TASK 4: DATA MANAGEMENT

Contents

Exhibits

TASK 4: DATA MANAGEMENT

Introduction

This task involves the management of water quality monitoring data and the preparation and submittal of formatted data. Basin Planning Agency data validation and verification are described in Task 2.

Efforts for the FY 2008-2009 biennium are focused on:

- Adopting data file formats and content required by the TCEQ's new water quality database, the Surface Water Quality Monitoring Information System (SWQMIS)
- Validating data to the level required by SWQMIS and the *SWQM Procedures Manual* (www.tceq.state.tx.us/compliance/monitoring/water/quality/data/wqm/mtr/swqm_procedures.html)
- Using SWQMIS's web-enabled Station Location Request (SLOC) functionality
- Adopting new TCEQ Program Codes.

Data Management

Basin Planning Agency Data Management

The Planning Agency Data Manager is responsible for preparing data sets of quality-assured data in the format contained in Exhibit 4A for submittal to the TCEQ. Data sets should not contain data from more than one QAPP.

The Planning Agency Data Manager is also responsible for preparing the Data Summary (Exhibit 4B) which contains basic identifying information about the data set and comments regarding inconsistencies and errors identified during data verification and validation steps or problems with data collection efforts.

Coding Outliers

The TCEQ Surface Water Quality Monitoring (SWQM) Team establishes a minimum and maximum value for each parameter above or below which a value is considered an "outlier" by the TCEQ. These values may represent a judgment of reasonableness by SWQM staff, or they may represent a certain percentile of historical data (e.g. 5^{th} and 95^{th}). Reported values that are found to be outliers should be checked against field and laboratory records to verify the correctness of the analysis as described in Task 2. The Planning Agency Data Manager should ensure that these outliers are flagged in the data set to show that they have been confirmed. If an outlier is not flagged, the SWQMIS data loader will find the anomaly and will not accept the data. All outliers must be flagged in the *Results* file by the inclusion of a "1" in the last field of the record (*Verify_flg*).

It is strongly recommended that Basin Planning Agencies develop an automated program to compare result values to the min/max levels. Because the min/max thresholds are subject to change, a file (*sw_parm*) containing all parameters and their min/max levels is updated weekly and made available for download (see Web Page Resources for Task 4). Planning Agencies should ensure they have the most recent version of *sw_parm* when preparing data for submittal (www.tceq.state.tx.us/compliance/monitoring/crp/data/crp-resources.html).

TCEQ Automated Data Review and Validations

The TCEQ performs a review of the data that encompasses both formatting, certain data validation, and error checks. When a data set is received by the Clean Rivers Program (CRP) Data Manager, a thorough check is performed to ensure the data format is compatible with SWQMIS requirements and any data anomalies are located and addressed before uploading the data to SWQMIS. A complete list of the validations performed by SWQMIS is listed in Exhibit 4C. It is important to note that SWQMIS will perform certain automatic validations including rounding values to the prescribed number of significant figures. Other validations may require correction by the Planning Agency, such as reporting water temperature to the nearest tenths. SWQMIS will never add a significant figure to a reported value (e.g. reporting 30 would be rejected until it is changed to 30.0, reporting 30.01 would auto-correct to 30.0).

CRP Project Manager Data Review

In addition to the initial automated review, a report is prepared for CRP Project Manager review as a check on contract requirements and data validity. This report contains:

1. Date range
2. *Tag_id* range
3. Count of records in the data set
4. *Storetcodes* submitted with the data set and the number of times each was reported
5. Minimum and maximum values submitted for each *Storetcode*
6. *StationIDs* for which data was submitted
7. *Source1*/*Source2*/*Program* code combinations in the data set
8. Outliers in the data set and their *Remark* field
9. A comparison of data reported against historical statistics at each station.

After review of the report, a list of errors and/or discrepancies is provided to the Basin Planning Agency for correction and, if necessary, resubmittal to the CRP Project Manager. When all errors and discrepancies have been reconciled, the data set is uploaded to SWQMIS.

To expedite data approval, it is recommended that the Basin Planning Agency develop similar automated review processes.

Data Correction Request Forms

If the Basin Planning Agency finds that water quality monitoring data are in error in its database, this fact should be communicated to the TCEQ so that the same corrections are made in SWQMIS. A SWQM Data Correction Request Form should be used to specify the applicable corrections. The forms should be submitted electronically to the CRP Project Manager. The form can be obtained from the Internet at (www.tceq.state.tx.us/assets/public/compliance/monops/water/wdma/dmrg/2005/2005dmrg_ch9.pdf). If a large number of errors or systematic errors are found which make use of the form unreasonable, contact your CRP Project Manager for alternate electronic reporting methods.

Using Source Codes, Program Codes, and Tag_id Prefixes

The *Tag_id* is the seven-digit alphanumeric that is unique to each sampling event and that links the measurement values to the sampling event. The *Tag_id* prefix is the unique letter code added to the beginning of the *Tag_id*. Each basin Planning Agency has a different *Tag_id* prefix which is inserted as the first one or two letters of the *Tag_id*.

Source1 codes identify the agency responsible for the QAPP, *Source2* codes identify the organization conducting the monitoring in the field, and *Program* codes identify the type of monitoring under which

the reported data was collected. For instance, "SR|LW|RT" would signify that the data was submitted under the Sabine River Authority (SR) QAPP and collected by the City of Longview (LW) without targeting any certain environmental condition (RT). The *Tag_id* prefix would be "J" which is assigned to SRA.

Requests for new *Source1*, *Source2*, or *Tag_id* prefixes should be coordinated with the CRP Project Manager. Lists of valid codes can be found in the *2006 Surface Water Quality Monitoring Data Management Reference Guide (DMRG)* (www.tceq.state.tx.us/compliance/monitoring/water/quality/data/wdma/dmrg_index.html).

The TCEQ has adopted a new set of Program Codes that reflect whether a sample was targeted at a certain environmental condition. Please refer to the DMRG for details.

Composite Samples

Composite samples require entries in several additional fields in the *Events* file (see Exhibit 4A). These fields are *Startdate*, *Starttime*, *Startdepth*, *Category*, and *Type*. *Category* must be one of four codes: *T* for time composites, *S* for space composites, *B* for both space and time composites, or *F* for flow-weighted composites. The *Type* field must be a two-digit number (including leading zeros, if necessary) indicating the number of grabs, *CN* for continuous, or *GB* when the number of grabs is unknown.

Data Submittal to the TCEQ

Water quality data collected during the biennium will be submitted to the TCEQ in pipe-delimited ASCII text files in TCEQ format (Exhibit 4A) as described in the workplan.

All data should be sent directly to the CRP Project Manager. Each data submittal must be accompanied by a Data Summary (Exhibit 4B) which explains data discrepancies and issues.

If the data have any errors upon review by the TCEQ, an explanation of the errors will be provided to the Basin Planning Agency. the data submittal is not considered a complete deliverable until all errors have been resolved and the dataset approved for upload.

The Basin Planning Agency should make current data available to the public. Newly available data should be added to the web at least twice annually. The Basin Planning Agency may choose to provide a link to the TCEQ water quality data viewer to satisfy this deliverable.

Biological Electronic Data Reporting

Electronic reporting of biological data represents a special case in that the sum of the sampling should be reported as separate events: field, benthics, nekton, and habitat. Each of these sampling events have unique *Endtime* and *Starttime* information and different composite information. In order to relate the events to each other, a comment should be included in the *Comment* field of each event which references the *Tag_id*(s) of the other events (e.g. "see X023847 for habitat data, X023848 for nekton data").

Requests for Station Identification Numbers (StationID)

As part of SWQMIS, requests for the creation of new monitoring stations will be handled via the internet. Specific instructions are included in the DMRG (see web address above).

CRP Data Management Training

The TCEQ may conduct data management training workshops, as needed, in response to new data management procedures or requirements. These workshops will typically be held in conjunction with other CRP training.

Exhibit 4A

Event and Result File Structure

EXHIBIT 4A
EVENT AND RESULT FILE STRUCTURE

Event File

Field Name	Type	Length	Required	Description/Format
Tag_id	A	7	Y	Alphanumeric with proper Tag_id prefix.
StationID	A	5	Y	TCEQ station identifier.
Enddate	A	10	Y	MM/DD/YYYY
Endtime	A	5	Y	HH:MM - 24 hour clock
Enddepth	A	varies	Y	Sample depth in meters.
Startdate	A	10	composite	MM/DD/YYYY - for composite samples.
Starttime	A	5	composite	HH:MM - for composite samples.
Startdepth	A	varies	composite	Start depth for composite samples.
Category	A	1	composite	T - time composite S - space composite B - time and space composite F - flow weighted composite
Type	A	2	composite	CN - continuous GB - grab where number of grabs is unknown ## - number of grabs (must be two digits)
Comment	A	varies	N	Avoid using delimiter characters (such as commas) in text.
Source1	A	2	Y	Entity under whose QAPP the data is being reported.
Source2	A	2	Y	Entity who did actual sampling.
Program	A	2	Y	Type of sampling.

Result File

Field Name	Type	Length	Required	Description/Format
Tag_id	A	7	Y	Relates measurement results to their sampling event.
Enddate	A	10	Y	MM/DD/YYYY - must match event enddate.
Storetcode	A	5	Y	##### - 5 digits with leading zeros
Gtlt	A	1	N	“<“ or “>”
Value	A	varies	Y	Using alpha type prevents the fixing of values to a predetermined number of decimal places.
LOD	A	varies	N	Used to report the Limit of Detection when reporting values below the Limit of Quantitation, otherwise leave blank.
LOQ	A	varies	N	Used to report the Limit of Quantitation when reporting values below the Limit of Quantitation, otherwise leave blank.
Remark	A	1	N	Refer to DMRG for allowable values and use.
Verify_flg	A	1	N	“1" or blank - used to indicate that a value is an outlier and has been confirmed as correct.

Exhibit 4B

Data Summary

EXHIBIT 4B
DATA SUMMARY

Data Information

Data Source: _____.

Date Submitted: _____.

Tag_id Range: _____.

Date Range: _____.

Comments

Please explain in the space below any data discrepancies including:

- Inconsistencies with AWRL specifications or Reporting Limits;
- Failures in sampling methods and/or laboratory procedures that resulted in data that could not be reported to the TCEQ; and
- Other discrepancies.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Planning Agency Data Manager: _____.

Date: _____.

Exhibit 4C

SWQMIS Validations

The following are the initial validations performed by SWQMIS:

Check for correct file formats as specified in the DMRG.
Check for validity of Source Codes and Program Codes.
Check for validity of StationIDs.
Check for validity of TagID prefixes.
Check the Result file for duplicate Parameter Codes on unique TagIDs.
Check the Event file for duplicate TagIDs.

The following are the validation rules and SWQMIS's automated behavior:

Rule	SWQMIS Action & Examples	Notes
Water temperature (00010) must be reported to the nearest tenth of a degree.	Auto-correction 30.11 corrects to 30.1 28.55 corrects to 28.6	Whole numbers will require a manual correction to add the tenths digit. SWQMIS will not add a significant figure to a reported value (e.g., 30 will require manual correction to 30.0).
pH (00400) must be reported to the nearest tenth of a pH standard unit.	Auto-correction 7.22 corrects to 7.2 6.88 corrects to 6.9	Whole numbers will require a manual correction to add the tenths digit. SWQMIS will not add a significant figure to a reported value (e.g., 7 will require manual correction to 7.0).
Dissolved oxygen (00300) must be reported to the nearest tenth of a mg/L.	Auto-correction 6.33 corrects to 6.3 4.19 corrects to 4.2	Whole numbers will require a manual correction to add the tenths digit. SWQMIS will not add a significant figure to a reported value (e.g., 6 will require manual correction to 6.0).
Specific conductance (00094) must be reported to three significant figures when the value exceeds 100.	Auto-correction 1014 corrects to 1010 1267 corrects to 1270	
Salinity (00480) must be reported to the nearest tenth of a part/thousand when the reported value is above 2.0.	Auto-correction 3.12 corrects to 3.1 7.77 corrects to 7.8	Whole numbers will require a manual correction to add the tenths digit. SWQMIS will not add a significant figure to a reported value (e.g., 3 will require manual correction to 3.0).
If the Station is a freshwater or inland (brine) location, do not report salinity.		Parameter must be removed by submitting entity.
Secchi disk (00078) must be reported to two significant figures.	Auto-correction	
Days since last significant precipitation (72053) must be reported as a whole number	Manual correction	
E. coli (31699) must be reported as a whole number with two significant digits.	Auto-correction 854 corrects to 850	
E. coli (31699) must not be reported as zero.	Manual correction	

Rule	SWQMIS Action & Examples	Notes
Enterococcus (31701) must be reported as a whole number and with two significant figures	Auto-correction 858 corrects to 860	
Enterococcus (31701) must not be reported as zero.	Manual correction	
Fecal coliform (31616) must be reported as a whole number with two significant figures.	Auto-correction 1214 corrects to 1200	
Fecal coliform (31616) must not be reported as zero.	Manual correction	
Fecal coliform (31616) must not be reported as TNTC.	Manual correction	
Flow (00061) values less than 10 and greater than 0.1 must be reported to the nearest tenth.	Auto-correction 8.62 corrects to 8.6	Whole numbers will require a manual correction to add the tenths digit. SWQMIS will not add a significant figure to a reported value (e.g., 4 will require manual correction to 4.0).
Flow (00061) values greater than 10 must be reported to the nearest whole number.	Auto-correction 15.6 corrects to 16	
Flow (00061) values less than 0.01 must be reported as <0.01.	Manual correction	
Flow severity (01351) must be a whole number in the range of 1 through 6.	Manual correction	
If Flow (00061) is reported as zero, then Flow Severity (01351) must be reported as 1.	Manual correction	
If Flow Severity (01351) is reported as 6, then Flow (00061) must not be reported.	Manual correction	
Composite samples must include all required fields.	Manual correction	
Each TagID in the Event file must have at least one reported Result.	Manual correction	
Each TagID in the Result file must have a TagID in the Event file or already in SWQMIS.	Manual correction.	
Startdate must be before Enddate. If the Startdate and Enddate are the same, then Starttime must be prior to Endtime.	Manual correction.	
Outliers must include a '1' in the Verify_flg field in the Result file.	Manual correction.	