



TASK 4: DATA MANAGEMENT

Contents

Introduction	4-2	
Data Management Roles	4-2	
Basin Planning Agency Data Management		4-2
TCEQ Data Management	4-2	
Preparing and Reporting Data	4-2	
Formatting Data	4-2	
Coding Data	4-3	
Biological Electronic Data Reporting		4-3
Composite Samples	4-3	
Data Review and Validation	4-4	
Data Review Checklist	4-4	
Validating Outliers	4-5	
SWQMIS Validations	4-5	
Submitting Data to the TCEQ	4-6	
TCEQ Data Review	4-6	
TCEQ CRP Project Manager Data Review		4-6
Other Data Management Considerations	4-6	
Creating New Sampling Stations and New Codes		4-6
Data Correction Requests	4-7	
Data on Planning Agency Web Pages	4-7	
CRP Data Management Training	4-7	

Exhibits

Exhibit 4A	Data Review Checklist	4-8
Exhibit 4B	Data Summary	4-10
Exhibit 4C	SWQMIS Validations	4-12

TASK 4: DATA MANAGEMENT

Introduction

This task involves the preparation of water quality monitoring data to ensure data quality and compatibility with TCEQ requirements.

Data Management Roles

Basin Planning Agency Data Management

The Planning Agency Data Manager is responsible for preparing data sets of quality-assured data in accordance with the Data Management Reference Guide (DMRG) for submittal to the TCEQ. Each data set submitted to the TCEQ should contain data collected under a single QAPP and should contain data collected no more than eight months prior to submission.

The Planning Agency Data Manager will review each data set using the Data Review Checklist (Exhibit 4A), and will prepare a Data Summary (Exhibit 4B) to be submitted with each data set. The Data Summary will contain basic identifying information about the data set, information regarding inconsistencies and errors identified during data verification and validation steps, and/or problems with data collection efforts.

TCEQ Data Management

The TCEQ staff receive and review data sets, and other requests for new codes / monitoring stations, or corrections to existing data. The TCEQ conducts automated reviews of incoming data sets and reviews data verification reports generated by SWQMIS against specifications in the QAPP.

Preparing and Reporting Data

Formatting Data

Data will be prepared for entry into the statewide water quality database (SWQMIS) by creating two text files that are related by a unique identification number (Tag ID) that is assigned to each sampling event. The first file has each sampling event (unique date, time, and place (depth & station)) and the second file has each measurement result that was collected and analyzed for that event (so there are many results for each event). These files, called the Events and Results files are formatted using the data dictionaries defined in Chapter 7 of the DMRG. When formatting the two files, the vertical bar typically found on the same key as the back-slash is used to delimit the fields. This vertical bar is called a "pipe", and so the text files are commonly known as "pipe-delimited" text files. It is important to use the "pipe" because there is a chance that other commonly used delimiters will be used in the "Comments" field and therefore cause a problem when loading the fields into a database.



The text files should only contain the data and no header rows, since these are problematic for the SWQMIS loader program.

Coding Data (Tag_id, Tag Prefix, Submitting Entity, Collecting Entity, and Monitoring Type)

The *Tag_id* is a unique alphanumeric identifier for each sampling event that links the sampling event in one file to the measurement values in the other file, in a one-to-many relationship. The *Tag Prefix* is the unique one or two digit letter code added to the beginning of the *Tag_id* and identifies the Basin Planning Agency that is preparing the data set.

A list of valid Tag Prefixes can be found in Chapter 5 of the DMRG at:

www.tceq.state.tx.us/compliance/monitoring/water/quality/data/wdma/dmrg_index.html

Submitting Entity codes identify the agency listed in the QAPP as the entity responsible for submitting the data to the TCEQ (QAPP Entity), *Collecting Entity* codes identify the organization responsible for collecting the data (field entity), and *Monitoring Type* codes identify the type of monitoring under which the reported data was collected.

Example: SR|LW|RT means 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 prefix would be "J" which is assigned to the Sabine River Authority.

Lists of valid codes can be found in Chapter 4 of the DMRG at:

www.tceq.state.tx.us/compliance/monitoring/water/quality/data/wdma/dmrg_index.html

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 shocked, nekton seined, nekton summary statistics, 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., This tag is for nekton seined. See X023847 for habitat, X023848 for nekton).

Composite Samples

Composite samples require entries in several additional fields in the Events file. 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 Review and Validation

The data review and validation process combines the data validation and verification requirements defined in Task 2 with those outlined in this task. The major considerations for this process involve checks for reasonableness and verifying that quality control limits were met. Data that does not meet specifications of the QAPP will not be submitted to the TCEQ. Each anomaly that causes a data point to not meet QAPP specifications will be described in the *Data Summary* (Exhibit 4B).

Data Review Checklist

The *Data Review Checklist* (Exhibit 4A) covers three main types of review: data format and structure, data quality review, and documentation review.

The *Data Format and Structure* section includes checks for required entries and formats. This section can be automated by developing a computer program that checks the database for outliers, other data anomalies, and some types of data transcription errors. This includes checks such as:

- Are there any duplicate Tag ID numbers and are Tag Prefixes correct?
- Do the sampling dates in the Results file match those in the Events file for each Tag ID?
- Are the codes for Submitting Entity, Collecting Entity, and Monitoring Type consistent with the entity and type of monitoring conducted?
- Are the sampling dates and times in the correct format with leading zeros (MM/DD/YYYY) and (HH:MM)?

See the Data Review Checklist, Exhibit 4A, for a complete list.

The *Data Quality Review* section includes checks specific to the acceptability of the data. This requires a more in-depth review of the data by personnel that understand the results of the laboratory analyses. This section includes checks such as:

- Are the required reporting limits consistent with those in the QAPP?
- Have outliers been confirmed and a code entered into the *Verify_flg* field?
- Do the laboratory results appear reasonable and acceptable when compared to other corollary data and/or historical measurements?
- Are all sampling sites defined in the QAPP?
- Are all parameter codes listed in the QAPP?

See the Data Review Checklist, Exhibit 4A, for a complete list.

The *Documentation Review* section includes checks of the quality control information that is developed and provided by the laboratory or field staff. This section includes checks such as:

- Are blank results acceptable as specified in the QAPP?
- Were control charts used to determine the acceptability of field duplicates?
- Were there any failures in field and laboratory measurement systems that were not resolvable and resulted in unreportable data?
- Have any anomalies been reported on the *Data Summary*?
- Was the laboratory's NELAP Accreditation current for the analysis conducted?

Validating Outliers

The TCEQ establishes a minimum and maximum value for each parameter above or below which a value is considered an “outlier” by the TCEQ. The min/max values represent a statistically derived range based on historical data (e.g., 1st and 99th percentile). Reported values that are found to be outliers should be checked against field and laboratory records to verify the correctness of the value 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 *Verify_flg* field.

A file containing all parameters and their min/max levels (Outliers_24feb09.xls) has been provided to each Planning Agency. Planning Agencies should ensure they have the most recent version of Outliers_24feb09.xls when preparing data for submittal.

SWQMIS Validations

Planning Agencies load data sets into the test environment of SWQMIS to insure that data is formatted correctly and meets data reporting requirements described in the DMRG and Table 3.1 of the *SWQM Procedures Manual*. Instructions for loading data are available for download on the CRP web page in the section related to *Data, Forms, and Map Resources*.

The SWQMIS Validator provides a report of errors in the data set relating to data structure, outliers, significant figures, and rounding for certain parameters. Some errors are over-written by the system and will not require a correction to the data files, while others are not. A full list of validations performed in SWQMIS is located in Exhibit 4C.

Once the data is free of errors, the SWQMIS Validator produces a report called the *Loading Validator Report* that contains:

- Date Range
- Tag_id Range
- Count of records in the data set
- Parameter codes submitted with the data set and the number of times each was reported
- The minimum and maximum values submitted for each Parameter code
- Stations for which data was submitted
- Number of events at each station
- Submitting Entity / Collecting Entity / Monitoring Type codes in the data set
- Outliers in the data set
- A comparison of data reported against the historical statistics at each station
- The ten highest and lowest values reported for each parameter

The report should be reviewed against the Monitoring Activities section of the quarterly progress report (see Exhibit 1D), the QAPP, and the *Data Summary* (Exhibit 4B) to ensure consistency in reported values and site locations.

Submitting Data to the TCEQ

Water quality data will be submitted to the TCEQ CRP Project Manager via email as a deliverable at least three times per year. The deliverable is due on March 1, August 1, and December 1 of each year. The March 1st date will ensure that all the data collected through November 30th of the prior calendar year will be submitted to the TCEQ for use in the Texas Water Quality Integrated Report. The August 1st date will ensure data is submitted prior to the close of the fiscal year, and the December 1st date ensures the TCEQ has an updated water quality database. The data submitted to the TCEQ needs to be up-to-date and therefore, each data submittal will include data collected no more than six months prior to submission.

Along with the water quality data, the Planning Agency's Data Manager provides the SWQMIS *Loading Validator Report* to the TCEQ CRP Project Manager (.pdf or .html file) with a *Data Summary* (Exhibit 4B) to the TCEQ CRP Project Manager. The *Data Summary* explains data discrepancies (e.g., missing measurements), describes field and lab issues, and indicates whether the Corrective Action Process has been initiated. *Corrective Action Status Reports* referencing Corrective Action Plans relating to the missing measurements will be submitted with the Progress Report after they are finalized. Additionally, when data cannot be submitted due to either sampling complications or QC failures, a running tally of this data should be maintained and reported on the Data Summary to determine overall percent completeness per data submittal.

TCEQ Data Review

TCEQ CRP Project Manager Data Review

The TCEQ CRP Project Manager utilizes the SWQMIS *Loading Validator Report* to ensure data meet QAPP specifications and contract requirements.

The *Loading Validator Report* is compared to the respective QAPP, the Monitoring Activities section of the quarterly progress report (see Exhibit 1D), and *Data Summary* to reconcile expected versus actual results. One of the most critical steps in this review is checking whether the required limits of quantitation listed in the QAPP correspond to the minimum values submitted in the data set. When discrepancies exist, the CRP Project Manager will ask the Planning Agency to reconcile the differences and make corrections as necessary. This may require additional review of the QAPP when methods, parameter codes, or required reporting limits are changed. When all errors and discrepancies have been reconciled, the data set is approved for upload to SWQMIS.

Other Data Management Considerations

Creating New Sampling Stations and New Codes

Requests for the creation of new monitoring stations will be handled via an internet connection to the SWQMIS interface. Each Basin Planning Agency has been given access to the Monitoring Stations



Module as well as the Reports Module of the SWQMIS interface. Specific instructions for requesting new sampling stations are included in Section 4.3 of the SWQMIS User's Guide at:

<http://www8.tceq.state.tx.us/SwqmisWeb/help/output/index.htm?page=html/441CreatingStationRecord.html>.

Additional helpful information for creating new stations or requesting changes to existing stations can be found in chapter 3 of the DMRG at:

http://www.tceq.state.tx.us/compliance/monitoring/water/quality/data/wdma/dmrg_index.html

Requests for new codes relating to Submitting Entity, Tag Prefix, Collecting Entity and parameter codes should be coordinated with the CRP Project Manager. The Monitoring Type codes have been set for specific data use purposes, and new ones are created only if there is a strong business need. Lists of the existing codes are available in the DMRG. The forms for submitting the requests can be found at:

http://www.tceq.state.tx.us/compliance/monitoring/water/quality/data/wdma/wdma_forms.html.

Data Correction Requests

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:

http://www.tceq.state.tx.us/compliance/monitoring/water/quality/data/wdma/wdma_forms.html.

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.

Data on Basin Planning Agency Web Pages

The Basin Planning Agency should make current data available to the public, and include TCEQ-collected data if feasible (a disclaimer should be provided on the website if the complete dataset for the basin is not available). 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 to satisfy this deliverable.

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 DATA REVIEW CHECKLIST



EXHIBIT 4A Data Review Checklist

This checklist is to be used by the Planning Agency and other entities handling the monitoring data in order to review data before submitting to the TCEQ. This table may not contain all of the data review tasks being conducted.

Data Format and Structure	✓, X, or N/A
A. Are there any duplicate Tag Id numbers in the Events file?	
B. Do the Tag prefixes correctly represent the entity providing the data?	
C. Have any Tag Id numbers been used in previous data submissions?	
D. Are TCEQ station location (SLOC) numbers assigned?	
E. Are sampling Dates in the correct format, MM/DD/YYYY with leading zeros?	
F. Are sampling Times based on the 24 hr clock (e.g. 09:04) with leading zeros?	
G. Is the Comments field filled in where appropriate (e.g. unusual occurrence, sampling problems, unrepresentative of ambient water quality)?	
H. Submitting Entity, Collecting Entity, and Monitoring Type codes used correctly?	
I. Sampling dates in the Results file match those in the Events file for each Tag Id	
J. Are values represented by a valid parameter code with the correct units?	
K. Are there any duplicate parameter codes for the same Tag Id?	
L. Are there any invalid symbols in the Greater Than/Less Than (GT/LT) field?	
M. Are there any Tag Ids in the Results file that are not in the Events file or vice versa?	
Data Quality Review	✓, X, or N/A
A. Are "less-than" values reported at the LOQ? If no, explain in Data Summary.	
B. Have the outliers been verified and a "1" placed in the Verify_flg field?	
C. Have checks on correctness of analysis or data reasonableness been performed? e.g., Is ortho-phosphorus less than total phosphorus? Are dissolved metal concentrations less than or equal to total metals? Is the minimum 24 hour DO less than the maximum 24 hour DO? Do the values appear to be consistent with what is expected for site?	
D. Have at least 10% of the data in the data set been reviewed against the field and laboratory data sheets?	
E. Are all parameter codes in the data set listed in the QAPP?	
F. Are all stations in the data set listed in the QAPP?	
Documentation Review	✓, X, or N/A
A. Are blank results acceptable as specified in the QAPP?	
B. Were control charts used to determine the acceptability of field duplicates?	
C. Was documentation of any unusual occurrences that may affect water quality included in the Event table's Comments field?	
D. Were there any failures in sampling methods and/or deviations from sample design requirements that resulted in unreportable data? If yes, explain in Data Summary.	
E. Were there any failures in field and/or laboratory measurement systems that were not resolvable and resulted in unreportable data? If yes, explain in Data Summary.	
F. Was the laboratory's NELAC Accreditation current for analysis conducted?	



EXHIBIT 4B

DATA SUMMARY



EXHIBIT 4B Data Summary

Data Set Information

Data Source: _____

Date Submitted: _____

Tag_id Range: _____

Date Range: _____

- I certify that all data in this data set meets the requirements specified in Texas Water Code Chapter 5, Subchapter R (TWC §5.801 et seq) and Title 30 Texas Administrative Code Chapter 25, Subchapters A & B.
- This data set has been reviewed using the criteria in the Data Review Checklist.

Planning Agency Data Manager: _____ Date: _____

Comments

Please explain in the table below any data discrepancies discovered during data review including:

- Inconsistencies with LOQs
- Failures in sampling methods and/or laboratory procedures that resulted in data that could not be reported to the TCEQ (indicate items for which the Corrective Action Process has been initiated and send *Corrective Action Status Report* with the applicable Progress Report).

Parameter	Tag Ids Affected	Type of Problem	Reason for Problem	Percent Loss*	Corrective Action (Y/N/SOP)
pH	XL12345	Post calibration	Equipment failure	4%	SOP
pH	XL12346	Post calibration	Forgot to write in log	4%	N
TKN	XL12351- XL12353	Laboratory analysis	LOQ Check Standard failed	10%	Y
TOC	XL12345- XL12350	Exceeded hold time	Sample received late in day and not set up next day.	10%	Y
Zinc	XL12365	Field equipment blank	Possible contamination	4%	N

* Percent Loss = # Data Points Lost / # Data Points Expected for that parameter in the data set.



EXHIBIT 4C

SWQMIS VALIDATIONS

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 Submitting Entity, Collecting Entity, and Monitoring Type Codes.
Check for validity of StationIDs.
Check for validity of TagID prefixes.
Check for validity of Parameter Codes.
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. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Auto-correction 30.11 corrects to 30.1 28.55 corrects to 28.6 30 corrects to 30.0	
pH (00400) must be reported to the nearest tenth of a pH standard unit. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Auto-correction 7.22 corrects to 7.2 6.88 corrects to 6.9 7 corrects to 7.0	
Dissolved oxygen (00300) must be reported to the nearest tenth of a mg/L. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Auto-correction 6.33 corrects to 6.3 4.19 corrects to 4.2 6 corrects to 6.0	
Specific conductance (00094) must be reported to three significant figures when the value exceeds 100. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	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. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Auto-correction 3.12 corrects to 3.1 7.77 corrects to 7.8 3 corrects to 3.0	
If the Station is a freshwater or inland (brine) location, do not report salinity. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).		Parameter must be removed by submitting entity.

Rule	SWQMIS Action & Examples	Notes
Secchi disk (00078) must be reported to two significant figures. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Auto-correction 0.351 corrects to 0.35	
Days since last significant precipitation (72053) must be reported as a whole number. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Manual correction	
If sample collected when raining or has rained within the last 24 hours, report a value of <1 for Days since last significant precipitation (72053). (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Manual correction	
E. coli (31699) must be reported as a whole number with two significant digits. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Auto-correction 854 corrects to 850	
E. coli (31699) must not be reported as zero. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Manual correction	
Enterococcus (31701) must be reported as a whole number and with two significant figures. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Auto-correction 858 corrects to 860	
Enterococcus (31701) must not be reported as zero. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Manual correction	
Fecal coliform (31616) must be reported as a whole number with two significant figures. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Auto-correction 1214 corrects to 1200	
Fecal coliform (31616) must not be reported as zero. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Manual correction	



Rule	SWQMIS Action & Examples	Notes
Fecal coliform (31616) must not be reported as TNTC. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Manual correction	
Flow (00061) values less than 10 and greater than 0.1 must be reported to the nearest tenth. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Auto-correction 8.62 corrects to 8.6 4 corrects to 4.0	
Flow (00061) values greater than 10 must be reported to the nearest whole number. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Auto-correction 15.6 corrects to 16	
Flow (00061) values less than 0.01 must be reported as <0.01. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Manual correction	
Flow severity (01351) must be a whole number in the range of 1 through 6. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Manual correction	
If Flow (00061) is reported as zero, then Flow Severity (01351) must be reported as 1. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	Manual correction	
If Flow Severity (01351) is reported as 6, then Flow (00061) must not be reported. (Rule is from SWQM Procedures Manual Volume 1 (10/2008), pages 3-14, Table 3.1).	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.	