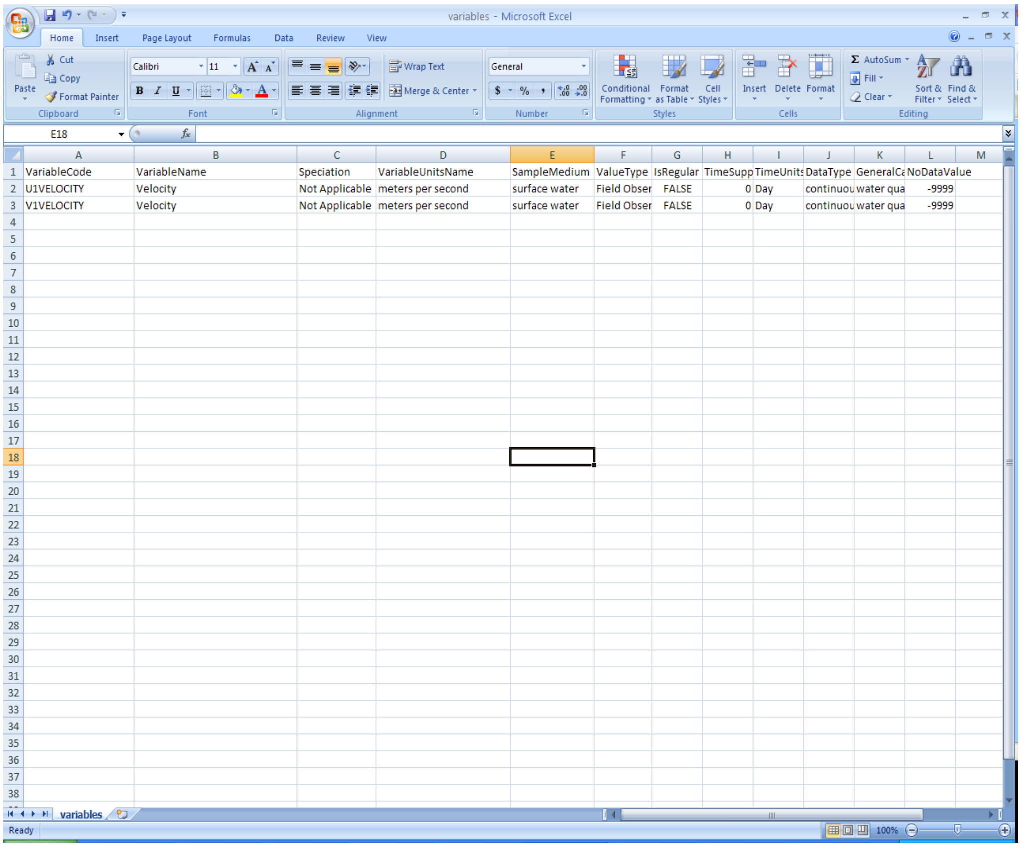
Now, fill in the VariableUnitsName column, by executing similar query

And copying the UnitsName field to VariableUnitsName column. If the unit

does not exist in Units table, then request its creation, see Section 8.6 and section 8.1 and then proceed to next step.



Repeat procedure from step 4, for all the variables that needs to be created and send it to domain expert to fill out remaining fields.



After the entire variables.csv sheet is created, follow section 8.2 to upload variables.csv

Preparing and Uploading Model Data – DataValues Meta-Information

Model data is available in netcdf file format. So I used Numeric Python (NumPy)

module to extract information from netcdf files. See section 4 for software requirements.

We have two models

hydro-dynamic model data (located in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \cd Data\Hydro Model)

water quality model data. (located in KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \cd Data\Water Quality Model)

Both model data has few variables that has 4-dimensions, few has 3-dimensions, 2-dimensions and 1-dimension.

To generate data for HydroModel 4-D variables

For Hydro-Model, to generate data for 4-D variables, keep the netcdf file in same directory as the program KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Python Scripts\Model_DataValue_4D.py and the sitesCodeTableSheet.csv

Change the year , month and timerange in code as shown below

```
#******************** MUST CHANGE BEFORE RUNNING THIS SCRIPT ***************

# timerange will be 59 for Jan,Mar,May,July,Aug,Oct & Dec files

# timerange will be 57 for Apr,June,Sept,Nov

# timerange will be 54 for Feb (no loop year)

timerange =59
```

Run the script, it takes some time to generate the data.

Output is created in Output_4D folder.

Run program in EEOS.Util.Batch folder CreateBatchFile_TEMP and CreateBatchFile_RestD which creates the batch file containing commands to upload datavalues.csv to SQL Server.

Double click the batch file to start uploading the data.

To generate data for HydroModel 3-D variables

For Hydro-Model, to generate data for 3-D variables, keep the netcdf file in same directory as the program KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Python Scripts\Model_DataValue_3D.py and the sitesCodeTableSheet.csv

Change the year , month and timerange in code as shown below

```
#******************** MUST CHANGE BEFORE RUNNING THIS SCRIPT ****************

# timerange will be 59 for Jan,Mar,May,July,Aug,Oct & Dec files

# timerange will be 57 for Apr,June,Sept,Nov

# timerange will be 54 for Feb (no loop year)

timerange =59
```

Run the script, it takes a while to generate the data.

Output is created in Output_3D folder.

Run program in EEOS.Util.Batch folder CreateBatchFile_3D.java which creates the batch file.

Double click the batch file to start uploading the data.

To generate data for 2-D variables

For Hydro-Model, to generate data for 2-D variables, keep the netcdf file in same directory as the program KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Python Scripts\Model_DataValue_2D.py and the sitesCodeTableSheet.csv

Change the file name and save the program

```
rootgrp = Dataset('ecom2005.12.cdf', 'r', format='NETCDF3')
```

Run the script, it takes a while to generate the data.

Output is created in Output_2D folder.Repeat steps b to c for every month file for that year and then go to next step

Run program in EEOS.Util.Batch folder CreateBatchFile_2D.java which creates the batch file.

Double click the batch file to start uploading the data.

To generate data for 1-D variables

For Hydro-Model, to generate data for 1-D variables, keep the netcdf file in same directory as the program KDL Server(Pradnya's profile) or F:\Pradnya\EEOS \Python Scripts\Model_DataValue_1D.py and the sitesCodeTableSheet.csv

Change the year , month and timerange in code as shown below

```
#******************** MUST CHANGE BEFORE RUNNING THIS SCRIPT ***************

# timerange will be 59 for Jan,Mar,May,July,Aug,Oct & Dec files

# timerange will be 57 for Apr,June,Sept,Nov

# timerange will be 54 for Feb (no loop year)
```

c. Run the script, it takes a while to generate the data.

Output is created in Output_1D folder. Repeat steps b to c for every month of that year.

Run program in EEOS.Util.Batch folder CreateBatchFile_1D.java which creates the batch file for entire year.

Double click the batch file to start uploading the data

To generate data for 4-D variables in WaterModel

Water Model had huge number of records – 120* 10*68*54 for around 19 variables each. Python does not allow to keep open file handles beyond certain limit. So I have divided the script in part1, part2 and part3 .

For Water-Model, to generate data for 4-D variables, keep sitesCodeTableSheet.csv and the netcdf file in same directory as the program \Python Scripts\WaterModel\WaterModel_DataValue_4D_part1.py or part2.py or part3.py

Change the year and daysInMonth (if leap year or not) in code as shown below

```
******************** MUST CHANGE BEFORE RUNNING THIS SCRIPT ****************

# timerange will be 120 for all years


timerange =120

year=2000;
```

c. Run the script, it takes a while to generate the data.

Output is created in WaterOutput_4DPartX folder(X= 1,2 or 3). Create the batch file for entire year.

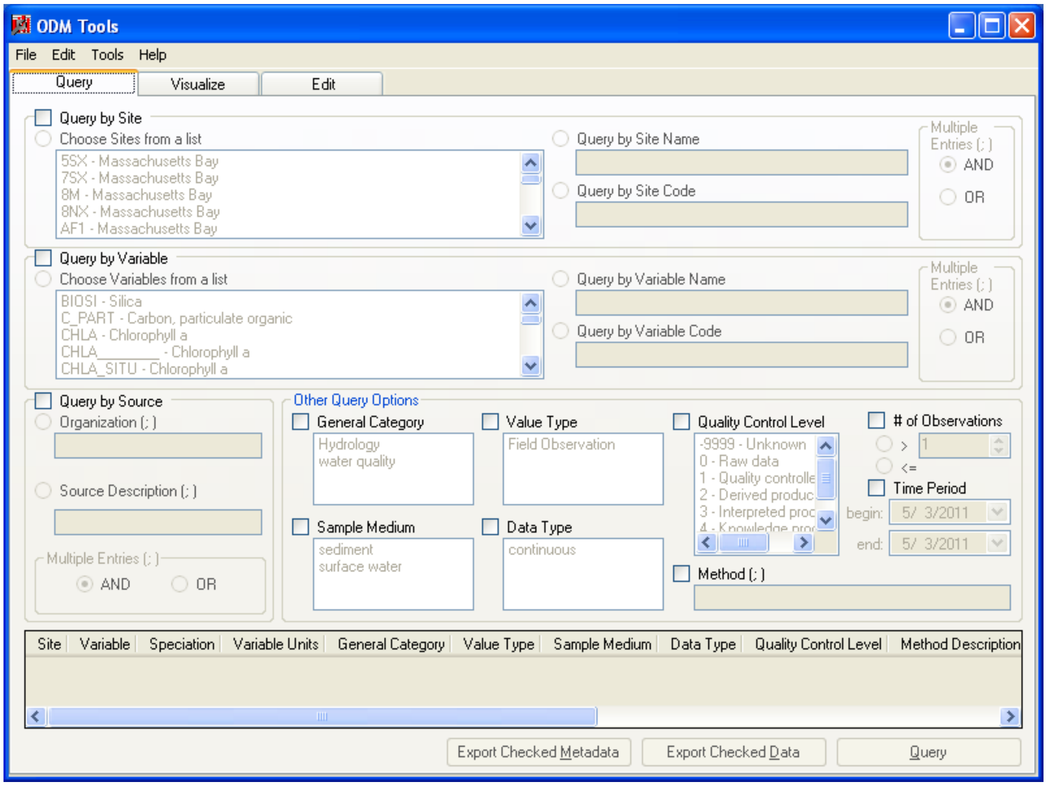
Double click the batch file to start uploading the data

Repeat above steps, to upload data for all other years.

Some Generic Information

Updating Controlled Vocabularies

Open the ODM Tool 1.1



Click Tools - > Quick CV Update

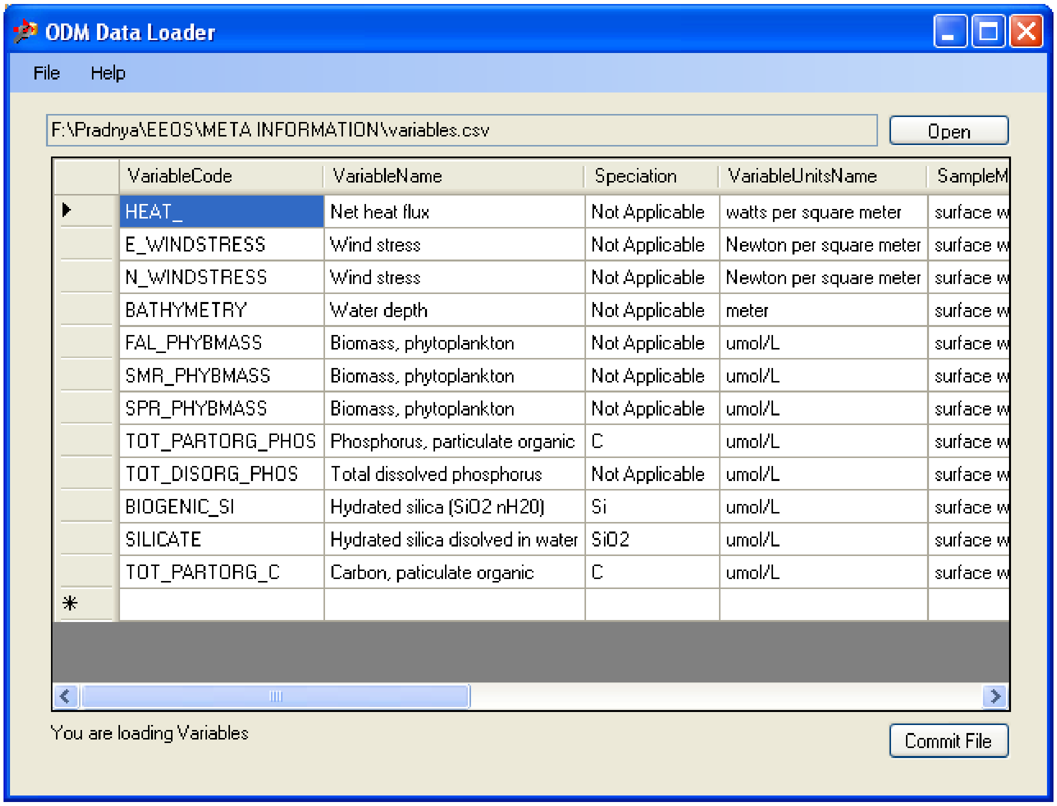
OD Database controlled vocabulary tables will be updated.

Uploading Variables information

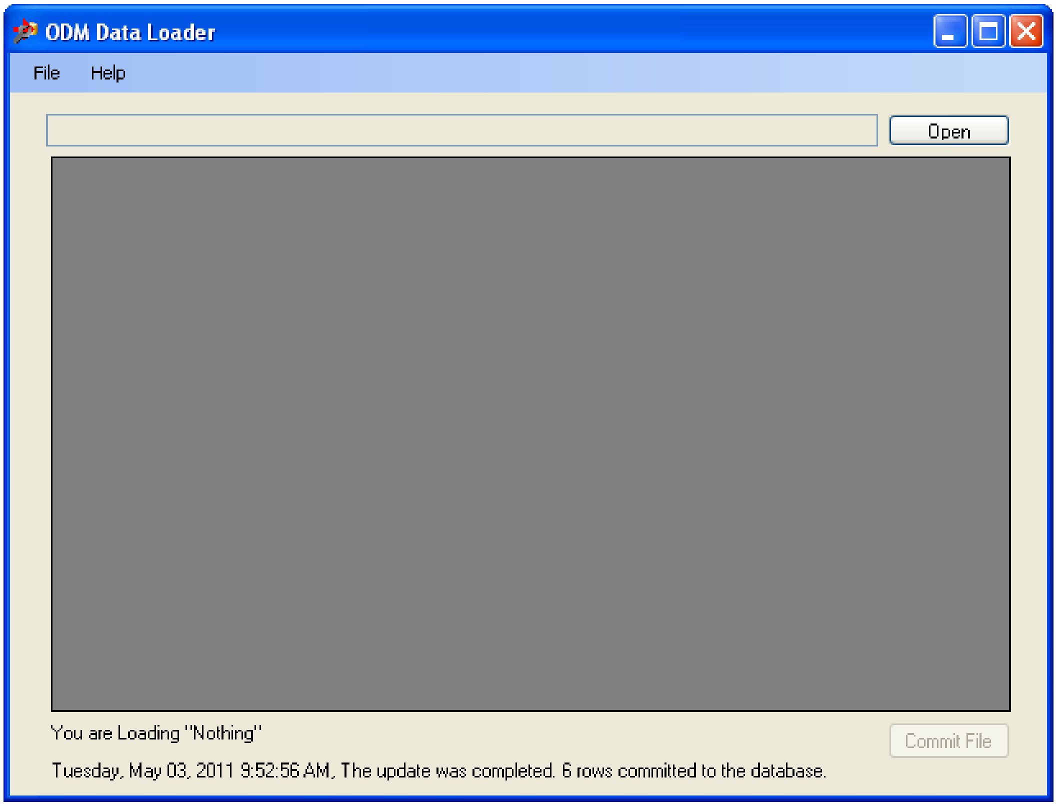
Open ODM_DL 1.1.3 tool (ODM Data Loader)

Specify the path of variables.csv file

Click on commit file button



After successful upload, following screen will be shown

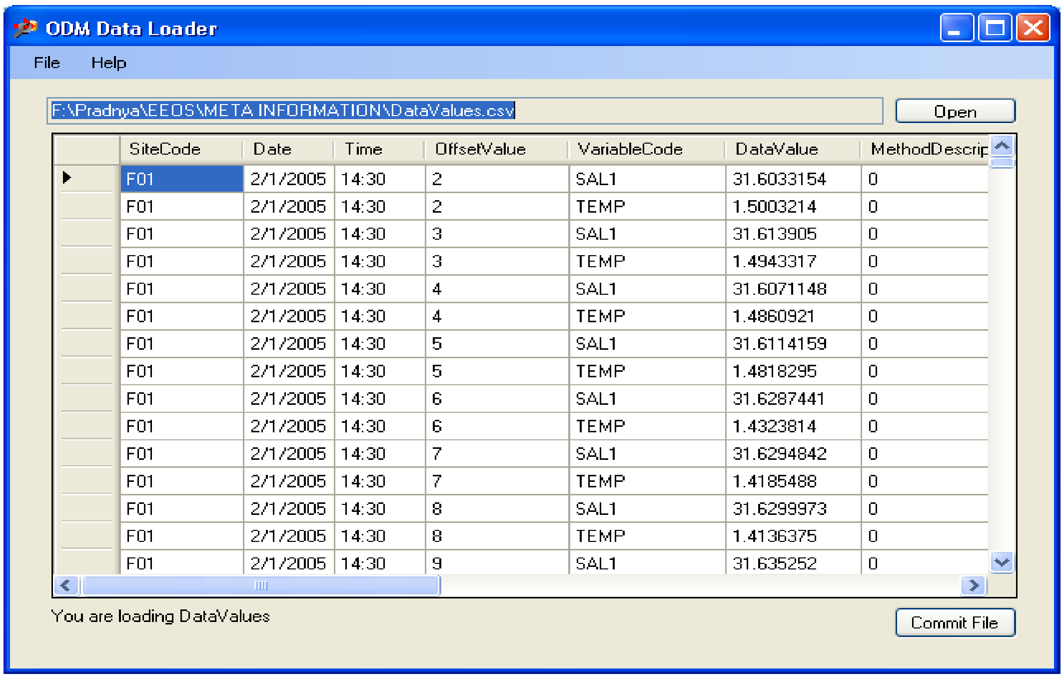

ODM Data Loader
File
Help
Open
You are Loading "Nothing"
Commit File
Tuesday, May 03, 2011 9:52:56 AM, The update was completed. 6 rows committed to the database.

Uploading DataValues Information

Open ODM_DL 1.1.3 tool (ODM Data Loader)

Specify the path of DataValues.csv file

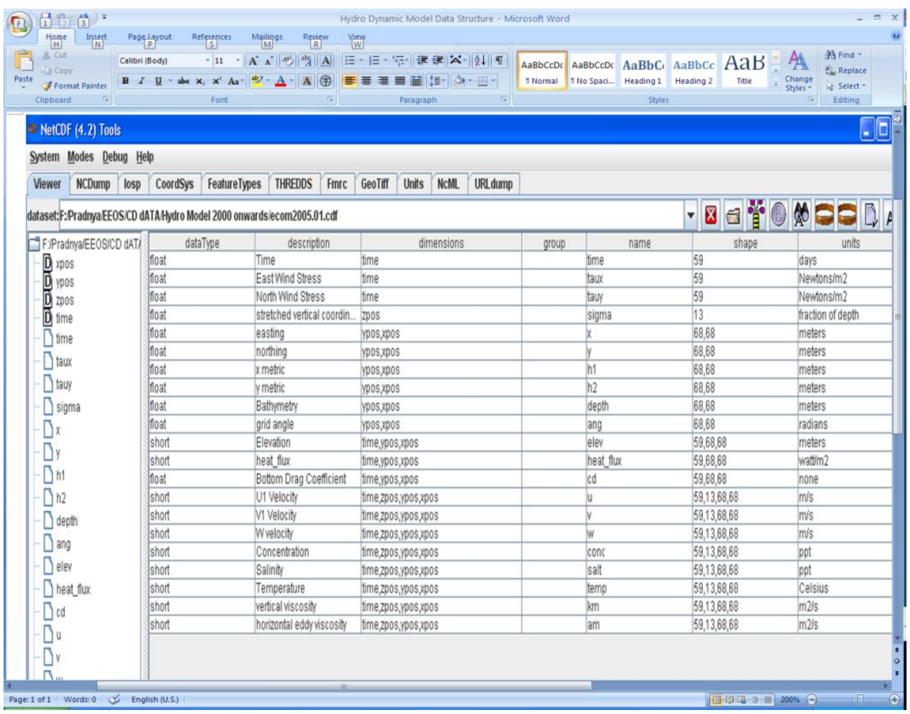
Click on commit file button to upload data



Viewing model data structure and fieldnames using ToolsUI

To view the structure of model data files, the field name used or the unit in which the data was measured, ToolsUI tool can be used.

Open ToolsUI, specify the .netcdf file whose structure is to be viewed (All Hydro-Dynamic Model files have same structure. Similarly, all Water Quality Model data files share same structure)



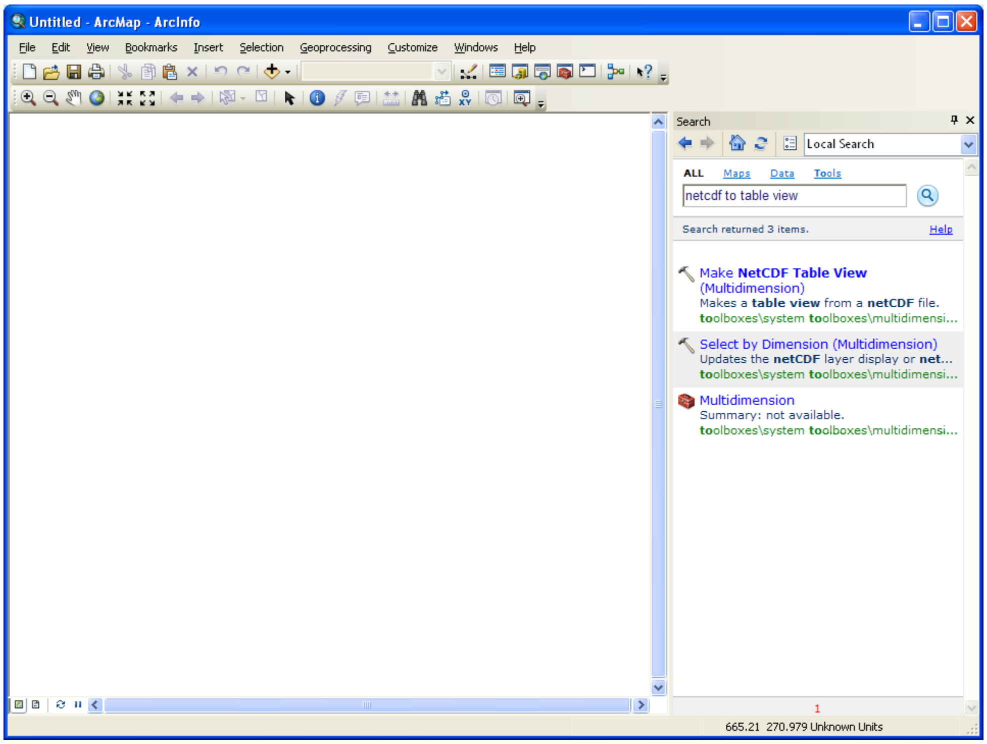
dataType	description	dimensions	group	name	shape	units
float	Time	time		time	59	days
float	East Wind Stress	time		taux	59	Newtons/m2
float	North Wind Stress	time		tauy	59	Newtons/m2
float	stretched vertical coordin...	zpos		sigma	13	fraction of depth
float	easting	ypos,xpos		x	68,68	meters
float	northing	ypos,xpos		y	68,68	meters
float	x metric	ypos,xpos		h1	68,68	meters
float	y metric	ypos,xpos		h2	68,68	meters
float	Bathymetry	ypos,xpos		depth	68,68	meters
float	grid angle	ypos,xpos		ang	68,68	radians
short	Elevation	time,ypos,xpos		elev	59,68,68	meters
short	heat_flux	time,ypos,xpos		heat_flux	59,68,68	watt/m2
float	Bottom Drag Coefficient	time,ypos,xpos		cd	59,68,68	none
short	U1 Velocity	time,zpos,ypos,xpos		u	59,13,68,68	m/s
short	V1 Velocity	time,zpos,ypos,xpos		v	59,13,68,68	m/s
short	W velocity	time,zpos,ypos,xpos		w	59,13,68,68	m/s
short	Concentration	time,zpos,ypos,xpos		conc	59,13,68,68	ppt
short	Salinity	time,zpos,ypos,xpos		salt	59,13,68,68	ppt
short	Temperature	time,zpos,ypos,xpos		temp	59,13,68,68	Celsius
short	vertical viscosity	time,zpos,ypos,xpos		km	59,13,68,68	m2/s
short	horizontal eddy viscosity	time,zpos,ypos,xpos		am	59,13,68,68	m2/s

As shown in image above, Concentration, Temperature etc are 4-dimensional variables, while elevation and angle are 3-dimensional variables

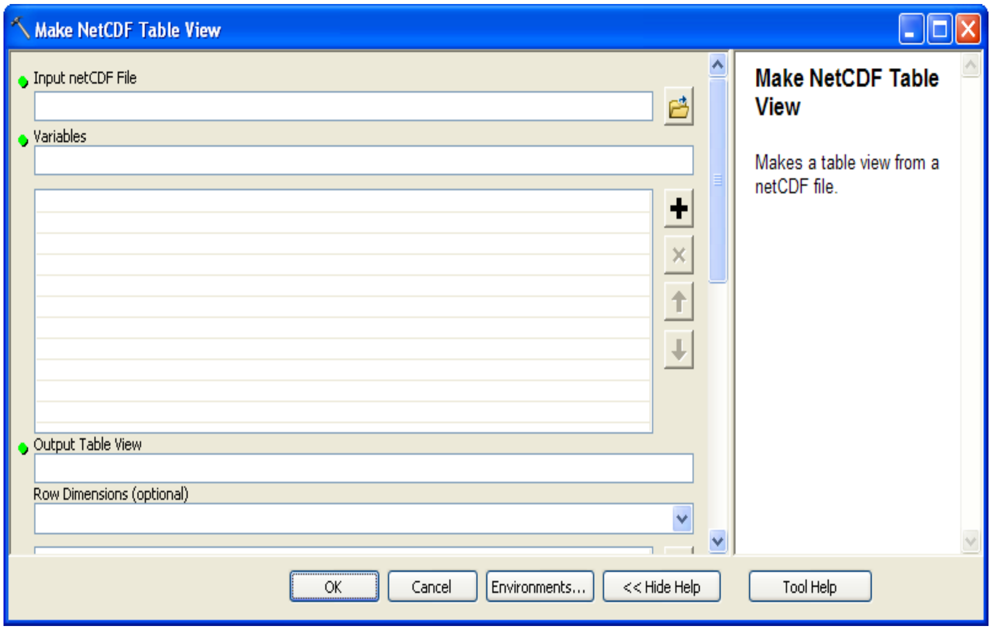
The dimensions column specify the dimension variables used to store the measured variable's data

Viewing model data contents through ArcMap

Open ArcMap 10 .In search box type netcdf to table view

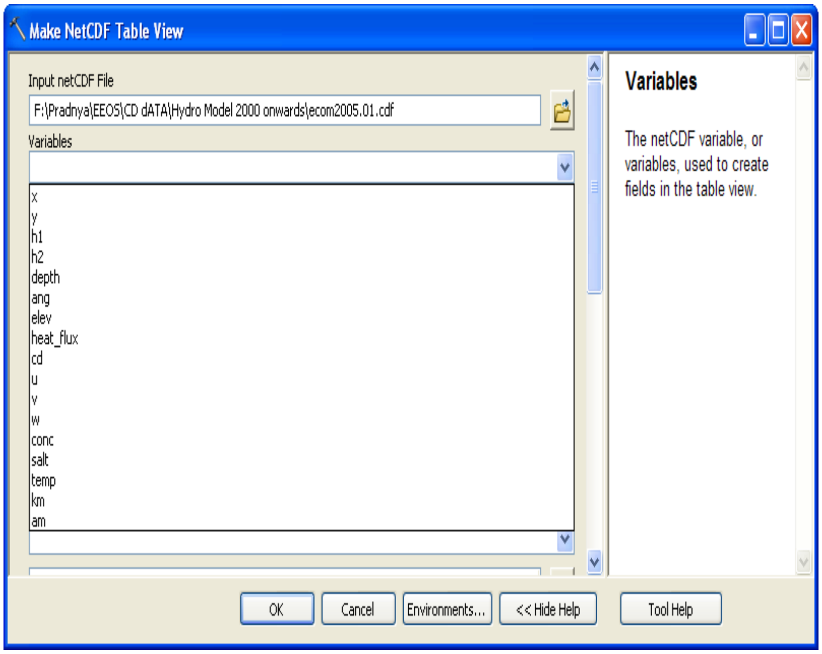


Click first search item which will display window below

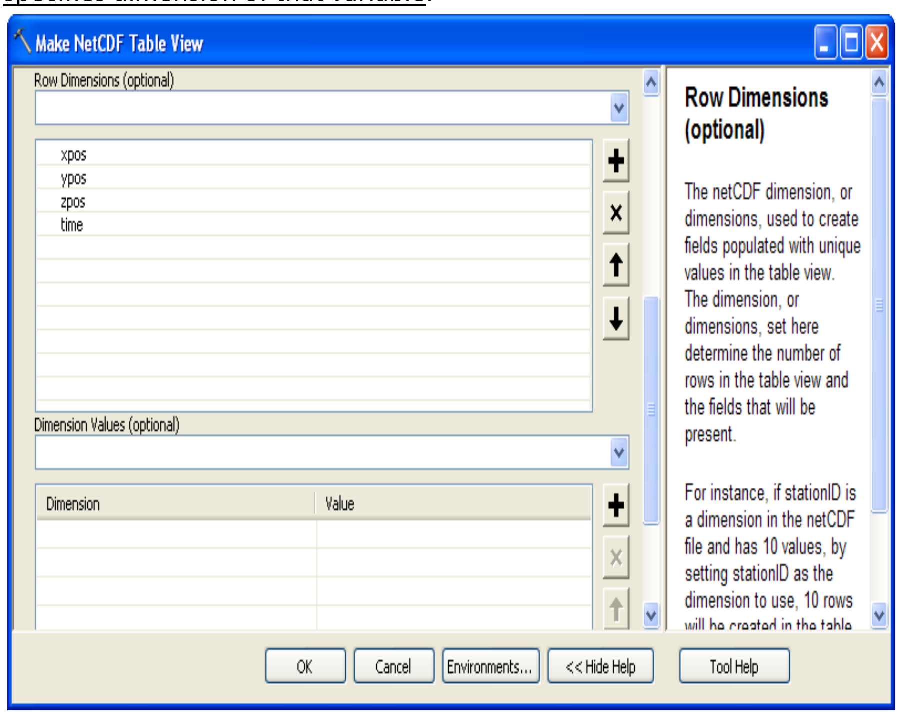


Open the required file (example is of Hydro dynamic mode file 2005-01) and click on drop down box to select the required variables (example is of all variables which have 4 dimensions)

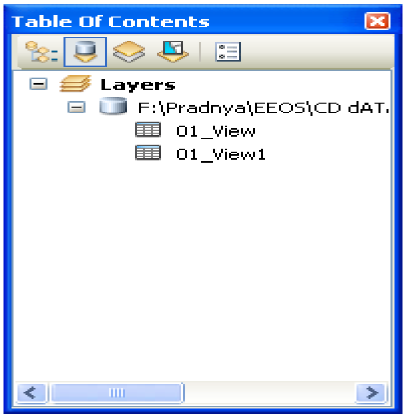
You view data, you should variables which have same dimension. For example – elevation, heat_flux and angle in HydroDynamic model are 3-D variables. If you mix up, variable with different dimensions then number of records selected by ArcMap will be of the variable having smallest dimension.So if along with elevation, heat_flux you select sigma, then only 13 records will be shown, since sigma is 1-D variable having 13 values. Use ToolsUI to view dimension information in section 8.4



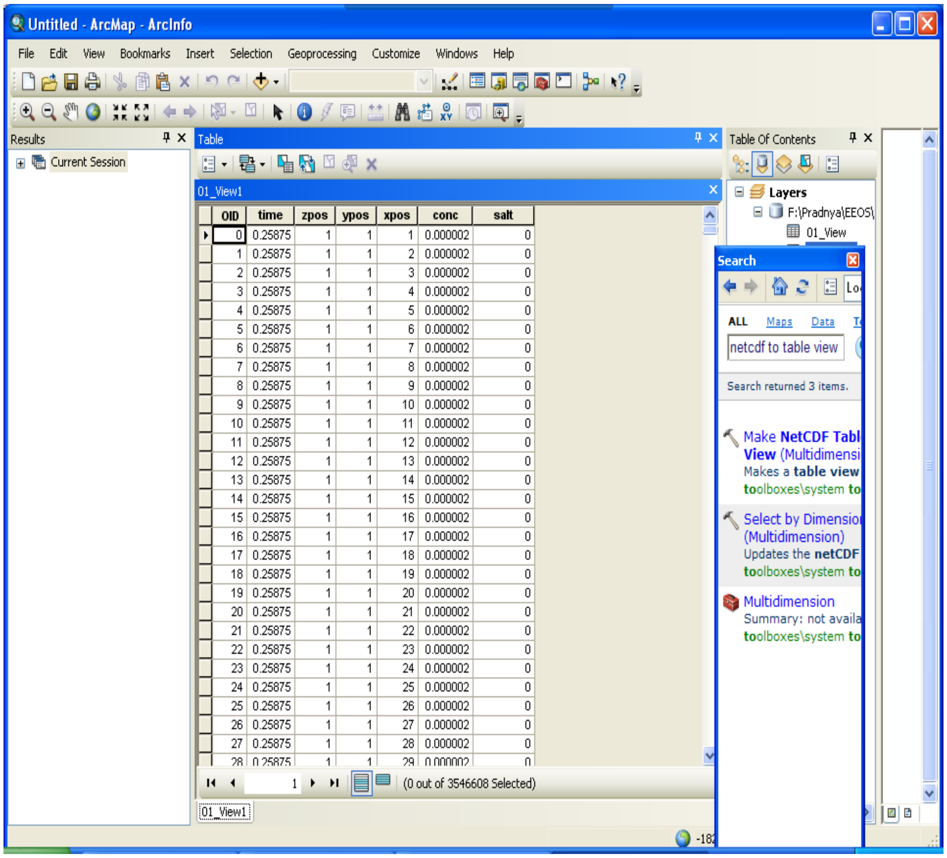
Specify row dimension (either through drop down box or by typing the correct dimension variable name) and then click OK. <u>See Section 8.4 to view structure of model data and dimension column specifies dimension of that variable</u>.



It will create table view (01_View1 in this example)

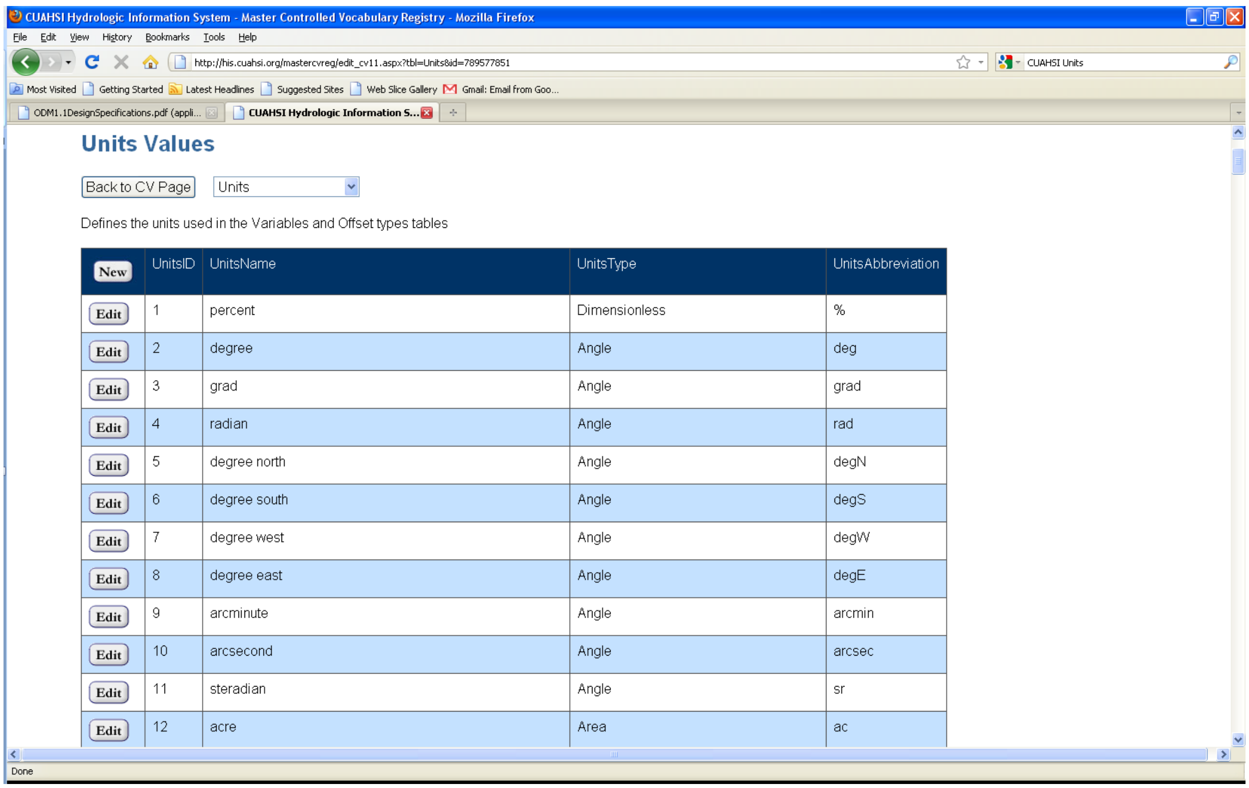


Right Click on view name and click open


Untitled - ArcMap - ArcInfo
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
Results
Current Session
Table
01_View1
OID time zpos ypos xpos conc salt
0 0.25875 1 1 1 0.000002 0
1 0.25875 1 1 2 0.000002 0
2 0.25875 1 1 3 0.000002 0
3 0.25875 1 1 4 0.000002 0
4 0.25875 1 1 5 0.000002 0
5 0.25875 1 1 6 0.000002 0
6 0.25875 1 1 7 0.000002 0
7 0.25875 1 1 8 0.000002 0
8 0.25875 1 1 9 0.000002 0
9 0.25875 1 1 10 0.000002 0
10 0.25875 1 1 11 0.000002 0
11 0.25875 1 1 12 0.000002 0
12 0.25875 1 1 13 0.000002 0
13 0.25875 1 1 14 0.000002 0
14 0.25875 1 1 15 0.000002 0
15 0.25875 1 1 16 0.000002 0
16 0.25875 1 1 17 0.000002 0
17 0.25875 1 1 18 0.000002 0
18 0.25875 1 1 19 0.000002 0
19 0.25875 1 1 20 0.000002 0
20 0.25875 1 1 21 0.000002 0
21 0.25875 1 1 22 0.000002 0
22 0.25875 1 1 23 0.000002 0
23 0.25875 1 1 24 0.000002 0
24 0.25875 1 1 25 0.000002 0
25 0.25875 1 1 26 0.000002 0
26 0.25875 1 1 27 0.000002 0
27 0.25875 1 1 28 0.000002 0
(0 out of 3546608 Selected)
01_View1
Table Of Contents
Layers
F:\Pradnya\EEOS\
01_View
Search
ALL Maps Data
netcdf to table view
Search returned 3 items.
Make NetCDF Tabl
View (Multidimensi
Makes a table view
toolboxes\system to
Select by Dimensio
(Multidimension)
Updates the netCDF
toolboxes\system to
Multidimension
Summary: not availa
toolboxes\system to

Sending Units Creation Request to CUAHSI

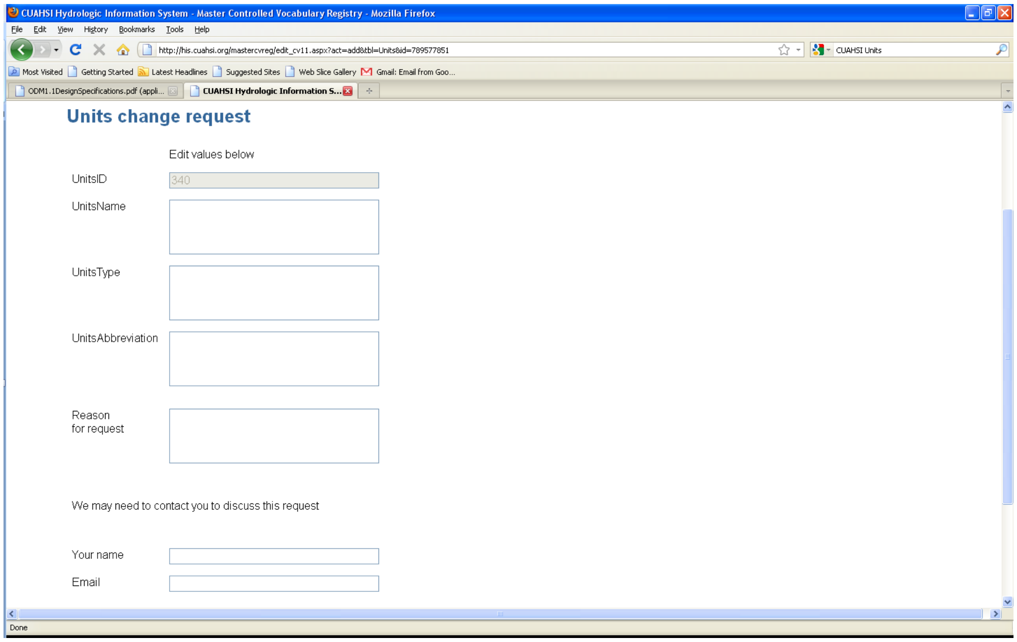
If you did not find the unit that means it does not exist in CUAHSI Controlled Vocabulary. So you need to request it to CUAHSI to create such unit .Please go to the link http://his.cuahsi.org/mastercvreg/cv11.aspx



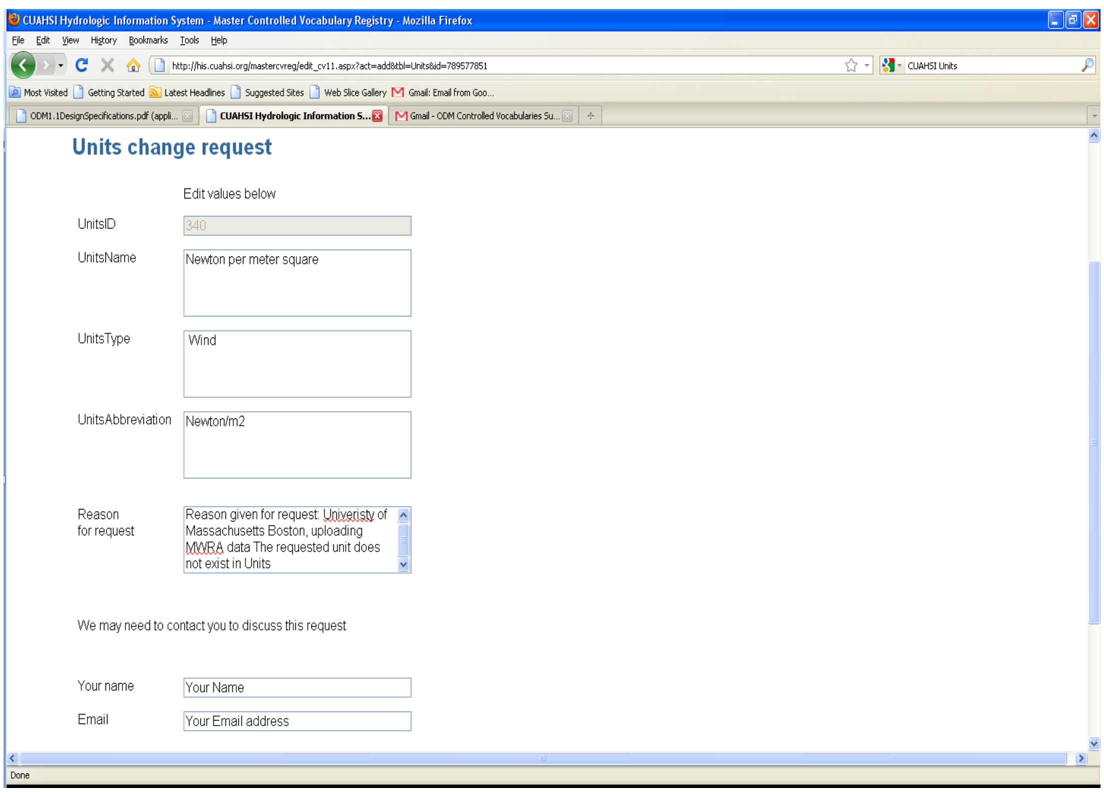
New	UnitsID	UnitsName	UnitsType	UnitsAbbreviation
Edit	1	percent	Dimensionless	%
Edit	2	degree	Angle	deg
Edit	3	grad	Angle	grad
Edit	4	radian	Angle	rad
Edit	5	degree north	Angle	degN
Edit	6	degree south	Angle	degS
Edit	7	degree west	Angle	degW
Edit	8	degree east	Angle	degE
Edit	9	arcminute	Angle	arcmin
Edit	10	arcsecond	Angle	arcsec
Edit	11	steradian	Angle	sr
Edit	12	acre	Area	ac

a. Click on New

b. Fill in the information required



Something like this



c. Also, provide your name and email address. You will get email looking like this

Thank you for your submission to ODM Controlled Vocabularies.

Table: Units
Request: Add New Entry
ID/Term Affected: 340

UnitsID: 340
UnitsName: Newton per meter square
UnitsType: Wind
UnitsAbbreviation: Newton/m2

Reason given for request: Univeristy of Massachusetts Boston, uploading MWRA data The requested unit does not exist in Units

d. Repeat steps from 3 to 6 for all the variables.

e. Follow up with CUAHSI for units or variable names creation in controlled vocabulary.

f. After CUAHSI updated controlled vocabulary, you will receive similar email

Thank you for submitting a Master Controlled Vocabulary change request. In response to your request the master controlled vocabulary at http://his.cuahsi.org/mastercvreg/ has been updated.

Table: Units
Request: Add
ID/Term Affected: 337

UnitsID: 337
UnitsName: Newton per square meter
UnitsType: Pressure/Stress

g) After you receive emails for all units and variables creation in controlled vocabulary, follow section 8.1 to update UMass Boston - Units table located on SQL Server.

Sending Variables Creation Request to CUAHSI

If you did not find the variable in VariableNameCV that means it does not exist in CUAHSI Controlled Vocabulary. So you need to request it to CUAHSI to create such variable .Please go to the link

http://his.cuahsi.org/mastercvreg/edit_cv11.aspx?tbl=VariableNameCV&id=821577965

a. Click on New

b. Fill in the information required

c. Also, provide your name and email address. You will get email looking like this

Thank you for your submission to ODM Controlled Vocabularies.

Table: VariableNameCV
Request: Add New Entry
ID/Term Affected: biogenic silica

Term: biogenic silica
Definition: biogenic silica

Reason given for request: Univeristy of Massachusetts Boston, uploading MWRA data The requested variable does not exist in VariablesCV

This request will be reviewed by a moderator and you will be notified once the status of the submission changes.

David Tarboton
Utah State University
david.tarboton@usu.edu

d. Repeat steps from 3 to 6 for all the variables.

e. Follow up with CUAHSI for units or variable names creation in controlled vocabulary.

f. After CUAHSI updated controlled vocabulary, you will receive similar email

Thank you for submitting a Master Controlled Vocabulary change request. In response to your request the master controlled vocabulary at http://his.cuahsi.org/mastercvreg/ has been updated.

Table: VariableNameCV
Request: Add
ID/Term Affected: Biogenic silica

Term: Biogenic silica
Definition: Hydrated silica (SiO2 nH20)

Administrator notes: Ok with edits - KATS

You may update your version of ODM with the latest terms by following the instructions at http://his.cuahsi.org/mastercvreg/cv11.aspx.

Kim Schreuders

g) After you receive emails for all variables creation in controlled vocabulary, follow section 8.1 to update UMass Boston - Variables table located on SQL Server

SQL Server User Name Password

UMASS-79B5C9487\SQLEXPRESS

login : sa

password: Passw0rd2011

Installed at

d:\program Files(x86)\Sql Server...

Current Status of Uploaded Data

Observational Data

As of 14th May 2011, all required observational data (specified by domain expert) in KDL Server or F:\Pradnya\EEOS followed by\CD dATA folder has been uploaded to the OD database in the SQL Server installed at location d:\program Files(x86)\Sql Server.

So, if you are using same machine in KDLab as I did, with SQL Server 2008 intact with OD database then you need not reload the observations data.

If SQL Server 2008 is not installed on your machine then refer section 4 and follow instructions to upload all required software. Once it has been done, configure the backup OD.bak file located at \db Backup to newly installed SQL Server 2008. Follow steps provided at the link below

http://msdn.microsoft.com/en-us/library/ms177429.aspx

If the actual UMass Boston Server has been step up by IT group, then in that case as well, you can use OD.bak or OD database to configure OD database on that machine.

The Observations Data is located at \CD Data \MWRA 1998-2005 folder. Domain expert specified what variables to be uploaded and the sheet names from which to upload data.

File Name	SheetName	Status	Comments
BFLUX_YYYY	All Data,PoreWater,Redox,Solid,Nutrients	Uploaded	

hydro_fluorlight_YYYY	data	**Uploaded**	
hydro_nutrients_YYYY	YYYY = year (1998,1999,… etc)	**Uploaded**	
hydro_productivity_YYYY	data	**Uploaded**	
phyto_sp_biomass_YYYY	data	**Uploaded**	
zooplankton_counts_YYYY	Not Required – as per domain expert	**Not Required**	

Model Data

As of 27th May 2011, **not all model data** (specified by domain expert) in \CD dATA folder **has been uploaded** to the OD database in the SQL Server installed at location d:\program Files(x86)\Sql Server.

So, if you are using same machine in KDLab as I did, with SQL Server 2008 intact with OD database then **you need to continue uploading** the model data where I left of. Please see table below.

If SQL Server 2008 is not installed on your machine then refer section 4 and follow instructions to upload all required software. Once it has been done, configure the backup OD.bak file located at \db Backup to newly installed SQL Server 2008. Follow steps provided at the link below

http://msdn.microsoft.com/en-us/library/ms177429.aspx

If the actual UMass Boston Server has been step up by IT group, then in that case as well, you can use OD.bak or OD database to configure OD database on that machine.

The Model Data is located at \CD Data \ Hydro Model 2000 onwards or \CD Data \ Water Quality Model folder. Domain expert specified what variables to be uploaded and the sheet names from which the data needs to be uploaded.

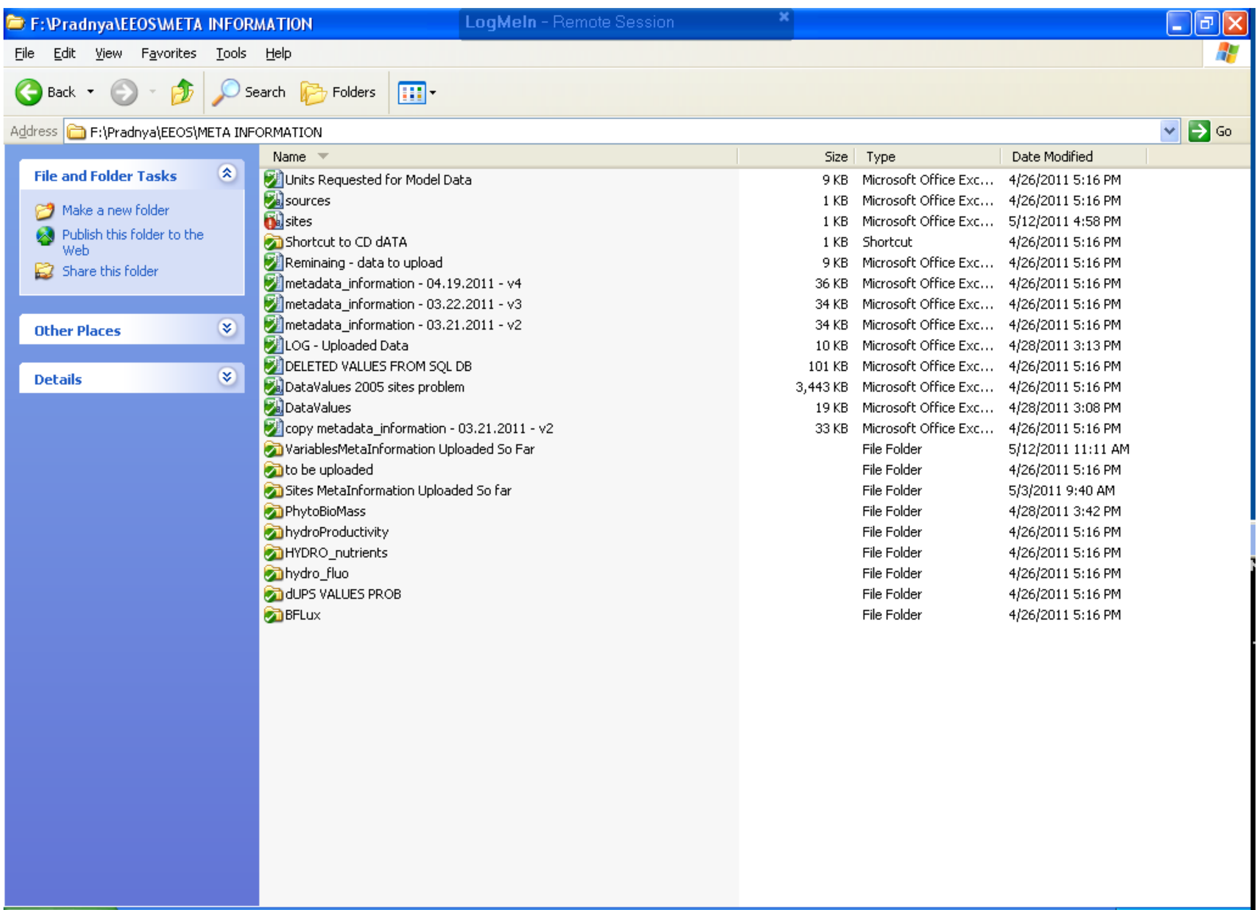
File Name	Variables	Status	Comments
Ecom2005.01(Hydro -Model)	Temperature, Elevation, Heat, Bathymetry	**Uploaded**	
Ecom2005.01(Hydro -Model)	Salinity (Uploaded till 18th file, need to upload from 19th file onwards See Section 10 for more details)	**Incomplete**	
Ecom2005.01(Hydro-Model)	U1Velocity,V1Velocity	**Not Uploaded**	
Ecom2005.02(Hydro -Model) to Ecom2005.12	Bathymetry	**Uploaded**	

Execute the Python scripts first to generate files and then use batch script to upload data. See section 7.3 to generate and upload model data

Reloading Data (Meta-information)

For some reason, if SQL Server 2008 is not able to configure OD database(from OD.bak), or SQL Server 2008 that I had installed with OD database is not available, then instead of re-executing programs to load data(DataValues.csv,sites.csv or variables.csv), you can re-use the saved files (sites.csv, variables.csv ,DataValues.csv) that I had generated by running programs.

Reloading Observations Data (meta-information)



The meta information folder has saved all files that I used to upload data.

Reloading sites meta-information

\META INFORMATION\Sites MetaInformation Uploaded So far contains all sites- partXXXX.csv

Rename the sites- partXXX.xsv to sites.csv and use ODM DataLoader (See section 8.2 or 8.3) to reload sites data .

Repeat steps for each saved sites file

Reloading variables meta-information

\META INFORMATION\VariablesMetaInformation Uploaded So Far contains all

Variables- partXXXX.csv

Rename the variables- partXXX.xsv to variables.csv and use ODM DataLoader (See section 8.2) to reload variables data.

Repeat steps for each saved variables file

Reloading DataValues meta-information

\META INFORMATION folder contains subfolders

BFLUX

PhytoBioMass

hydro_fluo

HYDRO_NUTRIENTS

hydroProductivity

Each of these folders contain DataValues-partXXXX.csv files.

Rename the DataValues- partXXX.xsv to DataValues.csv and use ODM DataLoader (See section 8.3) to reload data.

Repeat steps for each saved DataValues file

Reloading Model Data (meta-information)

For some reason, if SQL Server 2008 is not able to configure OD database(from OD.bak), or SQL Server 2008 that I had installed with OD database is not available, then instead of re-executing programs to load data(DataValues.csv,sites.csv or variables.csv), you can re-use the saved files (sites.csv, variables.csv ,DataValues.csv) that I had generated by running programs.

Reloading sites meta-information

\META INFORMATION\Sites MetaInformation Uploaded So far contains all sites- partXXXX.csv (if you have already executed it in section 10.1 then skip this part)

Rename the sites- partXXX.xsv to sites.csv and use ODM DataLoader (See section 8.2 or 8.3) to reload sites data .

Repeat steps for each saved sites file

Reloading variables meta-information

\META INFORMATION\VariablesMetaInformation Uploaded So Far contains all

Variables- partXXXX.csv (If you have already reloaded these files in section 10.1 then skip this part)

Rename the variables- partXXX.xsv to variables.csv and use ODM DataLoader (See section 8.2) to reload variables data.

Repeat steps for each saved variables file

Reloading DataValues meta-information

Current Status of uploaded data

File Name	Variables	Status	Comments
Ecom2005.01(Hydro - Model)	Temperature, Elevation, Heat, Bathymetry	**Uploaded**	
Ecom2005.01(Hydro - Model)	Salinity(need to upload 19th file onwards in\Python Scripts\Output4D\Salinity folder)	**Incomplete**	
Ecom2005.01(Hydro-Model)	U1Velocity,V1Velocity	**Not Uploaded**	
Ecom2005.02(Hydro - Model) to Ecom2005.12	Bathymetry	**Uploaded**	

I have created files for the year 2005 (only) in \Python Scripts folder by name Output_4D) .You need to run scripts for other months for year 2005 for 3D and 2D variables. For all other years you need to execute the Python scripts first to generate datavalues.csv files and then use batch script to upload data See section 7.3

contains the DataValues_VVVV_N.csv (where VVVV= Variable Name and N= number of sheet) . Use batch programs to reload the data. See section 7.3 (1.e and 1.f)

Currently the upload is very slow, takes about 30-60 minutes to upload a .csv file containing about 70,000 records. The logged ticket is below:

http://hydroserver.codeplex.com/workitem/8264

In case, the issue is not resolved, then a program could be written which does job as the data loader tool, which will read the generated .csv file, do ALL validations and constraint checks the tool currently does and then write and update data to the required tables in the SQL Server.

APPENDIX B - Environmental Indicator Selection

HEPATITIS A

Hepatitis A Virus (HAV) is a virus that replicates in human liver cells, causing an immune responses that can interfere with liver functioning.{{702 Jacobsen, Kathryn H. 2009; }} There are an estimated 1.5 million clinical cases of HAV worldwide, with many more undocumented (sub-clinical) cases. {{706 Wasley, A. 2006}} Symptoms of HAV infection range from gastrointestinal problems to jaundice, but children often exhibit mild or non-specific symptoms.{{702 Jacobsen, Kathryn H. 2009; }} HAV is transmitted via the fecal-oral route either by direct contact with an infectious person or by ingestion of contaminated food or water. People infected with HAV can shed the virus in their stool several weeks before the onset of symptoms. {{702 Jacobsen, Kathryn H. 2009}}

Childhood exposure or vaccination is important because the severity of disease typically increases with age; while not usually fatal, HAV infection can cause an extended illness with a long convalescence period. {{702 Jacobsen, Kathryn H. 2009}} Unvaccinated children and young adults are the most vulnerable to HAV infection. {{702 Jacobsen, Kathryn H. 2009}} Routine childhood vaccination against HAV was first recommended in the United States starting in 1999 for states with elevated infection rates{{706 Wasley, A. 2006}} and was expanded nationwide in October 2005 for all children aged 12–23 months.{{706 Wasley, A. 2006}} In Massachusetts, among children aged 19-35 months the HAV vaccination rate is estimated at just below 50%.{{707 Wooten, K.G. 2010}}) A low childhood exposure/ vaccination rate translates into a higher 'adult susceptibility' rate. {{702 Jacobsen, Kathryn H. 2009}}

Hepatitis A virus (HAV) has no known animal reservoir, so its presence in the environment is correlated with human waste. {{706 Wasley, A. 2006}} Although specific to humans, HAV can persist in the environment in an infectious form for weeks, perhaps months, at a time.{{706 Wasley, A. 2006; 719 Rzeżutka, A. 2004;715 Bloch, A.B. 1990; }} Any water body receiving human sewage, even if partially treated, could be contaminated with HAV, because it can survive in groundwater and soils, and on nonporous surfaces.{{719 Rzeżutka, A. 2004}} HAV can remain present in groundwater for at least 3 months, and can possibly cause infection through drinking water supplies, even when they have acceptable levels of fecal coliforms.{{715 Bloch, A.B. 1990}} Because of this viral resiliency the most important determinants for the risk of HAV transmission in an area have to do with the presence and amount of human fecal material and the exposure risk of immunologically naïve people (i.e., socio-economic factors). Despite the availability of a vaccine HAV continues to pose a health threat to unexposed persons in both developed and developing countries. While not a high profile disease in all parts of the United States, an investigation of one HAV outbreak associated with raw shellfish in Panama City, Florida, suggested that even approved shellfish harvesting areas could be contaminated with

HAV.{{714 Desenclos, J-C. A. 1991}} While HAV might be a risk heading for the sunset, it is not over the horizon yet.

Environmental determinants of Hepatitis A Virus	**References**
Relative Humidity (for surfaces)	{{717 Mbithi, J.N. 1991}}
Air Temperature	{{717 Mbithi, J.N. 1991}}
Rainfall (for influence on runoff level, virus survival)	{{719 Rzeżutka, A. 2004}}

Initial socio-economic determinants of Hepatitis A Virus	**References**
Level of sewage treatment	{{702 Jacobsen, Kathryn H. 2009}} {{706 Wasley, A. 2006}}
Number and flow level of Combined Sewer Overflows	Inferred from: {{702 Jacobsen, Kathryn H. 2009}}
Riverine discharge (HAV can survive in freshwater)	{{719 Rzeżutka, A. 2004}}
Loading / Leachate from Septic systems	{{719 Rzeżutka, A. 2004}}
Bather population at beach	inferred from: {{717 Mbithi, J.N. 1991}} {{702 Jacobsen, Kathryn H. 2009}}
Beach debris and sanitation: sanitary plastics, etc (HAV can survive on polystyrene)	{{717 Mbithi, J.N. 1991}}, and inferred from: {{702 Jacobsen, Kathryn H. 2009}}
Location of bather facilities (showers, lavatories) and relevance of input from these sources to beach	Inferred from: {{717 Mbithi, J.N. 1991}}{{719 Rzeżutka, A. 2004}} {{702 Jacobsen, Kathryn H. 2009}}

Vibrio parahaemilyticus

Vibrio parahaemolyticus is a cosmopolitan species of bacteria found in marine and estuarine waters.{{692 Joseph,S.W. 1982;677 Martinez-Urtaza, J. 2008; 675 Su,Yi-Cheng 2007; 657 Iwamoto, Martha 2010; }} *V. parahaemolyticus* can cause food poisoning when consumed humans and has been linked to seafood-borne outbreaks since it was first isolated by Japanese scientists in 1950.{{692 Joseph,S.W. 1982; }} *V. parahaemolyticus* was first identified in the United States in 1971 after large outbreaks of gastroenteritis associated with consumption of improperly cooked crabs, since then, sporadic outbreaks related to the consumption of raw shellfish or cooked seafood have been reported in various US coastal regions.{{675 Su,Yi-Cheng 2007; }} *V. parahaemolyticus* is a leading cause of human gastroenteritis associated with seafood consumption in the US and an important seafood-borne pathogen worldwide.{{675 Su,Yi-Cheng 2007; }} The symptoms of *V. parahaemolyticus* infection include acute gastroenteritis and vomiting; severe cases can be fatal.{{692 Joseph,S.W. 1982;657 Iwamoto, Martha 2010; }} While there are harvesting and processing procedures designed to reduce the risk of *V. parahaemolyticus* in raw seafood such as oysters {{657 Iwamoto, Martha 2010; }}, these measures do not protect recreational bathers from water-column exposures to *Vibrio* species. Although rare, there have been documented cases of bloodstream infections resulting from the exposure of an open wound to seawater.{{657 Iwamoto, Martha 2010; }}

While *V. parahaemolyticus* has been of interest to the public health community for over 50 years, investigation into controlling environmental factors remains limited due to the complex nature of the marine system, the variety of *Vibrio* species and *V. parahaemolyticus* strains (both pathogenic and non-pathogenic types), and its wide distribution in habitats both tropical and temperate. Pioneering work by Rita Colwell and collaborators in the 1970s and early 1980s demonstrated the seasonal cycle *of V. parahaemolyticus* populations in Chesapeake Bay. {{692 Joseph,S.W. 1982; }} In humans, *Vibrio* infections also follow a seasonal pattern, with most cases corresponding with the warmer temperatures of summer and early fall.{{657 Iwamoto, Martha 2010; }} Due to the link between outbreaks and warmer seawaters temperature, climate change will present a new challenge to areas not previously considered at-risk for *V. parahaemolyticus* contamination. Notably, in recent years, outbreaks of infections due to pathogens usually associated with seafood harvested from warm waters were reported from more northerly areas of the USA that had not previously reported outbreaks. The first outbreak of *V. parahaemolyticus* infections in Alaska was reported in 2004, during a period when waters in the area experienced historically high temperatures.{{657 Iwamoto, Martha 2010; }} Surveillance for *Vibrio* infections (known collectively as vibriosis) began in 1988 in the US Gulf Coast states but did not become a nationally notifiable disease [to the Center for Disease Control and Surveillance, CDC] until January 2007.{{657 Iwamoto, Martha 2010; }}

It is known that populations of *Vibrio parahaemolyticus* are responsive to the supply of nutrients in the environment.{{682 Gode-Potratz, C.J. 2010;657 Iwamoto, Martha 2010; }} However, the state of current understanding was summarized in 2008 as "little information is available about the environmental variables governing the dynamics of *V. parahaemolyticus* populations..."{{677 Martinez-Urtaza, J. 2008; }} None of the literature reviewed to date has incorporated socio-economic/ land-based/ anthropogenic drivers as explanatory variables for the observed changes in *V. parahaemolyticus* range or population behavior.

Environmental determinants of *Vibro parahaemolyticus*	**Reference**
Salinity – surface	{{680 Cabrera-Garcia, M.E. 2004;679 Garcia, K. 2009; ; 676 Alam, M. 2009;677 Martinez-Urtaza, J. 2008; }}
Salinity – at 15m depth	{{677 Martinez-Urtaza, J. 2008; }}
Seawater Temp – surface	{{680 Cabrera-Garcia, M.E. 2004;679 Garcia, K. 2009; ; 675 Su,Yi-Cheng 2007;676 Alam, M. 2009; ; 677 Martinez-Urtaza, J. 2008;657 Iwamoto, Martha 2010; }}
pH of water	{{676 Alam, M. 2009; }}{{863 Hayat Mahmud, Z. 2006; }} {{692 Joseph,S.W. 1982; }}
Windspeed	{{677 Martinez-Urtaza, J. 2008; }}
Wind direction	{{677 Martinez-Urtaza, J. 2008; }}
Downwelling	{{677 Martinez-Urtaza, J. 2008; }}
Upwelling	{{677 Martinez-Urtaza, J. 2008; }}
Rainfall (influences salinity)	{{677 Martinez-Urtaza, J. 2008; }}
Calcium	{{682 Gode-Potratz, C.J. 2010; }}
Iron	{{682 Gode-Potratz, C.J. 2010; }}
Sediment suitability/presence	{{680 Cabrera-Garcia, M.E. 2004;675 Su,Yi-Cheng 2007; }}
Organic matter in suspension (turbidity)	{{680 Cabrera-Garcia, M.E. 2004; }}

Initial socio-economic determinants of *Vibrio parahaemolyticus* infection risk	**Reference**
Consumption of raw seafood, especially molluscan shellfish	{{657 Iwamoto, Martha 2010; }}
Exposure of open wounds to seawater	{{657 Iwamoto, Martha 2010; }}

ANTIBIOTICS

Antibacterial compounds are naturally manufactured and released by bacteria as they compete with each other for resources. Thus, the evolution of resistance to antibiotics is a natural process within the microbial world. However, the selection process for resistance has been sped up by the release of synthetic antibiotics. These synthetic antibiotics take the form of "medicines, their metabolites or degradation products" and reach the environment through wastewater, the use of antibiotic-laced manure in agriculture, or by pasture-reared animals excreting on land, followed by surface water run-off, driftage or leaching in deeper layers of the earth.{{752 Kemper, Nicole 2008; }} "The global market consumption of antibiotics was recently estimated at between 100,000 and 200,000 tons/year. Most of these compounds are not completely metabolized [within human or animal patients] and excreted residues can reach urban sewage treatment plants, where they can contaminate waste, surface and ground-water."{{757 Zuccato,Ettore 2010; }} Since human waste passes through treatment plants not designed to break down antibiotics, the sewage treatment plants themselves are considered major contributors to the spread of antibiotics in the environment.{{757 Zuccato,Ettore 2010; }} Sewage treatment plants can also act as point sources for the release of antibiotic resistant bacteria populations into the environment after resistant bacteria have evolved within and been excreted from human hosts.

Antibiotics as a class are chemically diverse, with different degradation pathways and rates. Some antibiotics persist a long time in the environment, especially in soil, while others degrade very fast.{{752 Kemper, Nicole 2008;750 Wright, Gerard D. 2010; }} There are about 250 different antibiotic entities registered for use in human and veterinary medicine, but usage patterns vary widely between countries.{{754 Kümmerer,Klaus 2009; }} On a global level, the ß-lactam antibiotics, which includes penicillins and related subgroups, make up the largest share of human use antibiotics. They account for approximately 50–70% of total antibiotic use.{{754 Kümmerer,Klaus 2009; }} Identifying the ecological effect of one single antibiotic would be challenging since wastewater streams contain a mixture of antibiotics/metabolites and populations of resistant bacteria. The effect of antibiotics or introduced resistant strains, and the change in risk for marine-sourced bacterial infections is unknown.

Antibiotics are typically found in the environment at sub-therapeutic concentrations, that is, strong enough to kill some but not all of the targeted bacteria, this situation promotes bacterial resistance. Once established, aquatic and soil ecosystems seem to act as reservoirs of antibiotic-resistant bacteria.{{757 Zuccato,Ettore 2010; }} Resistant organisms in water can represent the part of normal aquatic microbial populations. Alternatively, they can be the result of contamination by anthropogenic sources such as runoff from agriculture/aquaculture.{{750 Wright, Gerard D. 2010; }} Contamination from these sources is a significant problem that is increasing resistance in pathogens such as *Escherichia coli* and *Salmonella* in the environment.{{750 Wright, Gerard D. 2010; }}

Of parallel concern, but with much greater uncertainty, is the effect of antibiotics on the marine ecosystem, starting with primary producers such as blue-green algae but possibly expanding to other trophic levels. One specific concern is the effect of antibiotics released into the environment on Cyanobacteria, which account for more than 70% of the total phytoplankton mass and are considered sensitive to antimicrobial agents.{{749 Baquero,Fernando 2008; }} High mortality of Cyanobacteria could cause cascading effects up the food web. Further complicating the picture is the effects of antibiotics in the presence of other anthropogenic contaminants such metals. "Metal pollution indirectly aids in the persistence of antibiotic resistance in bacterial communities due to a combination of the stability of metals in terrestrial and aquatic environments and commonly occurring co- and cross-resistance to metal toxicity and antibiotics."{{785 Rose, J.M. 2009; }}

Environmental determinants of antibiotic burden	**Reference**
pH of water	{{754 Kümmerer,Klaus 2009; }}
Riverine discharge to an area (carrying wastewater from upstream)	{{754 Kümmerer,Klaus 2009;757 Zuccato,Ettore 2010; }}
Air temperature	{{754 Kümmerer,Klaus 2009; }}
Total light / Solar radiation	{{754 Kümmerer,Klaus 2009; }}

Initial socio-economic determinants of antibiotic burden	**References**
Level of wastewater treatment	{{754 Kümmerer,Klaus 2009;757 Zuccato,Ettore 2010; }}
Untreated wastewater release (Combined sewage overflows, storm drains, septic system leachate, etc)	{{754 Kümmerer,Klaus 2009;757 Zuccato,Ettore 2010; }}
Aquaculture, presence / absence	{{754 Kümmerer,Klaus 2009; }}
Level of antibiotic use in geographic measurement area (watershed, state, country, etc)	{{754 Kümmerer,Klaus 2009;750 Wright, Gerard D. 2010; }}
Level of non-medical antibiotic use	{{754 Kümmerer,Klaus 2009; }}

DOMOIC ACID

Domoic acid (DA) is a water-soluble excitotoxin produced by members of the *Pseudo-nitzschia* genus of diatoms in marine waters worldwide. When consumed by humans in high enough doses, DA can cause amnesiac shellfish poisoning (ASP). {{674 Pulido, O.M. 2008}} The symptoms of ASP range from moderate memory impairment and gastrointestinal upset to severe central nervous system toxicity, in certain cases DA poisoning has been fatal.{{674 Pulido, O.M. 2008; }} The hippocampus area of the brain has been shown to be highly sensitive to DA, but is only one of multiple potential targets.{{674 Pulido, O.M. 2008}} Domoic acid was first identified as the natural toxin responsible for outbreaks of ASP in shellfish consumers after the 1987 Canadian outbreak involving over 200 people. {{674 Pulido, O.M. 2008}} This outbreak, and identification of domoic acid as the causative agent, led to the establishment of safety standards for levels of DA in seafood. In the marine environment, DA has been linked to intoxication and mortality events in sea birds and sea lions.{{674 Pulido, O.M. 2008}}

Despite our understanding of the chemistry and cellular effects of domoic acid, our understanding of critical environmental factors that incite *Pseudo-nitzschia* to produce Domoic Acid remain limited. Most research into *Pseudo-nitzschia* bloom dynamics has been carried out in waters along the Pacific coast of North America. Nutrient runoff (rich in phosphates, nitrates, and silicates) has been proposed as a cause of *Pseudo-nitzschia* blooms on the Pacific coast of North America and in the Gulf of Mexico.{{730 Trainer, V.L. 2000}} Trainer et al. (2000) also suggested that nutrient inputs from upwelling, combined with wind transport of *Pseudo-nitzschia* cells, are important factors in contributing to harmful diatom concentrations.{{730 Trainer, V.L. 2000; }}{{777 Angus, T. 2011; }} Additionally, Ladizinsky (2003) has suggested that DA production in Monterey Bay, CA was “associated with excess copper in runoff from anthropogenic sources.”{{732 Ladizinsky, NL 2003}}{{777 Angus, T. 2011; }} Along the Atlantic coast, researchers have performed fieldwork to correlate DA production and bloom dynamics of *Pseudo-nitzschia* species in the Bay of Fundy with environmental conditions.{{764 Kaczmarska, I. 2007; }} The value of such research goes beyond an increased understanding of the natural world; the economic impacts of domoic acid can be significant for certain communities, shellfish areas have been closed multiple times due to unsafe levels of DA in the water.{{764 Kaczmarska, I. 2007; }}

Beyond the environmental factors that drive DA production by *Pesudo-nitzschia* species, there are evolving risk factors for human exposure due to the changing availability and consumption of seafood. Since DA biotransfers, but does not bioaccumulate, planktivorous sea creatures present a greater risk for domoic acid intoxication to consumers. Given the overfishing of larger carnivorous species, the greater availability/consumption of lower-trophic-level fish (e.g. sardines, anchoveta), and the increasing intake of seafood among both developed and rising-income countries {{88 Food and Agriculture Organization of the United Nations 2009; }}{{778 Food and Agriculture Organization of the United Nations (Fisheries and Aquaculture Department) 2010; }}{{777 Angus, T. 2011; }} the potential exists for more frequent (albeit low level) exposure to Domoic Acid. The use of wild-caught fish as feedstock for aquacultured products presents another, poorly characterized, risk vector for the introduction of Domoic Acid into foods destined for human consumption.{{777 Angus, T. 2011; }}

Environmental determinants of domoic acid	**References**
Nutrient Input	{{730 Trainer, V.L. 2000}}
Dissolved Inorganic Nitrogen	{{730 Trainer, V.L. 2000}}
Urea	{{731 Cochlan, W.P. 2008}}
Phosphorus	{{780 Pan, Y. 1996;781 Pan, Y. 1996; }}
Copper	{{732 Ladizinsky, NL 2003}}
Inorganic silicate depletion	{{733 Ramsdell, J.S. 2010}}, {{734 Anderson, C.R. 2006}}, {{735 Ragueneau, O. 2006}}
Iron	{{773 Wells, M.L. 2005; }}
Upwelling	{{782 Bates, S.S. 2006; }}{{767 Kudela, R. 2005; }}{{730 Trainer, V.L. 2000}}
Sea surface temperature	{{758 Chavez, F.P. 2003; }}
Epiphytic bacteria (growing on Pseudo-nitzschia)	{{779 Bates, S.S. 2004; }}{{762 Kaczmarska, I. 2005; }}

Initial Socio-economic determinants of domoic acid	**References**
Seafood Consumption, overall amount/frequency	{{778 Food and Agriculture Organization of the United Nations (Fisheries and Aquaculture Department) 2010; }}{{777 Angus, T. 2011; }}
Seafood Consumption, trophic level / species	{{778 Food and Agriculture Organization of the United Nations (Fisheries and Aquaculture Department) 2010; }}{{777 Angus, T. 2011; }}
Aquaculture	{{778 Food and Agriculture Organization of the United Nations (Fisheries and Aquaculture Department) 2010; }}{{777 Angus, T. 2011; }}
Land-use changes	{{777 Angus, T. 2011; }}

FECAL COLIFORMS

Fecal coliforms are a group of bacteria that live in the intestinal tract of humans and other mammals. These bacteria are shed by the millions in fecal material, and if ingested or inhaled can cause mild to severe illness.{{696 World Health Organization 1999; }} Waterbodies can become contaminated with fecal coliforms by point sources such as wastewater treatment facilities and combined sewer overflows, or by nonpoint sources such as leaky septic tanks, urban runoff, agricultural runoff, boat discharge, from bathers themselves, and from local animal populations.{{789 Halliday, E. 2011; }} Every year, bathing in coastal waters polluted with fecal contamination is estimated to cause more than 120 million cases of gastrointestinal illness and 50 million cases of respiratory disease around the world.{{694 Shuval 2003;}}

The association between recreational water contaminated with fecal coliforms and human illness from exposure to such water has been well recognized since at least the 1920s.{{696 World Health Organization 1999; }} However, improved detection technology that can quantify levels of other waterborne pathogens, including viruses and viable-but-not-culturable (VBNC) bacteria, has led to new concerns about the adequacy of current safety levels. When newer non-traditional results are compared to standard fecal coliform monitoring results they have raised questions about suitability of fecal coliforms as the sole indicator organism for the health risk of recreational waters.{{696 World Health Organization 1999; }} In general, "coliforms cannot be used to monitor nonfecal contamination or to indicate the presence of pathogens." {{792 Gonzalez, Alessandra M. 2010; }} What is needed is a more comprehensive understanding of the interplay between anthropogenic releases, environmental conditions, and exposure risk. As one researcher noted, "the continuing practice of implementing [monitoring for] fecal indicator organisms without understanding their persistence and survivability in the environment has hindered the ability to determine their significance in water and to accurately assess human health risks."{{794 Ferguson, D. 2011; }} An understanding of bacterial survival determinants is important for accurately assessing public health risks because "fecal indicator bacteria include strains that may survive in a particular setting for a period of time after introduction (persistence) and, perhaps more important, to grow, replicate, and adapt in nonhost environments (naturalization) including water, soil, and vegetation."{{794 Ferguson, D. 2011; }}

Fecal coliforms in the water column continue to be one of the most widely collected biological indicators, and these data sets are not without utility. In fact, a "meta-analysis of twenty-two epidemiological studies conducted from 1953 - 1996 at beaches around the world suggests a causal dose-related relationship between gastrointestinal symptoms and recreational water quality as measured by bacterial indicator counts. Among these studies, *Enterococcus* spp. emerged as the indicator bacteria best correlated with health outcomes in marine systems, whereas *E. coli* was best correlated with health outcomes in fresh water systems." {{795 Prüss, Annette 1998;}}{{789 Halliday, E. 2011; }} This suggests that monitoring for fecal coliforms (including *E. coli*) will continue to play a role in public health assessments of water quality.

While fecal coliforms have clear association with sewage releases, sands and sediments at estuarine beaches and coastal wetlands are another potential source of the fecal coliforms found in recreational waters. Particulate matter, to which fecal coliforms may attach, naturally settles out in these environments and may be resuspended during tidal or high flow conditions.{{789 Halliday, E. 2011; }} Thus, in addition to exposures from contact with the water column there is increasing interest in the presence of disease-causing organisms in beach sand. "The first beach sand epidemiological study {{796

Heaney, C.D. 2009;}} showed that "sand contact activities", including digging in sand or being buried in sand, were positively associated with enteric illness."{{789 Halliday, E. 2011; }}

Environmental determinants of fecal coliforms	**Reference**
Salinity	{{696 World Health Organization 1999; }}
Sea water temperature at surface (SST)	{{696 World Health Organization 1999; }}
pH of water	{{696 World Health Organization 1999; }}
Wind speed	{{696 World Health Organization 1999; }}
Wind direction	{{696 World Health Organization 1999; }}
Rainfall	{{696 World Health Organization 1999; }}
Riverine discharge to area	{{696 World Health Organization 1999; }}
Organic matter in suspension (turbidity)	{{696 World Health Organization 1999; }}
Air temperature	{{696 World Health Organization 1999; }}
Current direction and speed, fresh and estuarine waters	{{696 World Health Organization 1999; }}
wave height	{{696 World Health Organization 1999; }}
Total light or radiation	{{696 World Health Organization 1999; }}
Tidal state and magnitude	{{696 World Health Organization 1999; }}
Animal population, presence of horses, donkeys, dogs, shore birds (recommend hourly observation)	{{696 World Health Organization 1999; }}
Release of bacteria from beach sand 'reservoir'	{{790 Shah, A.H. 2011; }}

Initial socio-economic determinants of fecal coliforms	**References**
Storm drains (presence, abundance)	{{696 World Health Organization 1999; }}
Combined Sewer Overflows (presence, volume)	{{696 World Health Organization 1999; }}
Bather population at each transect point (recommend hourly observation)	{{696 World Health Organization 1999; }}
Beach debris and sanitation: sanitary plastics, visible grease balls, algae (recommend daily observation)	{{696 World Health Organization 1999; }}
Location of bather facilities (showers, lavatories) and relevance of input from these sources to beach	{{696 World Health Organization 1999; }}
Boats anchored or moored within 1 km of beach	{{696 World Health Organization 1999; }}