8-2006

NHEP Data Management Plan

Phil Trowbridge
New Hampshire Department of Environmental Services

2006

New Hampshire Estuaries Project

Data Management Plan

Prepared by:

Phil Trowbridge
NHEP Coastal Scientist
New Hampshire Department of
Environmental Services

August 2006

Introduction

A goal of the New Hampshire Estuaries Project (NHEP) and its monitoring program is to promote a cooperative effort by all agencies and organizations who participate in monitoring activities, in order to maximize the usefulness of current monitoring efforts and available data. To achieve this goal, it is necessary to effectively manage the large volume of existing information as well as new information that will be developed through the NHEP monitoring program.

Data and information about NH's estuaries now exists in multiple formats within a variety of organizations. Existing monitoring programs are designed to meet the missions of the various implementing organizations. The organizations use different procedures and protocols for data collection, analysis, storage, and reporting. Coordination of data management among organizations is currently limited.

This Data Management Plan contains protocols for data reporting to the NHEP to facilitate data integration. Different protocols will be applied to different types of data (e.g., chemical, geospatial, and biological). The protocols will be considered contract requirements for NHEP monitoring programs and recommended guidelines for other partners. This plan also includes protocols for conducting quality assurance tests on water quality data to ensure the integrity of the NHEP indicators.

Protocols for Data Reports to the NHEP

Chemical

For all data on chemical concentrations in water, sediment, soil, and tissue, the NHEP's goal is to integrate the data into a centralized database at the NH Department of Environmental Services. The NHDES Environmental Measurement Database contains all NHDES data plus data from Great Bay Coast Watch and a growing list of other NH monitoring organizations. This database is accessible via the internet at http://www.des.state.nh.us/OneStop/. The NHEP believes that compiling data in the Environmental Measurement Database will save NHEP staff time for State of the Estuaries reports and will make the data accessible to other researchers.

Georeferencing — For each station in the datatable, the following information should be provided at a minimum:

1. A unique "StationID", which is an alphanumeric combination of 15 or less characters
2. The station type (Estuary, River, Lake, Pipe, etc.).
3. The latitude and longitude of the station (DD MM SS format).
4. The town in which the station in located.

5. The method used to determine the latitude and longitude (dGPS, GPS, map interpolation, etc.).
6. The datum used to determine the latitude and longitude (NAD27, NAD83, etc.)

The station information form provided in Appendix A can (but does not have to) be used to report this and additional information about the station.

These station details are not needed for long-term Great Bay National Estuarine Research Reserve (GBNERR) monitoring stations in Great Bay and the Piscataqua River. However, the following station naming convention from the GBNERR program should be used.

Location	StationID
Adams Point	GRBAP
Chapman's Landing	GRBCL
Squamscott RR Bridge Sonde	GRBSQ
Lamprey River Sonde	GRBLR
Oyster River Sonde	GRBOR
Central Great Bay Sonde	GRBGB
Coastal Marine Lab Pier	GRBCML

Format

Data should be provided in Microsoft Excel spreadsheets or comma delimited text files. Data for the concentrations of chemicals in water, sediment, or soil should be in a format compatible with the DES Environmental Measurement Database. This database uses a "one result per row" format. Therefore, the spreadsheets should have the following columns at a minimum. An example table is provided in Appendix B.

Column Name	**Description**
StationID	Station identifier
Category	The category of the activity (routine, replicate, etc.).
Medium	Sample medium (e.g., water, sediment, soil)
Date	Date the activity began, usually the date the sample was taken.
Time	Time the activity began, usually the time the sample was taken.
Personnel	Person(s) conducting the activity.
Depth	Depth to activity.
DepthUnits	Units for depth to activity.
Parameter	Name of parameter that was analyzed (e.g., Dissolved Oxygen)
ResultNumeric	Numeric results for the parameter.

ResultQualifier	Qualifier for the results (example: <, >, >= , ND, U, J, etc.)
Units	Units for the results.

Content

Data provided to the NHEP should have undergone quality assurance checks by the principal investigator and be considered final. Data that do not meet data quality objectives from quality assurance project plans or standard operating procedures should be excluded from the dataset. Field duplicate samples should be included in the dataset but laboratory duplicates should not.

Documentation (metadata)

All laboratory results should be accompanied by the name of the laboratory and the analytical method used. The analytical methods should be a reference to a Standard Methods number, an EPA method number, or some other citation. A quality assurance project plan or standard operating procedure can be provided to supply this information. If the laboratory or the method for a parameter is not the same for the whole dataset, then the metadata should make it clear which laboratory and method was used for each result.

Geospatial

The NH Estuaries Project requests that all contractors engaged in geospatial data development activities conform to a set of basic standards governing data structure, format, and documentation. These standards, defined by NH GRANIT, will ensure that all data may be utilized by GIS users in the state and the region.

For further information about GRANIT, the statewide GIS clearinghouse, please see www.granit.sr.unh.edu.

Georeferencing

All data should be referenced to New Hampshire State Plane feet, North American Datum (NAD) 83.

Format

The preferred formats for data submission are those directly readable by ESRI software, including shapefiles (*.shp), export files (*.e00), and geodatabases (*.mdb). Other acceptable formats include Autocad drawing files (*.dwg), Autocad exchange files (*.dxf), and Microstation design files (*.dgn).

If you are unable to provide data in any of the above, please email the GRANIT database manager (granit.sr.unh.edu) to inquire about other options.

Content	Please ensure the spatial integrity of all vector polygon data, including closure of all polygons, absence of sliver polygons, absence of dangling arcs, etc.
Documentation (metadata)	Each data set must be accompanied by a comprehensive metadata record that conforms to the Federal Geographic Data Committee (FGDC) "Content Standard for Digital Geospatial Metadata" (FGDC-STD-001-1998), June, 1998. For further information on this standard, see www.fgdc.gov/metadata/metadata.html Many software packages provide tools for the development of FGDC-compliant records. If you do not have access to an appropriate tool, or would like to see an example of a completed metadata record, please email the GRANIT database manager (granit.sr.unh.edu) for assistance.

Biological

NHEP uses a variety of biological data to calculate environmental indicators for State of the Estuaries reports. For example, shellfish standing stock estimates are calculated from oyster and clam quadrat density data. Biological data will not be compiled in a centralized database because the datasets are often so different.

Georeferencing

For each station in the datatable, the following information should be provided at a minimum:

1. A unique "StationID", which is an alphanumeric combination of 15 or less characters.
2. The station type (Estuary, River, Lake, Pipe, etc.).
3. The latitude and longitude of the station (DD MM SS format).
4. The town in which the station is located.
5. The method used to determine the latitude and longitude (dGPS, GPS, map interpolation, etc.).
6. The datum used to determine the latitude and longitude (NAD27, NAD83, etc.)

The station information form provided in Appendix A can (but does not have to) be used to report this and additional information about the station.

These station details are not needed for data that are reported for major features such as the Nannie Island oyster bed or the Middle Ground clam flat. Instead, these data can just be reported for the name of the feature.

Format	Data should be provided in Microsoft Excel spreadsheets or comma delimited text files.
Content	Data provided to the NHEP should have undergone quality assurance checks by the principal investigator and be considered final. Data that do not meet data quality objectives from quality assurance project plans or standard operating procedures should be excluded from the dataset.
Documentation (metadata)	All results should be accompanied by either a quality assurance project plan or a standard operating procedure that document the methods used to generate the data.

Protocols for Data Quality Assurance Tests

Water Chemistry Data

Water chemistry data provided by laboratories should be quality assured using the steps listed below.

1. Check that data has appropriate metadata from the laboratory
   - Analytical methods used by the laboratory
   - Units for data
   - Name and contact information for the laboratory
   - Results of quality control tests (e.g., lab duplicates, matrix spike duplicates, continuing calibration checks, analysis of standard reference materials)

2. Censor data values below detection
   - Get the method detection limit for the parameter from the laboratory.
   - If the laboratory reported any values that are less than the reporting detection limit (RDL), then these values should be considered below detection and replaced by the RDL with a "U" qualifier. The RDL is the lowest calibration standard used for the test. For any values that are reported between the RDL and the method detection limit (MDL, if available), the results should have a "J" qualifier.

3. Calculate differences between field duplicate and field replicate samples
   - Compute the absolute value of the difference between the two samples.
   - Compute the relative percent difference between the two samples (absolute difference between the samples divided by the average of the two samples).

4. Compare the absolute differences and relative percent differences to the data quality objectives (listed below)

PARAMETER	Typical RDL	Absolute Difference DQO	RPD DQO
WATER TEMPERATURE	NA	1 degC	30%
SALINITY	NA	1 ppt	30%
DISSOLVED OXYGEN	NA	0.5 mg/L	30%
DISSOLVED OXYGEN SATURATION	NA	5 %	30%
TOTAL FECAL COLIFORM	1 cts/100ml	10 cts/100ml	30%
ENTEROCOCCUS	1 cts/100ml	10 cts/100ml	30%
ESCHERICHIA COLI	1 cts/100ml	10 cts/100ml	30%
CHLOROPHYLL A, CORRECTED FOR PHEOPHYTIN	0.2 mg/L	5 mg/L	30%
PHEOPHYTIN-A	0.2 mg/L	5 mg/L	30%
SOLIDS, SUSPENDED	1 mg/L	10 mg/L	30%
CARBON, SUSPENDED	0.125 mg/L	1 mg/L	30%

PARAMETER	Typical RDL	Absolute Difference DQO	RPD DQO
NITROGEN, AMMONIA AS N	0.005 mg/L	0.05 mg/L	30%
NITROGEN, NITRITE (NO2) AS N	0.005 mg/L	0.05 mg/L	30%
NITROGEN, NITRITE (NO2) + NITRATE (NO3) AS N	0.005 mg/L	0.10 mg/L	30%
NITROGEN, DISSOLVED	0.10 mg/L	0.25 mg/L	30%
NITROGEN, SUSPENDED	0.025 mg/L	0.10 mg/L	30%
PHOSPHORUS, ORTHOPHOSPHATE AS P	0.005 mg/L	0.025 mg/L	30%
PHOSPHORUS, DISSOLVED	0.025 mg/L	0.025 mg/L	30%
SILICA AS SIO2	0.1 mg/L	2 mg/L	30%

** The absolute difference DQOs were developed by reviewing the median value for each parameter in estuarine samples in NH and by reviewing the standard error observed in duplicate samples collected by the GBNERR monitoring programs in 2002-2004. The value was selected such that the error would be small compared to the overall data set (i.e., less than the median value) but was not so small that it was unrealistic to achieve (i.e., greater than 2 standard deviations of the absolute differences).*

5. Disqualify replicate pairs that fail the data quality objectives
   - If all of the station visits in the dataset have replicates, and if a pair of replicate samples fails both data quality tests, then both replicate samples should be rejected and removed from the database. If the pair only fails one of the data quality tests, then the pair will be retained.
   - If only a few samples in the database are replicated (e.g., 10% of station visits), then the pairs of replicate samples that fail both data quality tests should be reviewed for systematic errors. For example, the failed replicate pairs should be grouped by parameter and by station visit. If either a parameter or a station visit appears to have systematic data quality problems, then all of the data for that parameter or station visit should be rejected and removed from the database. However, if the failed replicate pairs occur randomly in the database, only the failed replicate pairs should be rejected and removed from the database.

6. Calculate summary statistics and box plots for each laboratory parameter to identify outliers
   - Use box plots or histograms to identify anomalous points in the dataset.
   - Compare summary statistics of the dataset to measurements made in the same waterbody or similar waterbodies to identify systematic errors.

7. Confirm tide stage assignments for samples
   - If the sampling design calls for samples to be collected a low and high tide, the tide stat assignment should be verified. The actual time of the high or low tide at the station should be compared to sample collection time to determine if they sample was collected at the right time. If the sample was collected between 3 hours before to 1 hour after the tide, then the sample time will be considered to be correct. If the sample was collected outside of this window, then it will not be associated with a tide stage.

Water Quality Data from In-Situ Datasondes

Water quality data from in-situ datasondes should be quality assured using the protocol listed below.

Introduction
Great Bay National Estuarine Research Reserve (GBNERR) and the University of New Hampshire (UNH) deploy datasondes throughout the Great Bay Estuary to monitor water quality during the ice-free season. The New Hampshire Estuaries Project and the Department of Environmental Services (NH DES) use measurements from the datasondes to determine whether water quality standards are being met in Great Bay for the State of the Estuaries Report and the Section 305(b) Surface Water Quality Assessments, respectively. A violation of water quality standards has implications for point source discharges, municipalities, and other sources of pollutants to the water body. Therefore, the datasonde data must pass certain quality assurance protocols.

GBNERR and UNH review the original data files and remove questionable data. Data and metadata for most of the deployments are available at http://cdmo.baruch.sc.edu/. The quality assurance process described in this protocol is only relevant for the stated objectives. The limitations placed on the data by these criteria do not restrict the use of the data for other purposes.

Assumptions
1. The generic metadata for the dissolved oxygen probes on the GBNERR/UNH sondes states that, "*The reliability of the dissolved oxygen (DO) data after 96 hours post-deployment for non-EDS (Extended Deployment System) data sondes may be problematic due to fouling which forms on the DO probe membrane during some deployments*." Therefore, DO measurements within the first 96 hours of the deployment will be presumed to be accurate unless proven otherwise by quality control (QC) measurements by another calibrated sensor. In contrast, DO measurements taken more than 96 hours post-deployment will only be considered useable for State of the Estuaries and 305b purposes if an end-of-deployment QC measurement proves that the sonde did not experience drift over the duration of the deployment.
2. Measurements of DO saturation with a calibrated YSI-85 or similar unit at the station at the same depth as the sonde will be considered to be a QC measurement. QC measurements should be completed at the beginning and the end of each deployment. When one sonde is being replaced by another within an hour, then one DO measurement can serve as the end-of-deployment measurement for one sonde and the beginning of deployment measurement for the other.

3. Dissolved oxygen readings from sonde and the QC measurements will be considered to "agree" if the absolute difference is less than or equal to 20 %sat.
4. The "beginning-of-deployment" sonde reading will be the average of the three sonde readings during the first hour of the deployment. The "end-of-deployment" sonde reading will be the average of the three sonde readings during the last hour of the deployment.
5. For all other parameters besides dissolved oxygen, the results retained in the datafile by the GBNERR or UNH project managers will be accepted as valid for State of the Estuaries and 305(b) purposes.

Quality Assurance Criteria and Process

Step 1: Based on the assumptions listed above, the DO data for each deployment will be evaluated using the QC measurements. The DO measurements in the deployment will determined to be acceptable for State of the Estuaries and 305(b) purposes according to the matrix in Table 1.

Table 1: Dissolved oxygen records in each deployment to be used for State of the Estuaries and 305(b) purposes based on the results of QC tests

		Post-Deployment QC Test Result Compared to End-of-Deployment Sonde Reading		
		Results Agree	*Results Disagree*	*Missing Data*
Pre-Deployment QC Test Result Compared to Beginning-of-Deployment Sonde Reading	*Results Agree*	All	First 96 hours	First 96 hours
	Results Disagree	None	None	None
	Missing Data	All	First 96 hours	First 96 hours

Step 2: The time series of DO (as %sat) will be plotted for each deployment to verify that the classifications from Step 1 are justified. If the DO data from a deployment passed QC tests in Step 1 but had obvious errors based on the plot, then DES may decide to reject the data from this deployment. Likewise, if there is a good explanation for why data from a deployment failed QC tests, then the NHEP and DES may decide to include the data from this deployment. Determinations of this sort should be documented in a memo.

Step 3: DO results that are determined to not be useful for State of the Estuaries and 305(b) purposes will be marked with a "N" in the ResultsValid field for DO in

the deployment datafile and then uploaded to the NH DES Environmental Measurement Database.

Step 4: A quality assurance memo will be prepared summarizing the determinations from this process.

Appendices

Appendix A: Sampling Station Identification Form

Appendix B: Example Table

Sampling Station Identification Form

Note: Shaded items are ultimately required.

Form Completed By:

Project

Station ID (15 char max) _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

Alias ID

Station Name

Transect

Town (no village names)

State (circle one)
NH ME
MA VT
Canada

Date Established ___/___/___

Station Type (circle one)

Air - Ambient	Culvert	Landfill	Soil Boring	Wetland - Estuarine, emergent
Air - Indoor	Drain Manhole	Land Runoff	Spring	Wetland - Estuarine, forested
Canal - Drainage	Estuary	Mine/Mine Discharge	Storm Sewer	Wetland - Estuarine, scrub-shrub
Canal - Irrigation	Facility - Industrial	Ocean	Tidal Swale	Wetland - Lacustrine, emergent
Canal - Transport	Facility - Municipal Sewage (POTW)	Pipe	Waste Pit	Wetland - Palustrine, emergent
Catch Basin	Facility - Other/combined	Reservoir	Waste Sewer	Wetland - Palustrine, forested
Channelized Stream	Facility - Privately owned non-industrial	River/Stream	Well	Wetland - Palustrine, moss-lichen
Combined Sewer	Lake	Riverine Impoundment		Wetland - Palustrine, scrub - shrub
Constructed Wetland	Land	Seep		Wetland - Riverine, emergent

Waterbody Name

Designated River Reach (list on other side)

Related Lake

Final Discharge Location (Used by Watershed Assistance)

Total Station Water Depth

Units (Circle one) in/ft/ cm/m

Station Description:

Directions to Station:

Date Located: ___/___/___

If located by GPS:

Latitude (Ex:DD MM SS.SS)

Longitude

GPS File Name

GPS Unit/Serial # (list on other side)

Corrected? Yes No

Locational comments:

If located by other method:

Method of Location (circle or enter):
Interpolation - Map
Interpolation - Photo
Interpolation - Satellite
Land-Survey
Other: _____

Map Scale (circle or enter)
1:24,000/25,000
1:100,000
Other: _____

Datum (circle or enter)
NAD 1927
NAD 1983
WGS 1984
Other:
_____

Elevation Information:

Elevation

Units ft/m

Method (Circle one)
Map Interpolation Digital (DEMs)
Differential Mode GPS
Absolute Mode GPS
Conventional Survey
Public Land Survey
Altimeter

Datum (circle or enter)
NGVDD 1929
NAVD 1988
WGS 1984
Local Tidal Datum
Mean Sea Level
Other:
_____

Site Diagram (or attach map with location marked)

Designated River Segments:

Ashuelot
Cold
Connecticut
Contoocook/North Branch
Exeter
Isinglass
Lamprey
Lower Merrimack
Pemigewassat
Piscataquog
Saco
Souhegan
Swift
Upper Merrimack

GPS Units:

Make	Model	Serial#	Section
Garmin	GPS III	40157743	Biomonitoring
Garmin	GPS III Plus	92186038	Watershed Assistance
Garmin	GPS III Plus	92177955	Water Quality
Magellan	320	23857	Shellfish
Trimble	GeoExplorer II	0010004LQ8	Biology
Trimble	GeoExplorer II	0010004LQ2	Biology
Trimble	GeoExplorer III	23970	Watershed Assistance
Trimble	ProXL	3450A00313	Data Management

APPENDIX B: EXAMPLE TABLE FORMAT FOR DATA REPORTS TO THE NH ESTUARIES PROJECT

StationID	Category	Medium	Date	Time	Personnel	Depth	DepthUnits	Parameter	ResultQualifier	ResultNumeric	Units
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	DISSOLVED OXYGEN		7.6	MG/L
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	ENTEROCOCCUS		65.5	#/100ML
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	ESCHERICHIA COLI		9.5	#/100ml
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	NITROGEN, AMMONIA AS N		0.018	mg/L
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	NITROGEN, NITRATE (NO3) + NITRITE (NO2) AS N		0.031	mg/L
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	NITROGEN, NITRATE (NO3) AS N		0.031	mg/L
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	NITROGEN, NITRITE (NO2) AS N	<	0.001	mg/L
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	PH		8	UNITS
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	PHOSPHORUS, ORTHOPHOSPHATE AS P		0.016	mg/L
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	SALINITY		29.7	PPT
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	SILICATE		0.276	mg/L
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	SOLIDS, TOTAL SUSPENDED (TSS)		7.5	mg/L
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	TEMPERATURE WATER		20	DEGC
ME02-0260A	ROUTINE	WATER	7/22/2002	13:15	J. DOE	0.5	M	TOTAL FECAL COLIFORM		12	#/100ml