Sediment Quality Database Technical Reference Manual

Version 1.0

SEPTEMBER 13, 2002

Los Angeles Region
CONTAMINATED SEDIMENTS
TASK FORCE

Los Angeles Basin Contaminated Sediments Task Force

Prepared by:

Data & Mapping Services

Peggy L. Myre
Exa Data &Mapping Services
907 Dayton Street
Edmonds, WA 98020
425-672-1947 (ph)
425-672-0748 (fax)

Steven M. Bay
Southern California Coastal Water Research Project
7171 Fenwick Lane
Westminster, CA 92683
714-372-9204 (ph)
714-894-9699 (fax)

Table of Contents

Acknowledgments

The database described in this document could not have been prepared without the assistance of many individuals, representing both public agencies and private companies, working in collaboration through the Los Angeles Basin Contaminated Sediments Task Force. The authors especially wish to thank programmer Kari Moshenberg, SCCWRP staff Doris Vidal, Larry Cooper, and Shelly Moore, and also Paul Johansen (Port of Los Angeles), Rick Cameron (Port of Long Beach), and Kathy Anderson (U.S. Army Corps of Engineers) for their assistance in creating and improving this database. Additional thanks to the staff of the following organizations for their assistance through the CSTF Sediment Thresholds Subcommittee:

California Coastal Commission*
California Regional Water Quality Control Board, Los Angeles Region*
United States Environmental Protection Agency*
Department of the Army*
County of Los Angeles*
Port of Long Beach*
City of Long Beach*
City of Los Angeles*
Amec Earth and Environmental, Inc
Anchor Environmental, L.L.C.
Hart Crowser, Inc.
Heal the Bay
MEC Analytical, Inc.
Santa Monica Bay Restoration Project
URS Corp.

* CSTF Signatory Agency

This document was prepared in partial fulfillment of task 2 of Agreement 00-103-140-0 with the State Water Resources Control Board, for $426,350. The agreement amount for this task represents compensation for other work in addition to that contained in this report.

Section 1

1 Overview of the Technical Reference Manual

The Sediment Quality Database (SQD) Technical Reference Manual ('manual') is intended as a companion to the User Guide. It contains technical details of the components of the SQD, including:

- Table and field definitions
- Database structure and relationships
- Form and code descriptions
- Suggestions for long-term database maintenance

The manual was created to allow the user to move efficiently to the particular item of interest. Accordingly, the file was created using Adobe® Acrobat (portable document format, or pdf), with linked documents for the different components. This main document contains a book marked Table of Contents intended to ease navigation through the different components of the manual. If you do not see the bookmarks, select "Show Bookmarks" under the Window menu item of Adobe Acrobat.

There is some overlap between the User Guide and this manual. The Technical Reference Manual is intended for the experienced database professional to provide the technical details necessary to understand and maintain the database. The User Guide contains an overview of the purpose of the database, and a layman's description of tables, fields, and conventions. It also contains guides on use of the graphical user interfaces (GUIs) that were created as part of the database.

1.1 Components of the Sediment Quality Guideline Database

Version 1.0 of the CSTF SQD was delivered not only with the database, but also a series of additional databases, files, and utilities discussed in the User Guide. Although this manual is primarily intended for the managing the SQD itself, we have also included some references and support for these other components. An outline of these components is provided here, and a summary of where additional information can be located.

1.1.1 SQD, EQT, and Comment Reporter

The SQD has been integrated with two utilities for the general user of the database. Full documentation on these utilities is provided in the User Guide. The Enhanced Query Tool (EQT) is a series of forms with Visual Basic (for applications; VBA) code that allows the novice user to build queries. These forms are dependent on the current structure and nomenclature of the database; should changes be made to the tables and/or fields, the EQT may not function properly. Documentation on the EQT forms and coding is provided in Section 2.

The Comment Reporter is another VBA utility that works both within the EQT, and independently. It is linked to the Comment form and table, and contains fields that are automatically populated from the database when opened. The Comments table is the table where the database will store maintenance and change information; more discussion of this function is in Section 4.

1.1.2 Metadata and the MetaManager

A second Access database was created to store metadata from this project. The format is compliant with the FGDC (Federal Geographic Data Committee) standard. The Access database has an interface (using forms and VBA) called the MetaManager that allows easier browsing and adding of metadata, as well as creation of exported text files compatible with the federal standard. A description of how to use the MetaManager is included in the User Guide. More information on the FGDC Content Standard for Digital Geospatial Metadata is available on the web at the URL: http://www.fgdc.gov/metadata/contstan.html.

1.1.3 Templates

Two types of templates were included to aid addition of new data to the database. The first include files that can be used as flat-file templates for a laboratory to provide data that can be appended into the SQD. There is also a utility to aid hand-entry of dredged material characterization data. The User Guide has detailed information for how to use the data entry forms. Further addition on how to append new data into the SQD is in Section 4 of this document.

1.1.4 Archived Data

With Version 1.0 of the CSTF SQD CD-ROM, we included two more sets of data. First, the source data were provided in the original formats, prior to merging. These data are provided to allow users to access the original data for questions regarding a specific study. This directory also includes all the processing notes for the electronic databases (summarized in the metadata record for each study). Please note that extensive quality review was conducted on the original data sets, these changes are NOT reflected in the original data sets.

The second database is called the Option 2 database. It includes data merged at the study and station levels only. It is provided so users can more easily include new or updated results from the original databases. No further documentation of the Option 2 database is included in the Technical Reference manual.

1.1.5 Shape Files

The dredged material characterization data commonly represented a composite of several geodetic positions. While a single position was selected to represent these areas (see Section 3), and the original coordinate for each individual core is stored (*tblCoreEvent*), the most representative format for these data was selected as a point theme shape file. Therefore, one shape file for each area was created using ESRI® ArcView, and saved using the naming convention of the study number concatenated with the name of the station/area. These files are provided in a subdirectory of the Version 1.0 CD-ROM. These were created using an ArcView script that selected a group of points with a unique station name, and then saving them as a separate shape file. For new data, these shape files could easily be saved by hand using ArcView. No additional documentation is provided on the shape files.

1.2 Overview of the Sediment Quality Guideline Database

1.2.1 Hierarchy

The SQD structure contains four levels of organization: Study, Station, Sampling, and Data (ERD_V1.pdf). This organization reflects the very different study design between dredging characterization and monitoring data. Study, station, and data tables have been merged for dredging and monitoring data. Sample information has been kept in two different sets of tables (Section 1.2.2).

The top level SQD hierarchy is the Study. Every dredging report, as well as every monitoring dataset, is one study. Each study has a unique identifier (***StudyID***) in the SQD. The tables *tblStudy* and *tblStudyReference* contain information about each one of the studies in the SQD. Appendix A in the User Guide summarizes the studies included in Version 1.0 of the SQD. The ***StudyID*** is equivalent to the ***DatasetID*** in the Metadata Database.

Every record in the SQD is related to a single geodetic point in the Station table (*tblStation*), although this point may represent an area for some of the dredging studies. More information about this is provided in Section 3 of this document.

Dredging (core) and monitoring (grab) sampling information is stored in separate tables (Section 1.2.2). The type of study is stored in many of the tables; this code represents whether the study or station was derived from dredging-related data ("D"), or monitoring-related data ("M"). The field name for the sample (the sediment, tissue, or water sample linked to a distinct set of results) is called the ***CompositeID***. Dredging data are also related to a table storing dredged material fate information (Section 1.2.4).

Finally, the lowest level of the SQD contains the data tables (Section 1.2.3). These tables are organized by information type (e.g., chemistry, toxicity, or infauna) and contain the results of measurements from both dredging and monitoring studies.

1.2.2 Sampling Information

Dredged material characterization data are commonly collected using long cores, with composite samples created from parts of several different cores. As an example, one sample may reflect the top half of five different cores (five different locations), representing an upper layer of a berth area to be dredged. Conversely, monitoring data typically are collected using a grab sampler, with one sample reflecting a single point location. The SQD structure reflects this difference, with several tables containing only dredging data (tables starting with 'tblCore'), and several tables reflecting only monitoring data (tables starting with 'tblGrab'). The naming convention can be deceiving in that there are grab sample data in the Core tables (e.g., reference grabs collected near the offshore disposal sites), and there are core data in the Grab tables. More accurately, tables with the preface tblCore contain only dredging-related information.

A master sample table was created (*tblSampleMaster*) that contains basic sample information for both core and grab data. This table increases the efficiency of querying, and simplifies the use of the SQD for novice users. It contains the basic information necessary to describe a sample event (e.g., date, sediment depth). The core and grab tables need only be accessed for field information specific to that type of data. Population of the master sample table can be a complex operation; more details are provided in Section 3 of this document.

1.2.3 Data Tables

Sediment, elutriate, and tissue chemistry data are all stored in the same table, called *tblChemistryResults*. The matrix of the result is stored in a field called 'MaterialCode;' options include SD (sediment), EL (elutriate), and TS (tissue). Method information is stored in the related *tblChemistryBatchInfo*.

There are two tables that contain toxicity results data:

- *TblToxicityResults* - contains raw replicate results, if available;
- *TblToxicitySumResults* - contains mean toxicity results, and includes identifiers for statistical significance of toxic response.

Associated toxicity laboratory information is stored in *TblToxicityBatchInfo*, and water quality information is stored in *TblToxicityWaterQuality*.

There are two tables that contain benthic infaunal data:

- *TblInfaunalAbundance* - contains infaunal species name and count;
- *TblInfaunalBiomass* - contains infaunal group and mass in grams.

1.2.4 Dredging-Related Information

The dredging characterization studies often contain information about the volume and expected suitability of the material for disposal. The table called *tblDredgeFate* contains this information, as well as related information from the dredging and regulatory agencies such as permit numbers, volumes dredged, and the location or type of disposal. Maintenance of this table is complex as it sometimes relates to a study, sometimes to a station, and sometimes to a sample. More information for maintaining the Dredge Fate table is in Section 3 of this manual.

Section 2

2 Technical Specifications of the SQD

This section is linked to a series of stand-alone documents in pdf stored in the same directory as the Technical Reference Manual that describe the technical specifications of the CSTF SQD CD-ROM. More descriptive language is used to describe the basic tables and fields in the User Guide. This section is meant to provide a database specialist with technical details of the database tables, fields, forms, and code.

2.1 Table Definitions

There are two types of tables in the SQD: the data table, and the Lookup List. Data tables start with the preface 'tbl' followed by the description of the data within that table (e.g., *tblChemistryResults*). These tables contain the actual data stored in the database. The Lookup List tables are reference tables that start with the preface 'luList' followed by the list number, and a descriptive name of the information stored in that table. Lookup lists store standardized definitions of codes and names of data stored in the data tables.

2.1.1 Data Tables

Documentation of data table definitions and properties is available in the file Table List.pdf. In addition, an Excel file with filtering option has been provided in file CSTFDataDictionary.xls. In this file, you can filter the table or field list by a table or field name or range. This is particularly useful to see where fields occur in different tables.

2.1.2 Lookup Lists

Documentation of lookup list definitions and properties is available in the file Table List.pdf.

2.2 Field Definitions

Documentation of field definitions and properties is available in the file Field List.pdf. In addition, an Excel file with filtering option has been provided in file CSTFDataDictionary.xls. In this file, you can filter the table or field list by a table or field name or range. This is particularly useful to see where fields occur in different tables.

2.2.1 Primary Keys

The SQG contains tables that are related on a multiple key index. A description of the primary keys in each table is in the Field List.pdf file, as well as shown graphically in the ERD (Section 2.3). The data tables are generally related on one or more of the following fields:

- ***StudyID***
- ***StationID***
- ***CompositeID***

2.3 Structure and Relationships

A formal entity relationship diagram (ERD) is embedded in the file ERD_V1.pdf. The ERD shows tables, primary and foreign keys, relationships between tables (using crow's feet notation). Related tables are grouped using colored boxes. Lookup list tables are shown in blue. The diagram itself is formatted for a large (poster sized) page; printing on regular paper will probably not be readable.

Most of the relationships have been enforced (requiring a parent record for each child record). In some cases there could be no enforcement; a description of these more complex relationships is provided in the sections below.

2.4.1 Relationships Between Core (Dredging Data) Tables

The three core tables have a complex relational structure that is discussed in more detail here. The three core tables are:

tblCoreEvent – This table stores every occasion of a core being collected, regardless of whether a sample was collected from it. This table includes both the core length recovered, as well as the total penetration depth (in cm). Core events document the cores collected at a station, and thus there is an enforced one-to-many relationship between *tblStation* and *tblCoreEvent*.

tblCoreSamples – This table stores the upper and lower depth of a sample collected from subdividing a core, but generally contains a record that links to CoreEvent only *if a core was subdivided or if the lower depth of the sample was not equivalent to core length*. This table provides the upper and lower depths of a sample collected in the *tblSampleMaster*. For every record in *tblCoreSamples*, there is a parent record (parent core) in tblCoreEvent; thus there is an enforced one-to-many relationship between *tblCoreEvent* and *tblCoreSamples*.

tblCoreComposite – This table stores information about composite samples, and includes data for samples collected from cores, as well as any other dredging-related sample. There may be core events for which there are no core composites (no samples collected from that core). In addition, because *tblCoreComposite* stores information about all samples in the database, there are records in *tblCoreComposite* that are not related to a coring event (e.g., for reference or control stations for which there are no cores). Therefore, there is no direct relationship between *tblCoreComposite* and *tblCoreEvent*. Because *tblCoreSamples* contains both the core event information (e.g., ***CoreID***), and the **CoreComposite** information for the samples collected from cores that ended up in a composite sample (e.g., ***CompositeID***), it serves as a bridge table between *tblCoreComposite* and *tblCoreEvent*. However, it is limited to those composites that were formed from multiple samples collected from a core (split cores). This is important to understand for any queries that are created to summarize coring information.

2.4.2 Laboratory Batch Information

There is not an enforced relationship between *tblChemistryResults* and *tblChemistryBatchInfo*. First, there are many chemical results for which there was no method information. For these records, the ***QABatch*** field is filled with NA, and there is no related record in *tblChemistryBatchInfo*. Second, there are multiple records within *tblChemistryResults* that must be linked to a single record in the batch table (multiple results in one batch); therefore, the chemistry results table cannot be indexed uniquely on the batch table's primary keys (***StudyID*** + ***QABatch***).

Similarly, there are not enforced relationships between *tblToxicityResults* and *tblToxicityBatchInfo* and *tblToxicityWaterQuality*. Laboratory and water quality information was only available for a subset of the data. The laboratory information is related to the raw toxicity result rather than the summary results because there was usually more information for the raw laboratory data.

2.5 Form and Coding Documentation

2.5.1 Forms

A list of the forms used for the EQT is available in the linked file Form List.pdf. Most of the code for the EQT and related utilities is embedded (VBA) in the event procedures of these forms. Documentation of the VBA code is also available (Section 2.5.2).

2.5.2 Code

Documentation of the VBA code embedded in the database module used for the EQT and related applications are provided in the linked file Module.pdf.

Section 3

3 Data Processing And Standards

This section provides details for population of certain tables that are more complex than appending new laboratory data. Here is detailed the logic process that occurred for population of certain fields, along with related assumptions. We also include some conventions for storing information; more information on conventions can also be found in the User Guide.

3.1 Stations and Station Locations

All dredging and monitoring data have a geo-referenced location in latitude/longitude coordinates (NAD83). For dredging data, the 'Station' may actually represent an area, such as a dredging polygon from which multiple cores were collected. If this is the case, the field called 'Area' in *tblStation* will be set to "True."

In order to assign each dredging area to a point location that could be stored in the station table, a single core location was selected to represent the area. The representative location was selected using an ESRI ArcView Avenue script; the median location (the point that was closest to all the other points) was selected from the points stored in *tblCoreEvent* with a unique combination of **StudyID+StationID** fields. Then a table of median points was created and later appended into the station table.

In addition to a point location, each station was assigned to one or more regions using several GIS (geographic information system) layers provided by SCCWRP and the POLA. The regions include:

- ***Hydrounit*** (Hydrologic Unit Code name or watershed description)
- ***Port*** (POLA or POLB)
- ***Regional Board*** (California Regional Water Quality Board)
- ***Locality*** (bays and harbors)
- ***County*** (California counties, primarily)
- ***Country*** (US/Mexico)

The Port, HUC, and Regional Board boundaries were provided by the POLA. The HUC name was used for classification, except for the Dominguez Channel watershed (a sub-classification of the HUC). County and country boundaries were obtained from the California Teale Data Center (see http://www.gis.ca.gov/). County boundaries include a three nautical mile extension into the water, so that the layers incorporate the water-based data. For stations located more than three miles offshore, the nearest county boundary was used.

The Locality regions were based on SCCWRP GIS (MapInfo) layers of bays and harbors in Southern California. Stations that were located upland were identified either from the ***Description*** field in *tblStation*, or stored with a locality of 'upland.' Control stations were identified as to location of the collected sediment (e.g., Tomales Bay); if this information was unavailable or inapplicable (e.g., water controls), the regions were classified as 'NA.'

3.2 Sample Master Table

Population of the Sample Master table (*tblSampleMaster*) was conducted after all of the dredging and monitoring data were populated. This table was added in order to make querying more efficient (e.g., less reliant on the complex core table structure). Basic sample information was included in the master table, including upper and lower sediment depths, grain size and TOC for that sample (for normalization purposes), and the sample date. Population of the upper and lower depths was complicated by the core table structures (Section 2.4.1). Several queries were created to populate these fields:

- Create of a table that stores the average upper and lower depth for a composite sample based on samples from multiple cores (summary query on *tblCoreSamples* based on a unique combination of ***StudyID***+***StationID***+***CompositeID***).
- Create of a table that stores an average lower depth for composites collected from cores that were not split horizontally (average of core length field for Horizon = 1).
- Update SampleMaster upper and lower sediment depth preferentially with average upper and lower depth derived from *tblCoreSamples*. For remaining core records, populate upper depth with 0 and lower depth with averages of core length from *tblCoreEvent*.
- Update upper and lower sediment depths for grab data from the GrabComposite table.

Other complexities occurred with multiple sampling dates for a composite sample (over several cores, the first date was selected), and for laboratory replicates of TOC and grain size (first value was selected). Percent fine grain size was calculated from the silt and clay fractions when fines was not reported.

3.3 Sample Types

The field ***SampleType*** stores information for control and reference samples, as well as duplicate and replicates. The codes are related to luList04, and used in both the SampleMaster table as well as all of the data tables. In some cases the SampleType stored in *tblSampleMaster* may be different from the type in the results tables. This is common for duplicate and replicate results of a single ***CompositeID***. A common example is for bioaccumulation data. Typically bioaccumulation data contained five replicate reported results for a single sample:

Berth142-A, Berth142-B, etc. These data are stored with the ***CompositeID*** Berth142 (to enable linking bioaccumulation data to sediment chemistry and toxicity results), and the replicate identifier stored in the ***LabRep*** field (A, B, etc.). Thus, the *tblSampleMaster* ***SampleType*** may be "Result" while the ***SampleType*** in tblChemistryResults is stored as TREP (Tissue Replicates). In essence, the ***SampleType*** has more meaning for the results/data tables; it was useful for querying and for the EQT, however, to be able to query reference and control data from tblSampleMaster.

In some cases, more than one ***SampleType*** applied to a particular result (e.g., duplicates of a reference or control sample). In this case, the control or reference information took precedence, as the laboratory replicate information is also stored in the ***LabRep*** field. This has implications towards the EQT filtering function, but not to the database itself. For chemistry data, laboratory duplicates and replicates were stored with a ***SampleType*** of DUP, and a labrep of "02" or another secondary identifier (original data were stored with original labrep identification). Tissue (bioaccumulation) replicates are stored with a ***SampleType*** of TREP. In some cases, there may have been a duplicate of a replicate. If, for example, a tissue replicate sample was analyzed twice, then ***SampleType*** = TREP, while ***LabRep*** = A and A2 for the two duplicates.

Field replicates are stored with a ***SampleType*** of FR. Finally, there were some samples for which only infaunal data were collected; these were given a ***SampleType*** of INF in *tblSampleMaster*.

3.4 Dredged Material Fate Table

This table combines two different types of information: first is suitability classifications (provided in dredged material characterization reports either for a station, composite, or core); second is fate information, which documents the actual fate of the dredged material characterized, commonly on a project or area (e.g., berth) basis. Because of the different levels of related information, this table could not be enforced, and is currently linked to the station table, although some records are based on a project (study), and there is composite information in some of the records.

Suitability classifications codes are all linked to the Lookup List 30, as described below:

- Suitability Recommendation – This field contains a text description of the overall toxicity of the material as well as suitability for open-water unconfined disposal. It is based on the chemical and biological data as reported by the characterization report's author.
- Recommend/Regulatory/Disposal Codes – These are standardized codes that describe the suitability of the dredged material for offshore open-water unconfined disposal. The Recommend code stores a summary code from the Suitability Recommendation; the Regulatory code stores the decision for suitability for offshore disposal according to the regulatory agencies; and the

disposal code stores a code for the type of disposal site ultimately used for placement of the sediment. All three codes are described in Lookup List 30.

Other project/fate information was provided by some of the project agencies. More information on the dredge fate table is in the User Guide.

3.5 Chemistry Data Standards

Results were provided as reported unless they were converted to match the standard units. Original data were converted so that units are standardized according to those listed in *luList18_ChemicalParameters*. The standard unit for the majority of organic chemicals is parts per billion (µg/kg). Exceptions are several dioxin/furan compounds (parts per trillion, ng/kg), and conventional organic tests such as TRPH and oil & grease (parts per million, mg/kg). Metals are stored as mg/kg, and grain size and total organic carbon (TOC) are stored in percent (%). Data and units for sediment are in dry weight; data and units for tissue are in wet weight.

For results reported as below detection, a "U" was stored in the qualifier field, and the reported value reflects the reporting limit. Both the reporting limit and method detection limit are also provided. The reporting limit is essentially a practical method detection limit, and also can vary if the sample was diluted (sample dilutions are also stored). The reporting limit is the concentration of a parameter that can be reliably reported in the presence of a moderate amount of sample-based interferences. Definitions of all standardized qualifiers are stored in luList13_QualifierCodes.

3.6 Statistical Significance of Toxicity

A description of the *tblToxicitySumResults* is in the User Guide. However, we provide additional detail for the creation of the statistical codes in the database. There are several different fields that document statistics conducted on the toxicity data:

- ***Mean*** (Average value of laboratory replicates);
- ***N*** (Number of replicates);
- ***StdDev*** (Standard deviation of replicates);
- ***PctControl*** (Mean value expressed as a percent of the negative control assigned to that batch of samples);
- ***SigEffect*** (Reported statistical significance from original report and/or database);
- ***Stat_Test*** (Statistical test used to calculate statistical significance);
- ***NormSigEffect*** (Results of a standardized statistical test between the sample and negative control conducted by the CSTF to provide a more consistent indicator of toxicity).

The codes used for statistical significance for both the ***SigEffect*** and

NormSigEffect fields are described in Lookup List 50. The codes differentiate between comparison to reference (SR/NSR) and control (SC/NSC). There is also a threshold value applied in some cases when comparing to reference. This most often is for sediment (solid phase) toxicity testing for dredging studies; if the resulting value is within 10% of reference, commonly there was no statistical analysis conducted (20% for amphipods, according to dredging guidelines). For more information on federal dredging testing and statistical guidelines (e.g., Green Book), see http://www.epa.gov/owow/oceans/gbook/index.html.

The ***NormSigEffect*** field reports the statistical significance of toxicity test results based on a pair-wise comparison to control. In most cases, there was only one control per test group (***QABatchID***), however in one batch two controls were run which offered two possible control samples to which the site samples could be compared. In the case of this batch, the six samples were compared to each of the control samples and each set of significance results were evaluated. For each sample, the significance results were consistent across the two controls. If the significance results had disagreed, the most conservative result would have been adopted (i.e., resulting in a "hit").

Statistical testing was done using a one-tailed, two-sample test procedure, either a parametric *t*-test (with or without equal variances) or a non-parametric analog, specifically a *t*-test performed on the normal scores (also referred to as normalized ranks or "rankits") rather than the data itself. For the sake of efficiency, three tests were performed on every sample-control pair: the *t*-test with equal variances, the *t*-test with unequal variances, and the non-parametric test. If the results from the three tests were in agreement (i.e., all results indicated that the site sample was significantly different or not significantly different from control, at a one-tailed $\alpha=0.05$) then the significance result was accepted without specifically determining whether the data were normally distributed or whether the variances were equal or not. If the results were not in agreement, then the data for the sample-control pair were evaluated further to determine the most appropriate test. Normality was tested using a Shapiro-Wilks goodness of fit test ($\alpha = 0.05$) on the residuals (i.e., the sample mean was subtracted from each observation). If the data were found to be normal, then equality of variances was tested using Levene's test ($\alpha = 0.05$) to determine the appropriate treatment of variances in the parametric *t*-test. The rankit-transformation tends to stabilize the variances so that inequality of variances is not an issue for the non-parametric *t*-test.

Substituting the normal scores for the original data in standard parametric tests (e.g., the *t*-test) provides a practical non-parametric test that is asymptotically efficient for non-normally distributed data. The use of the rankit-transformation is discussed by Conover (1980; p. 317), and is recommended as the non-parametric test for dredging applications by WES (Clarke and Brandon, 1996). Levene's test is essentially an analysis of variance on the absolute value of the sample residuals; WES recommends it for the test of equality of variances in dredging applications (Clarke and Brandon, 1996).

Percentage data are binomially distributed, and as such it is sometimes recommended that the survival or normality proportions be automatically transformed using the arcsine square-root transformation to theoretically approximate a normal distribution. The arcsine square-root transformation was not automatically applied to the proportion data in this database for several reasons: 1) when proportions are in the intermediate range (i.e., between 20% and 80%) then the binomial distribution approximates the normal distribution and the raw proportion data can be used in the test for normality; 2) when proportions are near the extreme ends of the distribution (i.e., near 10% or 90%) few samples pass the normality test, especially when arcsine-square root transformed, so a non-parametric test should be used; and 3) in this database survival or normality proportions were sometimes greater than one; proportion values greater than one are undefined under the arcsine square-root transformation. For these reasons, it was determined that the untransformed proportions would be used for statistical testing; if the untransformed data failed the normality test then the non-parametric was used. An evaluation of the survival data in the database revealed that the final toxicity significance results for 5 samples in the database would have been affected by use of an automatic arcsine-square root transformation, using the statistical approach described in previous paragraphs; this represents less than 0.5% of the total number of testable survival endpoint samples.

3.7 Required Fields and Nulls

In general, nulls were avoided in the SQD. Although many of the fields are not required in the database (other than key fields), most were populated with relevant information, or with a standard default value if no information was available. This population effort was conducted so the user could know the difference between the different types of the meaning of null (e.g., no information available vs. not applicable).

Required fields and conventions used for nulls and defaults can be found in the file CSTFDataDictionary.xls. Required fields are noted in the table, as well as the default value for valid null fields. Fields that are linked to lookup tables do not always have default nulls noted in the dictionary because the field must match a code in the lookup table. Commonly the null value is one of the options in the lookup table (e.g., UNK for Unknown, NA for Not Applicable, etc.). Default values are used for these mandatory fields when no other information is available. Nulls were acceptable in comment fields, and in fields where null is an implicit not applicable (e.g., null fields in the Qualifier field indicate that there was no qualifier for that result).

In order to avoid having blank fields in the SQD, a series of conventions were adopted to handle missing or unavailable information:

- Missing or unreported numerical information has a –99 in the field;
- Missing text information has an “NA” in the field; the exception is if there the field links to a Lookup List that has a specific code for missing, inapplicable, or unreported data;
- Missing or unreported dates are filled with 1/1/1900;
- Missing or unreported times are filled with 00:00.

Section 4

4 Long-Term Maintenance

4.1 Adding New Data

The intent of the SQD is to be updated on a regular basis with corrections and additions to the data. Although anyone can modify the information contained in the database, this activity is strongly discouraged. For publicly distributed databases such as the SQD, it is important that changes be made solely by the database manager. In this way, changes can be documented and all users will have access to the same information. Submission of data by others is encouraged by the CSTF.

In the 'Templates' folder of the CSTF SQD CD-ROM, two files are provided for appending new electronic data into the SQD. First, the file called 'SQD New Data.mdb' is a blank version of the database that can be used to append new data into the main CSTF SQD. An Excel file with multiple worksheets is provided as an extra format, and contains more fields for metadata type information. These are discussed below in Section 4.2.1.

For hand-entering dredging related data, we have also provided a database called 'Dredging Data Entry.mdb' containing a series of forms used to enter dredged material characterization data into the database. As part of this project, a series of data entry screens were created to enter the data more efficiently. This utility was originally intended for internal use, to increase the efficiency and accuracy of data entry. It is provided here as an additional resource for the CSTF and associated users. Although it was not intended for public distribution, we have provided detailed documentation on how to use the tool, and any problems that you may encounter in using it (Section 4.2.2).

After using the templates, data can be appended to the master database if desired, using Access' internal Update Query function.

4.1.1 Electronic Data

A version of the SQD with empty data tables (complete Lookup Lists are included) is provided on the CD-ROM for appending new electronic data to the SQD (SQD_NewData.mdb). The blank database includes two table templates for parsing electronic data, which usually comes in a single table or "flat file." These tables (*tblGrabImport* and *tblCoreImport*) can be used to merge new electronic data into the SQD. The tables contain the basic required fields for sediment grab data (one station/one sample), and for dredging-related coring data (one station/multiple positions). There is no tool or template to automate the append process; it should be managed by a database professional with experience in standardized naming conventions using the Technical Reference Manual. As always, a MetaManager record should be established for each new study appended to the database.

In addition to the Access database, we have also provided an Excel template that also contains the basic field information for storing study, station, and sediment chemistry data (Chemistry_GrabTemplate.xls).

4.1.2 Entering Data from Hard-Copy Reports

The second file provided in the Templates folder on the CD-ROM is also a copy of the master database (empty data tables, with Lookup Lists). This file (Dredging Data Entry.mdb) includes a series of linked forms for entering dredged material characterization data from hard-copy reports.

4.2 Comment Database and Updates File

The *tblComments* table is intended to store data quality or functionality comments from the users, especially during the first year use test. Users should use the Comment Reporter function in their personal copy of the database (see User Guide). If they find an error in the SQD or bug in the programming, they can fill in the date, their name, and a description of the error. These comments are stored in *tblComments*, which then can be forwarded to those responsible for maintaining the database at sqd@sccwrp.org.

The master *tblComments* has fields to store how the issue was addressed by the database manager, and will serve as a running record of corrections to the database. Each delivery of the revised database will include a table called *tblUpdate*. The fields in the update file will contain the table name, all of the key fields required to link the table to the table requiring updating, the old value, and the new value. The client database can then update their version by using an update query with this file. After each delivery, the *tblUpdate* should be archived and then cleared out for the next update.

4.3 EQT, Menu Bar, and Forms Maintenance

When the SQD is first opened, the Enhanced Query Tool (EQT) will automatically open. The EQT enables users to extract data from the SQD without having to learn Access or develop their own queries. There is currently no plan to modify this utility; however, if any changes are made to the database structure, the code of the EQT will have to be edited. A guide to the EQT is included in the User Guide.

The CSTF User Tools Menu bar can be maintained through the Menu Tools feature of Access. It currently has two menu items: the "Main Form" open up frmMain. The "Insert a Database Comment" menu item will open up the database comment form.

4.4 Database Updates

Revisions to the SQD will be released on an annual basis, or more frequently if major additions or changes to the database have occurred. Future versions of the database will be delivered as both revised data tables and as an update file. The data tables can be imported to replace the older tables. For more experienced users who have their own queries and tables, the update file should be used instead of replacing the data tables so that changes can be made to existing tables without deleting any information that may have been added to their personal copy.

Section 5

5 References and Glossary

5.1 References

Conover, W.J. 1980. Practical Nonparametric Statistics, Second edition. John Wiley & Sons, New York.

Clarke, J.U. and D.L. Brandon. 1996. Applications Guide for Statistical Analyses in Dredged Sediment Evaluations. Technical Report D-96-January 1996. U.S. Army Corps of Engineers, Waterways Experiment Station, Vicksburg, MS.

5.2 Glossary

Append – To add data to the database, usually through an Append Query.

Control – Used in the context of a form or database code, a control is function imbedded within a form that allows the user to interface with the data. Examples are (check box, etc.).

Data Table – Tables that store raw and summary data.

ERD – Entity-Relationship Diagram; a diagram that shows how data and lookup tables are related.

EQT – Enhanced Query Tool

FGDC – Federal Geographic Data Committee

Form – Object within the database used as an interface between the user and the data tables, commonly have code imbedded in the form controls.

GIS – Geographic Information System

Index – A table is indexed, or sorted, on a key or series of keys in order to make a query run more efficiently.

Lookup Table – Dictionary table that contains codes and code definitions for fields stored in the Data Tables.

Metadata – Federally-required information about a spatial dataset.

MetaManager – Utility that stores metadata.

Module – Object within the database that stores code for utilities and applications.

Null – A field with no information (this is different then a field with a value of zero).

Primary Key(s) – The field or fields that make each record within a table unique.

Query – A query is a statement that is used to extract data from a database. Also, within Access, it is an object that is used to store query statements. Can also be used as a verb (to query the database).

Relationships – A relationship defines how two or more tables are related through key fields. Common types of relationships between two tables include one-to-one (each record in one table is related to one record in the other), or one-to-many.

Table – Object within the database application used to store information.

Update – To modify data to the database, usually through an Update Query.
VBA – Visual Basic for Applications

Appendices

The following appendices are all available as separate electronic files. If viewing this document in Adobe Acrobat, you can click on the links below to open up and view the files.

I. Description of All Tables of the SQD (Table List.pdf)

II. Description of All Fields of the SQD (Field List.pdf)

III. Description of All Forms Used in the EQT (Form List.pdf)

IV. Description of All Modules Used in the EQT (Module.pdf)

V. Entity Relationship Diagram (ERD_V1.pdf)

VI. CSTF Data Dictionary (CSTFDataDictionary.xls)