PORTLAND HARBOR INTERIM DATABASE (PHIDB) DESIGN SUMMARY

Version 2.0

Prepared for

State of Oregon—Department of State Lands

775 Summer Street NE # 100, Salem, Oregon 97301

Prepared by

Integral Consulting, Inc.	GeoEngineers, Inc.
719 2nd Avenue, Suite 1450	5820 S Kelly Avenue, Suite B
Seattle, Washington 98104	Portland, Oregon 97239

December 16, 2021

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ATTACHMENTS

ACRONYMS AND ABBREVIATIONS

ASAOC	Administrative Settlement Agreement and Order on Consent
CERCLIS	Comprehensive Environmental Response, Compensation and Liability Information System
DBMS	database management system
DDx	The sum of 6 ortho- and para-substituted isomers of dichlorodiphenyltrichloroethane (DDT) and its breakdown products DDE and DDD
EDD	electronic data deliverable
ERD	Entity-Relationship Diagram
EPA	U.S. Environmental Protection Agency
FIPS	federal information processing standard
GIS	geographic information system
GPS	global positioning system
HUC	hydrologic unit code
IDL	instrument detection limit
MDL	method detection limit
MIS	multi increment sampling
NAVD88	North American Vertical Datum of 1988
NGVD29	National Geodetic Vertical Datum of 1929
PCB	polychlorinated biphenyl
PHIDB	Portland Harbor interim database
PHSS	Portland Harbor Superfund Site
QA	quality assurance
QC	quality control
QAPP	Quality Assurance Project Plan
RAL	remedial action level
SDG	sample delivery group
SMA	sediment management area
SQL	structured query language
SRID	spatial reference system identifier
TIC	tentatively identified compound
USGS	United States Geological Survey
VOC	volatile organic compound
WGS84	World Geodetic System 1984
WKT	Well-known text format for spatial data

1 PURPOSE AND GOALS

This document describes the data tables that will be used to implement the Portland Harbor interim database (PHIDB), including explanations and rationale for the design of this data structure. This design summary is intended to document the data management system's ability to accommodate the environmental data used to support decision-making at the Portland Harbor Superfund Site (PHSS). Activities to be conducted in relation to data collection, compilation, and quality assurance review will be carried out by both the U.S. Environmental Protection Agency (EPA) and performing parties; this document does not include descriptions of those parties or the processes to be carried out. The focus of this document is the database design.

1.1 PURPOSE OF THE PORTLAND HARBOR INTERIM DATABASE

The PHIDB is intended to provide a centralized and standardized repository for the environmental characterization data, and related administrative data, relevant to remedial activities at the PHSS. Essential capabilities of the database, and related tools, standards, and processes, are to:

- Ensure that the data collected by various parties are consistent in structure, unambiguous in representation of information, and are structured to establish data integrity. Data integrity is supported using primary keys, foreign keys, and check constraints within the database.
- Ensure that data are available to EPA and performing parties in a single system that allows integration and summarization of information across different sites or study areas within the overall PHSS.
- Ensure that a standard set of rules is applied to summarize data and calculate derived quantities such as total polychlorinated biphenyls (PCBs) and total DDx.
- Allow export of data to a customized version of a Scribe[1] database
- Store administrative data

The PHIDB is intended to include information for making decisions about conditions and the result of actual or potential remedial actions within the PHSS. Environmental investigations typically generate copious supporting documentation in the form of plans, reports, photographs, field logs, geologic boring logs, and laboratory deliverables. The PHIDB is not designed to store these types of accessory documentation, though some of that documentation may be made available on the public Portland Harbor Environmental Data Portal (http://ph-

[1] Scribe is an EPA data management application

public-data.com/). The PHIDB is also not designed to store all of the types of quality control information generated by laboratories (e.g., calibration information and the results of analyses of spiked samples, blanks, and certified reference materials). The PHIDB does store the results of replicated analyses of environmental samples; although replicate analyses are used for quality control evaluations, they also provide additional information that allows more precise estimation of true environmental concentrations.

The PHIDB is expected to become a candidate component of a more comprehensive information management system that will be the repository for both Harbor-wide and site-specific environmental data and institutional controls.

1.2 DESIGN GOALS

Data types that are expected to be collected, and therefore to be accommodated in the PHIDB, include but are not limited to:

- Analytical chemistry measurements of sediment, tissue, surface water, porewater, and groundwater;
- Organism abundances;
- Results of toxicity tests; and
- Geographic data (e.g., site boundaries, and sampling locations)

Design goals include:

1. Accommodate environmental characterization data to be collected during investigations at the PHSS conducted under an Administrative Settlement Agreement and Order on Consent (ASAOC) with the EPA.
2. Store detailed quantitative and categorical data produced by the performing parties of each investigation as practicable to maximize the use of the data for multiple purposes.
3. Accommodate administrative and oversight information such as descriptions of ASAOCs, identities of performing parties and their relationships to the environmental characterization data.

2 DESIGN DEVELOPMENT PROCESS

2.1 INITIAL EPA DRAFT STRUCTURE AND COMMENTS

In October 2019, EPA Region 10 solicited review and feedback on an initial draft of a PHSS Scribe database structure from the following entities:

- Northwest Natural
- State of Oregon (State; Department of Environmental Quality and Department of State Lands [DSL])
- Port of Portland
- River Mile 11 East Group
- Yakama Nation Fisheries
- Five Tribes

During the comment period, DSL reached out to EPA to offer assistance in developing an interim database structure (i.e., this PHIDB) that might better meet the longer term needs of the project as well as populate the EPA's Scribe management database. On behalf of DSL, Cascadia Associates LLC (Cascadia) and Integral Consulting Inc. (Integral) reviewed the comments received by EPA and have worked to incorporate these comments into the design of the PHIDB where applicable. Table A-1 of Attachment A provides a summary of the comments received by EPA, responses to each comment, and, where applicable, actions taken in the development of the PHIDB design to incorporate the comment. As seen in Attachment A, many of the comments have formed the backbone of the development of this PHIDB.

2.2 ADDITIONAL DESIGN DRIVERS

In addition to the comments provided by reviewers of EPA's original draft design, the current PHIDB data model has incorporated features derived from the development team's experience using and developing environmental databases. Some of the additional goals that have helped to drive the design and implementation are:

- Establishment of a single centralized authoritative data system, using a client-server database management system (DBMS) to provide equal access to interested parties and to help reduce the possibility of fragmentation of data into multiple and possibly conflicting systems.
- A focus on automatic establishment and enforcement of data integrity. Data structures that are optimal for efficient storage and enforcement of data integrity are often not the same as the formats to be used for particular analyses. Priority has been given to

efficiency and data integrity because the flexibility of a relational database allows data to be easily transformed into other formats, such as those to be used for various data analyses.

- Accurate modeling of the complexity of environmental sampling designs. Although field replication, subsampling, sample splitting, sample compositing, laboratory replication, and laboratory re-analyses may be a small part of many sampling designs, they can add considerable complexity to the accurate representation of relationships between different samples and different analytical results. The PHIDB has been designed to unambiguously incorporate these elements into its data model.
- Accurate representation of spatial data related to sampling. Environmental samples may be collected at single points, composited over multiple points, or collected over the full extent of a transect or area. Whereas most environmental databases represent sampling locations only as a single latitude/longitude coordinate, the PHIDB is designed to accommodate the full spatial complexity of sampling designs. PHIDB also serves as a repository for the spatial extent of Site and Field Sampling Events.

3 STRUCTURAL ELEMENTS AND RATIONALE

3.1 INTRODUCTION AND OVERVIEW

The database tables in the PHIDB represent two broad classes of information: (1) environmental sampling data, and (2) administrative information related to the environmental data and to other EPA activities associated with the Portland Harbor site. The database tables within each of these broad classes model several different types of information and relationships between different information that are fundamental elements of these classes. The following subsections of this design document highlight important elements and their relationships. The complete model is shown in the tables referenced in following sections; the text of this section does not describe every table and column in detail but provides a rationale and explanation for the major elements of the database design.

The PHIDB data model for administrative data is based on the tables and information that were described in the original draft database design prepared by EPA.

The PHIDB data model for environmental data conforms generally to the paradigm found in most environmental databases, that is: multiple sampling locations, multiple samples per location, and multiple analytical results per sample. The data model expands on this paradigm to provide additional flexibility to accommodate a variety of sampling designs and different data types.

An overview of the database structure is shown as an entity-relationship diagram (ERD) on Figure 1. This figure shows the major data tables and the primary key columns for those tables. Following sections provide more detail on the structure and relationships of subsets of this diagram. A complete ERD diagram for all the PHIDB data tables and column names is shown on Figure 2. The names for PHIDB tables that make up the core of the data model are prefixed with "d_", and the names of all lookup tables containing valid values are prefixed with "l_". The complete data and lookup table names, relationships and column format specifications are provided in Figure 2. The prefixes are not included in this document when referencing the data and lookup tables to improve readability.

3.2 ADMINISTRATIVE INFORMATION

Administrative information includes descriptions of ASAOCs, performing parties, field sampling events related to remedial design, and the relationships between them. These elements, and the relationships between them, are illustrated in the overview of administrative tables ERD shown on Figure 3. Detailed descriptions of the columns in each of these database tables are shown in Tables 1 through 8. Key aspects of the information represented by this component of the data model include:

- There will be multiple ASAOCs. Each one will have a unique identifier, a docket number, and an effective date. This information will be stored in the ASAOC table (see Table 1).
- There are multiple sites within the overall PHSS. Each site is expected to consist of a river segment defined by upper and lower river miles to the nearest tenth of a mile and east/west bank riverbank designations. The Site table in PHIDB (Table 2) will contain a unique identifier for each site, a description of the site, the Comprehensive Environmental Response, Compensation and Liability Information System (CERCLIS) identifier for the PHSS, and spatial extent of the site boundary.
- An ASAOC may apply to multiple sites and/or to multiple parties. Sites and parties may also be related to multiple ASAOCs. ASAOCs and sites will be linked in a many-to-many relationship using the ASAOC Sites table (see Table 3). Each party, or group of parties, that is associated with an ASAOC will be identified in the Party or Party Group tables, and the group name will be stored with the ASAOC description (see Tables 4, 5, and 6). The Organization (Table 7) contains the names of contractors, validators, and field event sponsors and is linked to the Party, Lab Result and Field Event tables.
- Multiple field sampling events may be associated with or mandated by each ASAOC. There may also be other administrative events that do not involve field sampling. Information about these two types of events is recorded in the Field Event and Admin Event tables (Tables 8 and 10, respectively). Each event, of either type, is uniquely identified by an event ID. Each field event will be characterized by the type of field activity to be conducted. The Event Element (Table 9) allows distinct subsets of sampling efforts to be identified (e.g., targeted sampling, random grid sampling).
- Valid values are established for the types of field activities—these include remedial design sampling, remedy effectiveness assessment, and post-remedial monitoring. Administrative events are also characterized by an activity type, and the valid values available differ from those for field events; examples include remedial action oversight and five-year review. The status of both field and administrative events can be recorded (e.g., pending, underway, or complete). The spatial extent of the field activity can also be recorded, because this may not exactly coincide with the spatial extent of the site(s).

The Field Event table (Table 8) links to the rest of the data tables that contain environmental characterization data, including sampling locations, samples, and analytical chemistry results.

3.3 SAMPLING LOCATIONS

Different sampling programs often use different nomenclature for naming sampling locations. These differences can raise issues of consistency or ambiguity in a database that will integrate data from multiple investigations. Sampling designs and procedures can also vary between investigations. The challenges that these variations raise include:

- The same location may be sampled by different investigations, and each investigation may assign a different name to the same location.
- Because sediment and soil surface elevations can vary over time, elevation data are recorded at the time of sample collection or assigned to collections from recent bathymetric survey data.
- Different investigations may assign the same name to different locations.
- Actual sampling locations (coordinates) may differ from the target[2] coordinates, particularly for samples collected from vessels.
- Sampling designs may include collection of material from single points, multiple depths, multiple points, transects, and areas; each of these types is represented by a different type of coordinate data.

The location and field event location database tables used to manage these challenges are illustrated on Figure 4, their structures are shown in Tables 11 and 12, and their use is described in the following sections.

3.3.1 Universal and Event-Specific Location Names

To accommodate the use of different location identifier naming schemes by different investigations (or field events), the PHIDB structure uses two tables to represent sampling locations:

- The Location table (Table 11) is used to record a unique identifier for every target sampling location. The corresponding target coordinates are also stored in the Location table.

[2] The terms "target coordinates," "design coordinates," and "canonical coordinates" all refer to the coordinates at which samples are intended to be collected. These are the coordinates that are ordinarily listed in a field sampling plan.

- The "Field Event Location" table (Table 12) is used to record an event-specific identifier for each actual[3] sampling location. The Event Location actual coordinates are stored in the "Field Collection" table.

There may be one or more event-specific location identifiers (in the "Field Event Location" table) associated with each unique identifier (in the "Location" table). The event-specific identifiers in the "Field Event Location" table are effectively aliases for the unique location identifier.

The "universal" location identifiers that are stored in the "Location" table can be used when data are aggregated across sampling events and to provide a consistent identifier to be shown on maps and in other data summaries.

3.3.2 Target and Actual Locations

In addition to storing unique location identifiers, the "Location" table also stores a single set of coordinates for each unique location in the "Loc Geom" column. These are regarded as the previously described target locations (i.e., the coordinates that are intended to be sampled during each sampling event).

Actual sampling locations are recorded in the "Field Collection" table rather than in the "Location" table that stores target, location identifiers and coordinates.

3.3.3 Standard Coordinate System

All coordinate data will be stored in one coordinate system formatted as decimal degrees in the WGS84 (World Geodetic System 1984) coordinate system and horizontal datum. The North American Vertical Datum of 1988 (NAVD88) is the preferred vertical datum for PHIDB and the National Geodetic Vertical Datum of 1929 (NGVD29) is also supported.

Coordinate information will be stored in several places and forms in the PHIDB:

- Site boundaries will be stored in the "Site" table as a set of coordinates defining a polygon (or multiple polygons).
- The spatial extent of each field event will be stored in the "Field Event" table as a polygon (or multiple polygons).
- Target sampling location coordinates will be stored as points in the "Location" table.

[3] The term "actual" refers to coordinates of the sample location. (e.g., ccollected with a Differential Global Positioning System)

- The actual locations of field sample collection will be stored in the "Field Collection" table as one or more points, lines, or polygons and correspond to the actual location identifiers in the "Field Event Location" table.

3.3.4 Point, Transect, and Areal Samples

As noted in the previous section, the PHIDB database will store some coordinate information as polygons and some as points. It is possible that some data will also be stored as lines (e.g., if samples are collected on transects). The spatial centroid of any of these types of features can be obtained from the database. In this way, the multiple types of geometries that are supported can all be simplified and represented by a single pair of coordinates representing latitude and longitude.

3.4 SAMPLES

This section briefly describes how the variability of sampling designs is represented in the database. Some of the levels of complexity described here are most commonly (or exclusively) encountered when sampling for chemical analyses, but most of the complexity described here is relevant to sampling conducted for any purpose.

The database tables used to record sampling information are illustrated in detail on Figure 2 and are generalized in Figure 5, and their use is described in the following sections.

3.4.1 Application of Sampling Categories in the PHIDB to Address Variability of Sampling Schemes

A mass or volume of material that is collected from the environment can be subdivided, and even combined with other samples, in several different ways:

- Field Compositing—Material that is collected from multiple locations or at multiple times, sometimes with multiple gear deployments or even different gear, is combined together to make a composite "sample."
- Replication—Gear is deployed multiple times at the same location and (more or less) the same time to produce replicate samples. Replicates may also be taken from a single gear deployment, such as multiple small cores from within a box core, multiple aliquots from a GoFlo bottle, or multiple fish from a trawl. Field replicates are intended to provide information on field variability plus sampling variability. There may be several levels of field replication in a sampling design. Following the terminology outlined below, both collections and samples may be replicated. Also, to some extent, whether or not samples (or collections) are considered to be field replicates depends on the data analysis to be

conducted, and two samples may be treated as field replicates for one analysis but not for another.

- Subsampling—Material collected from a single gear deployment is frequently subdivided. One example is the multiple horizons of a soil or sediment core: the core itself may be considered to be a sample, but the individual horizons are also samples. Another example is tissue samples: a fish that is collected may be regarded as a sample, but individual tissues from that fish may be removed and analyzed separately, so each tissue would be a subsample. Although these examples illustrate only a single level of subsampling, some sampling designs may require several levels or steps of subsampling. Subsamples may be created in the field or afterward (e.g., in an analytical laboratory).
- Duplication for quality control (QC) checks or splitting for analysis by separate parties—For chemical analyses, typically one in every 20 "samples" is split into two "samples" and both of these splits (or duplicate samples) are submitted for laboratory analysis. The purpose of such QC duplicates is to evaluate variability due to transportation and laboratory analyses. Samples may also be split between parties or laboratories to provide confirmatory analyses. The material that is split is presumed to be uniform and homogenous, so that differences between splits do not represent differences in field conditions. For this reason, some types of samples are homogenized or mixed in the field prior to being split.

These different categories of "samples" need to be distinguishable because the differences are important for some types of data analyses. These different logical categories also can have different attributes. The PHIDB table structure supports these distinctions to enhance clarity, interpretability, and integrity of the sampling data. Figure 5 illustrates structural components that provide this support, and the following subsections identify the aspects of samples, sample categorization, and potential distinctions and differing terminologies to help clarify how "samples" are considered in the PHIDB. The subsections discuss:

- The terminology adopted in the PHIDB for the sample categories identified above; and
- Application of this terminology and standardized representation of samples and sample categories in the PHIDB.

3.4.1.1 PHIDB Terminology for Standardized Representation of Sample Categories

The terminology for the sample categories introduced in Section 3.4.1 used with the PHIDB is as follows:

- Field Collection—This term is applied to the material obtained from a gear deployment, or by multiple gear deployments if material is combined or composited. A collection is the mass (or volume) of material that is procured before any subsampling or splitting is done. Separate collections from the same location and (approximately) the same time may be considered to be replicate collections. A collection may be subdivided into multiple samples (field samples, described below). A sediment core is an example of a collection: the entire core is the collection, and individual horizons from the core are field samples. Another example of a collection is a fish trawl: the entire trawl would be the collection, and individual fish, or groups of fish, from the trawl would be distinct field samples. A unique identifier should be assigned to each collection. Often there is no distinction that needs to be made between a collection and a field sample, and in those cases the field sample ID can also be used as the collection ID.

- Field Sample—This term is used to identify the material that will be considered to be representative of a single point in space and time for the purpose of data analysis and interpretation. The term "interpretive sample" can be used to describe the conceptual purpose or use of each field sample; the term "field sample" describes the physical aspect and is used in the database design and implementation. A field sample is the mass of material obtained after any subsampling of the collection has been performed and before any splitting has been performed. Environmental (e.g., chemical) conditions within each field sample are assumed to be uniform. The field sample therefore is the level of detail most often of interest to a data user. In most cases, material from each unique location, date, and depth corresponds to a field sample. If no subsampling is done, the field sample will be the same as the collection. Each field sample must be uniquely identified to allow efficient data summarization and analysis.
- Analytical Sample—This term is used to refer to one of the fractions of material that results from splitting a field sample. These fractions are also referred to as splits or duplicates. Each analytical sample must be uniquely identified because laboratories must analyze and report the data for each split separately. In addition, data validators must be able to distinguish between the results produced by the laboratory for different analytical samples and also be able to identify which analytical samples are splits of the same field sample. When a field sample is not split, then the field sample and the analytical sample are one and the same. When a field sample is split then only the splits (analytical samples) will have a physical representation, and the field sample will be conceptual rather than physical. The field sample remains an important conceptual level because the data from the multiple analytical samples will be combined and used to characterize that field sample for the purpose of data interpretation.
- Most of the common interpretations of the word "sample" are made explicit by this terminology and the corresponding structural representation in the PHIDB.

3.5 IMPLEMENTATION OF THE STANDARDIZED REPRESENTATION OF SAMPLES

Sampling designs differ in complexity and therefore in their representation of material as a collection, interpretive sample, or analytical sample. All environmental sampling types can be accommodated by this standardized representation, however. The detailed implementation of this representation is illustrated in Figure 2 and described in the following sections.

The PHIDB design includes separate tables for collections, interpretive samples, and analytical samples (Figure 5). This structure allows the database system to automatically enforce data integrity rules that cannot be enforced when the information that defines collections, interpretive samples, and analytical samples is all combined into a single sample table. For example, a sediment core is collected at a single location at a single time, and so this information is stored only once in the collection table and applies to all interpretive samples (e.g., horizons) from that core. If the information relating to collections and interpretive samples was combined into a single table, the sampling location and time would be repeated as many times as there are interpretive samples. Using separate tables for collections, interpretive samples, and analytical samples prevents such errors from occurring. When data are summarized for analysis and interpretation, these tables can be joined together by the database software so that they appear to be a single sample table to the data user.

Sample identifier schemes used for field sampling programs do not necessarily use clearly distinct identifiers for collections, interpretive samples, and analytical samples. Frequently a single sample identifier is used, and for chemistry analyses, that corresponds to an analytical sample. While retaining the sample identifiers used in the original investigation, the PHIDB design allows the use of additional identifiers to unambiguously distinguish collections, interpretive samples, and analytical samples. The electronic data deliverable (EDD) Specifications for the PHIDB (Attachment B) outlines distinct identifiers to be recorded for collections, interpretive samples, and analytical samples.

3.5.1 Collections

Collection information is stored in the "Field Collection" table (Table 14). The primary key of the "Field Collection" table consists of two columns: the field event identifier and a collection identifier. The field event identifier must be defined in the "Field Event" table. The relationships of the collections table to other sampling tables are illustrated in the simplified ERD on Figure 5. Descriptions of the columns and column attributes for this table are shown in Table 14. Other important features of the field collection table are described in the following sections.

3.5.1.1 Collection Attributes

Collection locations. The collection location name is identified by a field-event-specific location identifier in the "Event Location" column. The actual coordinates at which the collection was obtained are to be recorded in the "Coll Coords" column in the "Field Collection" table. This column can contain a single pair of latitude and longitude values (a single point), multiple points, a line (e.g., a sampling transect), a set of lines, a polygon (e.g., a sampling area), or multiple polygons.

Additional descriptive text can optionally be recorded in the "Sub-Location" column. This can be useful if, for example, two different types of substrate are present within the area of uncertainty of the positioning method used; the "Sub-Location" value could be used in such a case to record values such as "hard bottom" and "soft bottom."

The area of spatial uncertainty around the target or actual coordinates can be indicated using the "Geo Method," "Horiz Accuracy Measure," and "Horiz Accuracy Units" columns.

Collection date. The collection date and time are required. Time zone information is also stored for the collection date. If the collection time is not known, it will default to 00:00 (midnight) on the specified day.

Collection method. Three columns are available to record different aspects of the collection method used. The "Event Element" column is intended to be used to identify, and distinguish between, planned subsets or elements of the field event. For example, event element values could be used to distinguish between samples collected from a random grid and other samples collected to characterize specific features or locations. No valid values are established for the "Event Element" column, though it is expected that the values used for a particular field event will have been defined in the relevant sampling and analysis plan.

The "Collection Method" column is used to identify the sampling gear used. Valid values are established for this column.

The "Collection Design" column is used to identify the method used to collect material. Examples of different methods include a point sample, a 5-point composite sample, and multi-increment sampling (MIS). Valid values are established for this column.

Collection depths. Upper and lower depths for the collection can be recorded. Depths may be recorded in different units; the units used must be within the established set of valid values. The "Vertical Reference Point" column identifies the zero point for the depth measurement; example values are the water surface, the sediment surface, and a well benchmark. Valid values are established for vertical reference point values. Collection depths may be different from sample depths under some circumstances. For example, when the collection is a sediment core, the lower collection depth should represent the bottom of the core, whereas the upper and lower sample depths should represent the depth of the sample within the core.

Collection material. The environmental material that is collected is recorded in the "Collection Material" column. This value is required and is controlled by a set of valid values. The collection material should be defined to be as specific as necessary to meet the needs for event-specific or overall data analyses. The valid values for collection materials reference a more general set of valid values for materials. For example, collection materials of "fish tissue" and "clam tissue" would share a valid value of "tissue." This two-level hierarchy of material codes allows data to be summarized using either generalized or specific data selection criteria.

Material codes are recorded separately for field collections and field samples because they may differ if there has been any physical fractionation, subsampling, or pretreatment of material in the field. For example, a collection material may be "whole fish" and that collection may be subdivided into two field samples with sample materials of "fillet" and "carcass."

Composites. The "Field Collection" table contains several columns that can be used to identify and describe sample composites. The "Composite Type" column is required and is controlled by a set of valid values. The set of valid values includes values that indicate the collection is not a composite, that it is a spatial composite, or that it is a temporal composite. The column "Composite Count" identifies the number of separate components (e.g., separate sediment grabs) that make up the composite. The columns "Composite Period" and "Composite Period Units" identify the length of time over which a temporal composite was collected.

Composites can also be described by identifying specific field samples that have been combined to create another field sample, as described in a subsequent section.

3.5.1.2 Collection Relationships

The following tables are directly related to Table 14, the "Field Collection" table as shown on Figure 5:

- *"Field Event Location" (Table 12; described in Section 3.3.1):* Each collection must be associated with one field event location identifier. Each field event location identifier may be associated with many collections.
- *"Collection Measurement" (Table 15; described below):* The "Collection Measurement" table allows miscellaneous quantitative information to be recorded for each field collection. A field collection may have zero or more measurements associated with it.
- *"Collection Observation" (Table 13):* The "Collection Observation" table allows miscellaneous observations to be recorded for each field collection.

- *"Field Sample" (Table 16; described in Section 3.5.3):* Each collection may have zero or more field samples associated with it.

The full details of these relationships are shown in Figure 2.

3.5.1.3 Collection Measurements

The "Collection Measurement" table (Table 15) provides a means to record quantitative information that does not have a designated column for it in the "Field Collection" table (Table 14). This information may include conditions at the time and place of sampling, such as pH, turbidity, percent cloud cover or water depth, or sample- or gear-specific measurements such as a core's penetration depth. This table allows the PHIDB to accommodate field data that are in addition to the core requirements of the database system. The structure of this table is shown in Table 15.

Measurements are identified by both a measurement type (e.g., "water depth") and a measurement method (e.g., "depth sounder"). Both of these are controlled by valid value lists. These valid value lists are expected to grow over time, depending on the type of information provided by individual field sampling events.

3.5.2 Collection Observations

The "Collection Observation" table (Table 13) provides a means to record categorical observation about a collection or about conditions at the time of sampling. This includes information such as color, odor, and sea state. The types of categories, and values within each category, are all defined as valid values.

3.5.3 Field Samples

Field samples, which may be subsets of field collections, are described in the "Field Sample" table (Table 16). The primary key of the "Field Sample" table consists of two columns: the field event identifier and a main sample identifier. The field event identifier must be defined in the "Field Event" table. The relationships of the "Field Sample" table to other sampling tables are illustrated in the simplified ERD on Figure 5 and the detailed ERD in Figure 2. Descriptions of all of the columns, and column attributes, for this table are shown in Table 16. Other important features of the "Field Sample" table are described in the following sections.

3.5.3.1 Field Sample Attributes

Collection part and preparation method. A field collection and a field sample may be one and the same, but when there are multiple field samples per collection, the "Collection Part" column describes how the collection has been subdivided to create each field sample. Valid values are defined for the "Collection Part" column; they will include values that indicate that the field sample is the entire collection, a vertical horizon from the (core) collection, a tissue from an organism, or a physical fraction such as a particular grain size class. The "Field Sample" table also includes a column named "Field Prep Method" that can be used to describe any sample preparation or transformation processes that were applied to derive an interpretive sample from a collection.

Sample material. The environmental material making up the sample is recorded in the "Sample Material" column. This value is required and is controlled by a set of valid values. The sample material may be different from the collection material (in the "Field Collection" table); for example, if sediment is collected and porewater is extracted in the field, the sample material and collection material will differ. The sample material should be defined to be as specific as necessary to meet the needs for event-specific or overall data analyses. The valid values for sample materials reference a more general set of valid values for materials. For example, collection materials of "fish tissue" and "clam tissue" would share a valid value of "tissue." This two-level hierarchy of material codes allows data to be summarized using either generalized or specific data selection criteria. The "Field Prep Method" column may also convey information about the process that resulted in a change in the sample material.

Sample date. The date and time (with time zone) at which a field sample was extracted or prepared from a field collection. The sample date may be equal to the collection date or may be later. The sample date is required. If the sample time is not known, it will default to 00:00 (midnight) on the specified day.

Sample depths. Upper and lower depths for the interpretive sample can be recorded. These depths should be relative to the upper depth of the collection and, therefore, also relative to the vertical reference point that applies to the collection. Depths may be recorded in different units; the units used must be within the established set of valid values. The "Vertical Reference Point" column identifies the zero point for the depth measurement; example values are the water surface, the sediment surface, and a well benchmark. Valid values are established for vertical reference point values. Collection depths may be different from sample depths under some circumstances. For example, when the collection is a sediment core, the lower collection depth should represent the bottom of the core, whereas the upper and lower sample depths should represent the depth of the sample within the core.

Sample mass and volume. The mass and volume can optionally be recorded for each sample. A measurement basis ("Dry" or "Wet") must be provided if a weight is recorded for a solid-matrix sample. These values may originate either from field activities or from a laboratory.

Original sample identifier. If the original investigation used a sample identification scheme that is not consistent with the paradigm of collections, interpretive samples, and analytical samples, consistency may be established during data loading by using a modified version of the original sample ID. If this is necessary, the original sample identifier will be recorded in the "Original Sample ID" column.

3.5.3.2 Field Sample Relationships

The following tables are directly related to Table 16, the "Field Sample" table as shown on Figure 5:

- *"Field Collection" (Table 14):* Each field sample must be associated with one field collection.
- *"Sample Composite" (Table 18):* Each field sample may be a composite of other field samples (see below).
- *"Sampling Reason" (Table 19):* There may be multiple reasons for the collection of each field sample (see below).
- *"Sample Measurement" (Table 20):* There may be any number of measurements made on each sample in addition to those that are explicitly incorporated into the data model.
- "Sample Observation" (*Table 17*): The "Collection Observation" table allows miscellaneous observations to be recorded for each field collection.
- *"Analytical Samples" (Table 21):* There should be one or more analytical samples for every field sample if analytical chemistry analyses are performed.

3.5.3.3 Composite Samples

In addition to the ways that composite samples can be identified at the collection level, composite samples can also be defined at the field sample level by explicitly linking different field samples together. If a field sample is formed by compositing other field samples, the "Sample Composite" table (Table 18) can identify all of the subcomposite samples that make up the composite samples. The mass or volume of each subcomposite sample can be recorded.

3.5.3.4 Sampling Reasons

Samples may be collected for various reasons or uses. These include remedial design, site characterization, risk assessment, post-remedial confirmation sampling, and post-remedial monitoring. Some samples may be collected for multiple reasons. The "Sampling Reason" table (Table 19) allows these reasons to be recorded as an aid to data selection and interpretation. Specification of sampling reasons is not required by the database structure. A valid value list is used for reason codes to ensure consistency and specificity.

3.5.3.5 Field Sample Measurements

The "Sample Measurement" table provides a means to record quantitative information pertaining to the sample or conditions at the time of its collection that does not have a designated column for it in the "Field Sample" table. This information may include measurements such as the fork length for fish and the temperature and conductivity of surface water. This table allows the PHIDB to accommodate field data that are in addition to the core requirements of the database system. Details of the structure are shown in Table 20.

When there is only one field sample for a collection, measurement information should be recorded in the "Sample Measurement" table rather than in the "Collection Measurement" table.

Measurements are identified by both a measurement type (e.g., conductivity) and a measurement method (e.g., field conductivity meter). Both of these are controlled by valid value lists.

3.5.4 Field Sample Observations

The "Sample Observation" table (Table 17) provides a means to record categorical observation about a field sample. This includes information such as color and odor. The types of categories, and values within each category, are all defined as valid values. When there is only one field sample for a collection, observations made on the sample should be recorded in the "Field Sample Observation" table rather than in the "Collection Observation" table.

3.5.5 Analytical Samples

Field samples that will have analytical chemistry analyses performed on them should have one or more associated analytical samples. If the field sample is split as part of the quality control program, there will be more than one analytical sample for the field sample. Analytical samples must have unique identifiers, composed of a combination of the "Field Event ID" and the "Analytical Sample ID" columns. The "Analytical Sample ID" may be the same as the corresponding field sample identifier ("Main Sample ID"). This will ordinarily be the case when samples are not split; when samples are split, one of the splits may have the same identifier, and the other may be distinguished by a suffix such as "DUP." Details of the structure are shown in Table 21.

Field samples that are not used for analytical chemistry analyses do not need to have analytical samples associated with them.

3.6 ANALYTICAL CHEMISTRY DATA

The results of chemical analyses performed on field samples are stored in the "Lab Result" table (Table 22). This table allows the analytical results for field samples to be stored, including laboratory replicates and reanalyses. The primary key of this table is composed of the following values:

- "Field Event ID" and "Analytical Sample ID"—These identify the analytical sample on which analyses were performed.
- "Lab SDG" and "Analysis"—The "SDG" value identifies the sample delivery group (SDG) in which the analyses were conducted. When samples are reanalyzed, the reanalysis results will have a different SDG identifier than the original result, so this identifier allows the different results to be distinguished. The "Analysis" value identifies the type of analyses performed, such as metals, semivolatile organics, or pesticides and PCBs. The SDG and analyte group are ordinarily part of the data package identifier used by the laboratory.
- "Analyte"—The chemical element or compound for which results are reported.
- "Preparation Method" and "Analytical Method"—These identify the methods that the laboratory used to conduct the analysis. Different methods may be used for the same analyte depending on concentration levels and precision desired.
- "Material Analyzed" and "Fraction Analyzed"—The "Material Analyzed" value may be the same as the "Sample Material" of the corresponding field sample, or it may be different if the laboratory has performed some kind of fractionation or subsampling of the field sample. The "Fraction Analyzed" distinguishes results reported for total or dissolved fractions of aqueous samples.
- "Lab Rep"—Laboratory quality control programs ordinarily include replicate analysis of five percent of the samples. In addition to their use for quality control purposes, the replicate analyses are an additional measurement that can be used to generate a more precise estimate of true environmental concentrations, and so the PHIDB allows all laboratory replicates to be stored.

In addition to the detailed results provided by analytical laboratories, the PHIDB can also store summarized results, such as calculated totals for PCBs and DDx. Important features of these capabilities are described in the following sections.

3.6.1 Original or "Raw" Data

Result value. Each analytical result is stored as a numerical value, with the number of significant digits and the units both stored in separate columns. Both the number of significant digits and the units are required. Significant digits are required so that they can be propagated through operations such as averaging and summing, and so that the results can be rounded appropriately. Results are not rounded on entry; the database stores as many digits of precision as is provided by the data source. This allows calculations to be carried out with the maximum precision available; results are only (optionally) rounded upon export.

Units are controlled by a set of valid values and the valid values each have an associated dimension and conversion factor. Dimensions are expressed in terms such as mass/mass (e.g., for units of milligrams per kilogram [mg/kg]) and mass/volume (e.g., for units of micrograms per liter [µg/L]), and the conversion factors allow values having different units but the same dimension to be converted to a standard unit during data summarization.

Qualifiers. Data qualifiers reported by the laboratory are recorded in a column named "Lab Qualifiers." If data validation is conducted, the completion of data validation and the data validation level are recorded in the "Validated" and "Validation Level" columns, and the data validation qualifiers (if any) are recorded in the "Result Qualifiers" column. In addition, the three Boolean columns "Undetected," "Estimated," and "Rejected" are used to encode the corresponding attributes of the reported value, whether those are derived from laboratory qualifiers or validation qualifiers. The use of Boolean columns allows for more reliable data selection and more efficient propagation of qualifiers through averaging and summing operations. In addition to the qualifier values, comments specific to the quality assurance review of the data can be stored. If data validation results in revision of the results (e.g., as a result of a change in corrections for laboratory blank samples) the original result can be stored in the "Original Lab Result" column.

Reportable flag. The "Lab Result" table includes a Boolean column to indicate whether or not the data value should be reported or used for analysis. When reanalyses are performed, one of the results may be considered to be more accurate or precise than another, and so preferred for analysis. The less accurate or precise value would have its "Reportable" attribute set to False.

3.6.2 Summarized Data

Calculated results such as total PCBs are stored in the "Calculated Chemistry" table (Table 23). Calculated results are stored separately from original laboratory data for the following reasons:

- Addition, recalculation, and deletion of summarized data will not cause any changes to the "Lab Result" table, reducing the possibility of accidental data alterations.
- Sums and other calculated results are expected to be summarized up to the level of the field sample, and consequently the "Calculated Chemistry" table should have (and does have) a different primary key than is used in the "Lab Result" table.
- The "Calculated Chemistry" table includes a column describing the calculation method. This column is part of the primary key and allows sums to be calculated and stored in

different ways. For example, different summing methods may be used when conducting ecological risk assessments and human health risk assessments.

This structure accommodates more than one scheme for calculating results. Different calculation methods can be established to calculate the same quantity (e.g., total PCB congeners) in different ways, if needed. In addition to the field sample ID and the calculation method, the primary key of the "Calculated Chemistry" table includes the material analyzed, the fraction analyzed, and the analyte. Qualifiers (as Boolean columns), units, and the measurement basis are also carried over from the "Lab Result" table to the "Calculated Chemistry" table. The calculated_chemistry table contains the four Boolean columns: "undetected," "estimated," "rejected," and "limited analytes." The first three of these qualifiers should be propagated through the process of summing individual analytical results up to the interpretive sample level, and the fourth qualifier should be applied if the number of analytes used for the sum is limited, as defined in Table 3 of the Portland Harbor Data Management Plan.

3.7 TOXICITY TEST DATA

Toxicity tests conducted on environmental samples are ordinarily carried out in batches. The structure for storage of toxicity test data in the PHIDB reflects this: one table is used to describe each batch ("Tox Test Batch"), and a second table is used to describe the results for each environmental sample within a batch ("Tox Test Result"). Complete descriptions of the columns in these database tables are shown in Tables 24 and 25. Important characteristics of a toxicity test batch are listed below.

Batch identifier. Each toxicity test batch is uniquely identified by a combination of the name of the laboratory conducting the test ("Lab Name") and a laboratory-specific batch identifier ("Tox Test Batch").

Toxicity test type and organism. Each batch is characterized by a single toxicity test type. Test types include, for example, the amphipod 10-day bioassay, the amphipod 28-day bioassay, the 72-hour echinoderm fertilization bioassay, and the *Neanthes* 20-day bioassay. The toxicity test type is recorded in the "Tox Test Type" column, which is controlled by a valid value list. Different species may be used for some bioassays—for example, different amphipod species are used in marine and brackish waters. The species used and its life stage are recorded in the "Taxon" and "Life Stage" columns, respectively.

Each toxicity test result is linked to the corresponding batch by the lab name and batch identifier. Other important characteristics of each toxicity test result are listed below.

Field sample. Toxicity test results are recorded for each field sample. Toxicity testing procedures do not ordinarily include splitting of field samples. The field sample is identified by the combination of the field event ID and the main sample ID.

Variables and measurements. Different types of results can be obtained from some types of toxicity tests. For example, the amphipod test can be used to measure either survival or growth. The type of result is recorded in the "Tox Test Variable" column, which is controlled by a valid value list. In addition, different types of measurements can be reported for some types of results. For example, the results of an amphipod survival test can be reported as either the number of amphipods alive at the conclusion of the test, as the percent mortality, or as the percent survival. For another example, growth can be measured either as the increase in mass of the organism(s) or as the percentage change in mass. The type of measurement used to report the results is specified in the "Tox Test Measurement" column, which is controlled by a valid value list.

Replicates. Toxicity testing protocols frequently require multiple measurements to be made for each sample, to allow an assessment of test precision. The "Replicate" column is used for the replicate identifier that distinguishes multiple measurements on the same sample.

3.8 SPECIES DATA

Species abundance data—specifically, sediment infauna abundances—are part of the sediment quality triad and may be collected during future studies in the PHSS. The structure for storing species abundance data is simple, consisting of a single table ("Species Abundance") that is linked to the "Field Sample" table. A complete description of the columns in the "Species Abundance" table is shown in Table 26. Important characteristics of species abundance data are listed below.

Species observed. Information on the type of organism for which an abundance measurement is recorded is represented in three columns in the "Species Abundance" table. These columns identify the taxon, the sex, and the life stage of the organism(s) found. All three of these values are required and are controlled by valid value lists. The identity of the organism observed is recorded in the "Taxon" column, and the lookup table for this value includes common elements of the taxonomic hierarchy from kingdom through subspecies, and it can accommodate a common name and the Integrated Taxonomic Identification System serial number.

Abundance value. A taxon's abundance is recorded as a numerical value with an associated unit code. The unit code is controlled by a valid value list. Abundances may be recorded not only as counts of organisms, but also as densities, such as the number of organisms per volume or per area.

The "Species Measurement" (Table 27) stores the taxon, sex, life stage, for a given species. Measurements are identified by measurement method (e.g., measuring board). Several columns are controlled by valid value lists. These valid value lists may grow over time, depending on the type of information provided by individual field sampling events.

3.9 WELL DESCRIPTIONS

The characteristics of wells used to collect groundwater samples can be recorded in the "Well" table. A complete description of the columns in the "Well" table is shown in Table 28. Important characteristics of well descriptions are listed below.

Well hole depth measure. The total depth of the well.

Well hole depth measure unit. The units for the well hole depth measure. The well hole depth measurement units apply to the screen depths as well.

Screened interval. The upper and lower depths of the screened interval(s) in water-bearing formations.

Well completion date. The date that the well was developed.

Well usage. The intended purpose of the well (e.g., drinking water, groundwater monitoring).

Vertical reference point. A description of the vertical reference point that should be used for water depth measurements (e.g., well benchmark, top of casing).

Well status and status date. An indication of whether the well is in use or has been abandoned. The date of the status assessment can also be recorded.

3.10 CRITERIA

Chemical criteria (e.g., remedial action levels [RALs]) that relate to PHSS are stored in the "Criteria" and "Criteriadef" tables (Table 29 and 30). The Criteria codes are stored in "Criteria," which contains the analytes, numeric values, and measurement basis for a given criterion. The citations for the criteria are stored in the "Criteriadef" table. Important characteristics of chemical criteria are listed below.

Criteria code. Each set of related criteria are identified, and distinguished from other sets of criteria, by a concise identifier. Different sets of remedial action levels for Portland Harbor project areas should each have its own criteria codes.

Material. Each criterion can be characterized by the material to which it applies, where the material may be sediment, groundwater, or surface water.

Resource. Each criterion can be characterized by the resource that it is intended to protect, where the resource may be human health, the benthic community, or other aquatic organisms.

Analyte. The chemical element or compound.

Criterion value. The concentration or other numerical value of the criterion.

Units. The concentration units for the criterion.

4 IMPLEMENTATION

4.1 DATABASE MANAGEMENT SYSTEM

The PHIDB database will be implemented using the PostgreSQL database management system (Postgres; https://www.postgresql.org/). Postgres is the selected candidate DBMS because:

- The PostGIS extension (http://postgis.org/) provides the most comprehensive set of features available for storage and manipulation of spatial data in a database management system.
- Postgres supports cascading updates and deletions through multiple pathways between tables, unlike other common client-server DBMS.
- Postgres allows the creation of custom data types and aggregate functions that simplify the propagation of qualifiers and significant digits while averaging and summing data.

The database will be implemented on the same Digital Ocean (https://www.digitalocean.com/) cloud server that currently hosts the Portland Harbor Environmental Data Portal. This will facilitate linking of documents to data if this feature is desired.

4.2 SPATIAL DATA SUPPORT

The specialized internal representation of spatial data that is supported by Postgres/PostGIS (i.e., multiple points, lines, or polygons in a single database column) is not supported by other software such as Microsoft Excel and Microsoft Access that will be used to transfer data to or from the PHIDB. Spatial data in other software will be represented in the following formats:

- Two columns for the latitude and longitude of the geographic centroid, in decimal degrees based on the WGS84 spheroid (SRID 4326).
- A text column containing the complete spatial geometry in the standard well-known text (WKT) format (https://www.ogc.org/standards/wkt-crs).

The numerical latitude and longitude values will allow spatial data to be used (e.g., viewed in tabular form or displayed on a map) when they are stored in alternate formats. The WKT value will preserve fidelity of the entire spatial representation when data are transferred to or from Postgres/PostGIS—potentially including Postgres/PostGIS databases other than the PHIDB.

4.3 CONVENTIONS

The following sections describe some specific decisions that have been made about the approaches that will be used for the implementation of the PHIDB data model in Postgres.

4.3.1 Schemas

Tables, views, and functions in a Postgres database can be created in different schemas within the database. Schemas are useful to segregate different types of functionality within a database. By default, in Postgres, tables, views, and functions are created in a schema named "public." However, for PHIDB, these objects will instead be created in a schema named "idb" (meaning "interim database," or perhaps ultimately, "integrated database"). The reasons for this are:

- *Security*. By default, all individuals who have access to the database have access to the "public" schema. Although these permissions will be removed from the PHIDB database, placing important data in a different schema, with a separate set of access controls, provides additional security over data access.
- *Separation of functionality*. The PostGIS extension, and other extensions that may be used, add data tables and functions to the "public" schema. By separating these components from the other components that are specific to the PHIDB data model, either set of components can be more easily revised or replaced without risk of affecting the other.

The tables used for the Portland Harbor Environmental Data Portal are also in a schema named "idb," and consistency in the schema name used will facilitate linking document information and data, if this feature is to be implemented.

4.3.2 Table and Column Names

Tables and column names in the attached data dictionary are identified using names with mixed case for readability. Table and column names in Postgres are case-sensitive, and Postgres automatically converts all table and column names to lowercase unless they are double-quoted. Double-quoting table and column names when writing structured query language (SQL) to load and summarize data decreases productivity and increases error rates, so all table and column names in PHIDB will be created using names that are all lowercase.

The names for all the tables that make up the core of the data model (Figure 5) will be prefixed with "d_", and the names of all lookup tables containing valid values will be prefixed with "l_". The complete data and lookup tables names and format specifications are provided in the Figure 2. These prefixes clarify the purpose and use of tables in SQL scripts, and also allow data and lookup tables to be easily distinguished in visual interfaces to the database. The most up-to-date data dictionary and valid value lists can be downloaded from the Portland Harbor Environmental Data Portal.

4.3.3 Column Widths

Column widths for identifiers (the field event ID, the collection ID, the main sample ID, etc.) are all specified as 50 characters. This is anticipated to be wide enough to accommodate identifiers typically used in environmental sampling programs. As appropriate, this width specification might be incorporated into guidance for developing sampling and analysis plans for the PHSS.

Column widths for most lookup values are specified as 24 characters. This is wide enough so that easily interpretable codes can be used for most quantities, including analytes.

4.3.4 Revision Tracking

Every database table will include columns that record the date and time at which a row was added (created), the date and time at which a row was edited, and the name of the data manager who made the latest change. These values will be automatically filled in as data are changed, and are therefore guaranteed to be correct and complete. This is internal tracking information only and will not be imported from EDDs or exported to Access or any other data formats.

4.4 USERS AND PERMISSIONS

A separate User Guide (Attachment C) contains specific information for users and permissions for accessing the database to download data.

Postgres' security model is based on users and groups. Users are individuals who can log in; groups are collections of users. Permissions to access or modify database objects (tables, views, and functions) may be assigned to both users and groups. Users inherit the permissions of all groups that they are members of.

This security model will be used as follows:

- Users will be required to log in using a password.
- No permissions will be assigned to individual users. Users will have only the permissions of the group that they are members of.
- Two groups will be used, one for database managers and one for database users. The database manager's group will have permissions to create tables, views, and functions, and to modify data in all data tables. The database user's group will have only read-only permissions on tables, views, and functions.

These standards will prevent unauthorized modifications to data in the PHIDB.

FIGURES

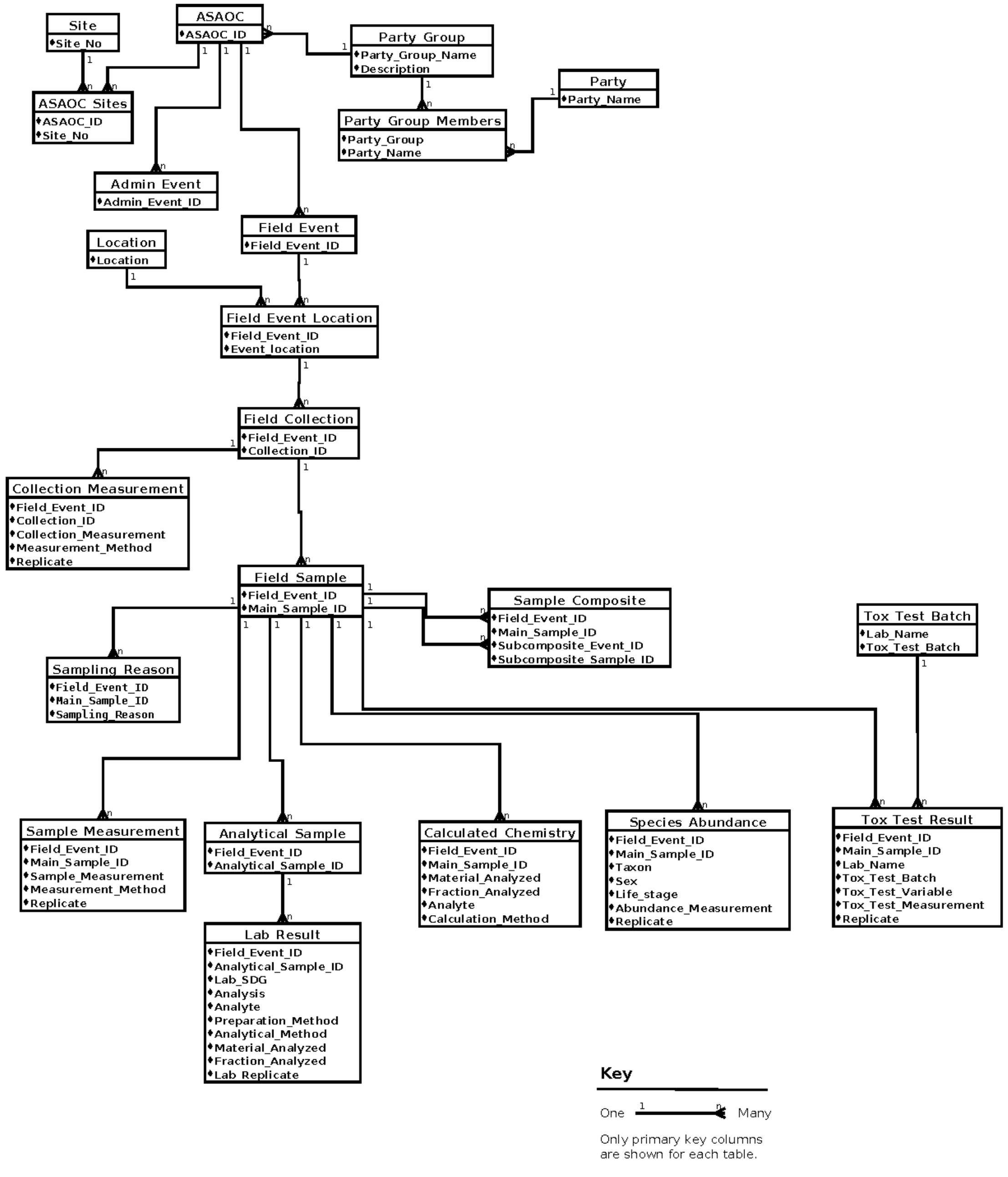
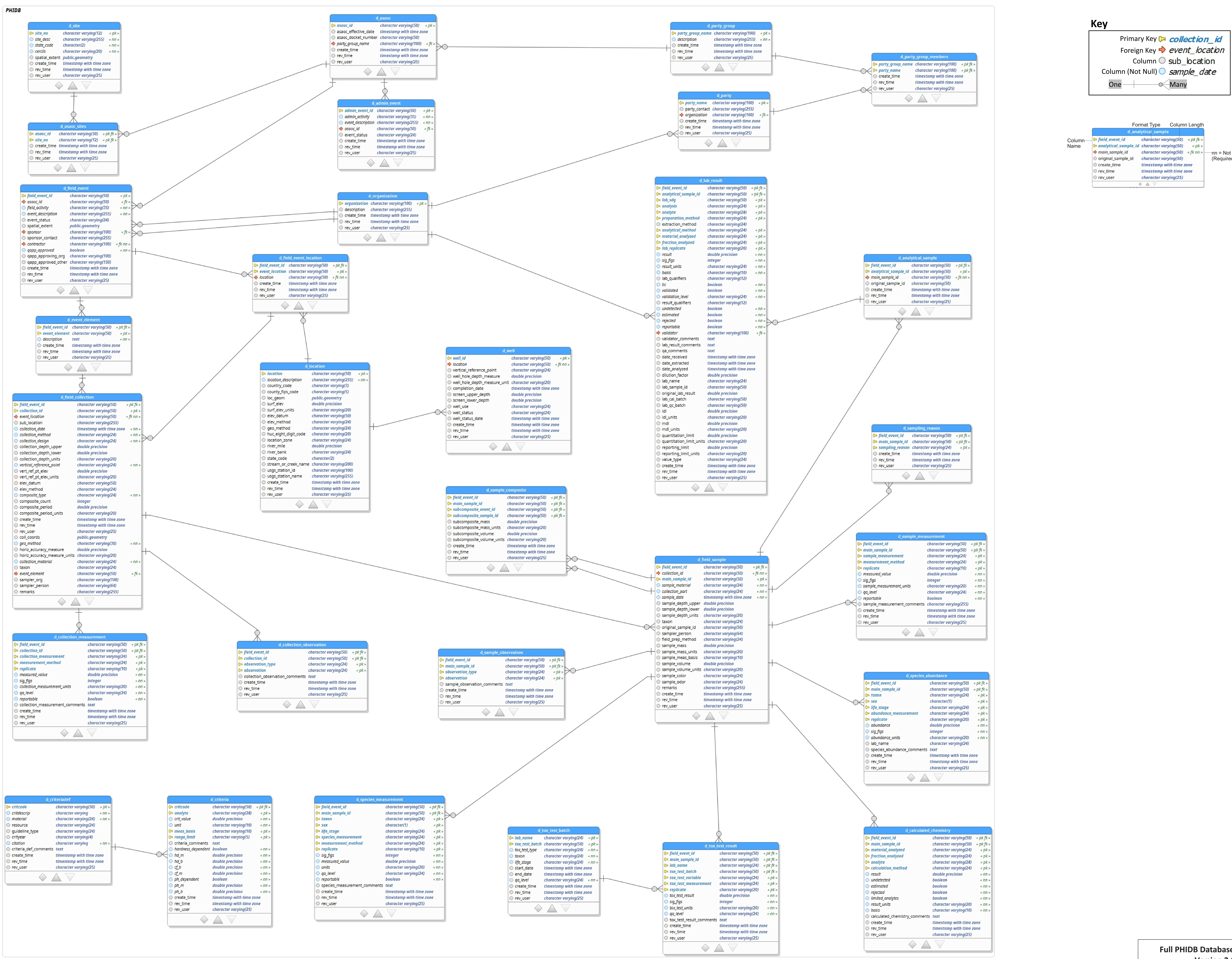


Figure 1
Overview of Database Structure



Full PHIDB Database Structure
Version 2.0
Entity Relationship Diagram

Figure 2

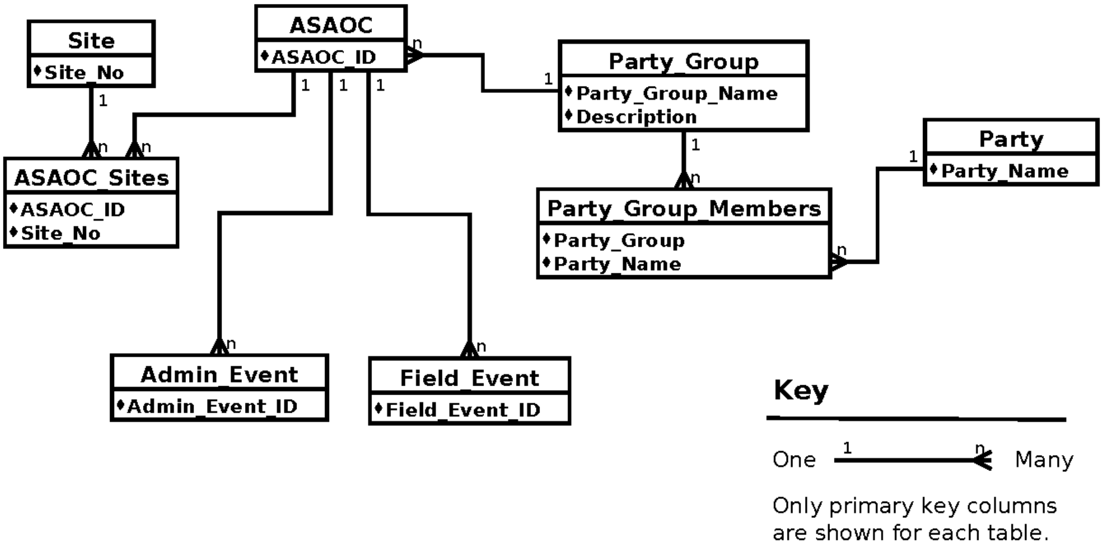


Figure 3
Overview of Administrative Tables Structure

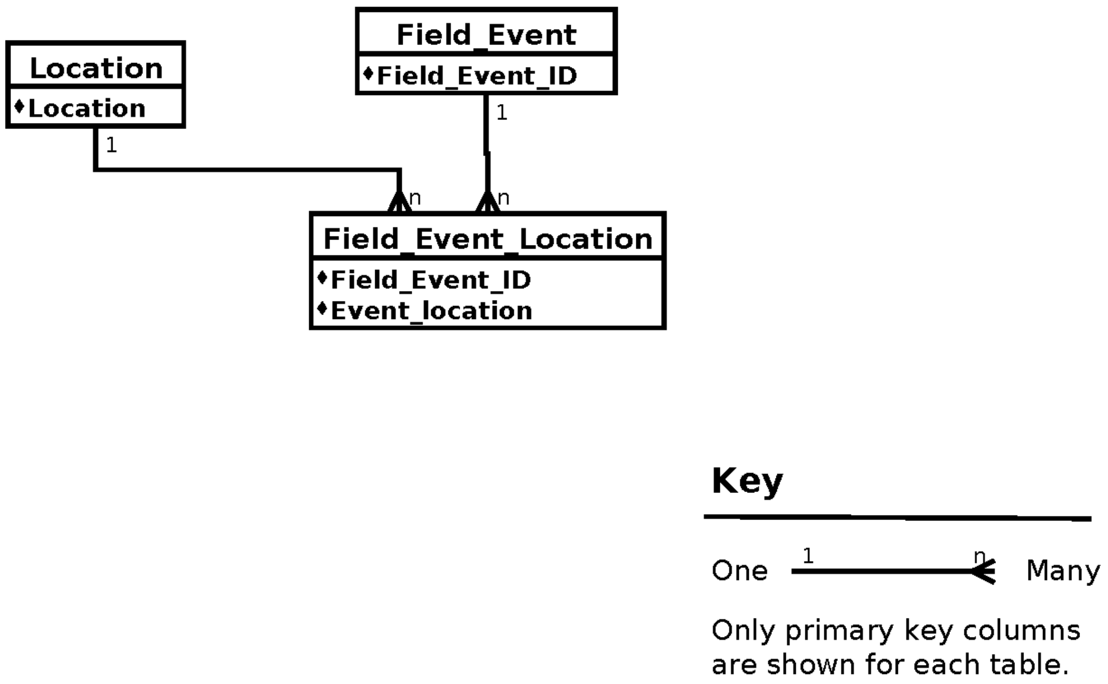


Figure 4
Overview of Locations
Tables Structure

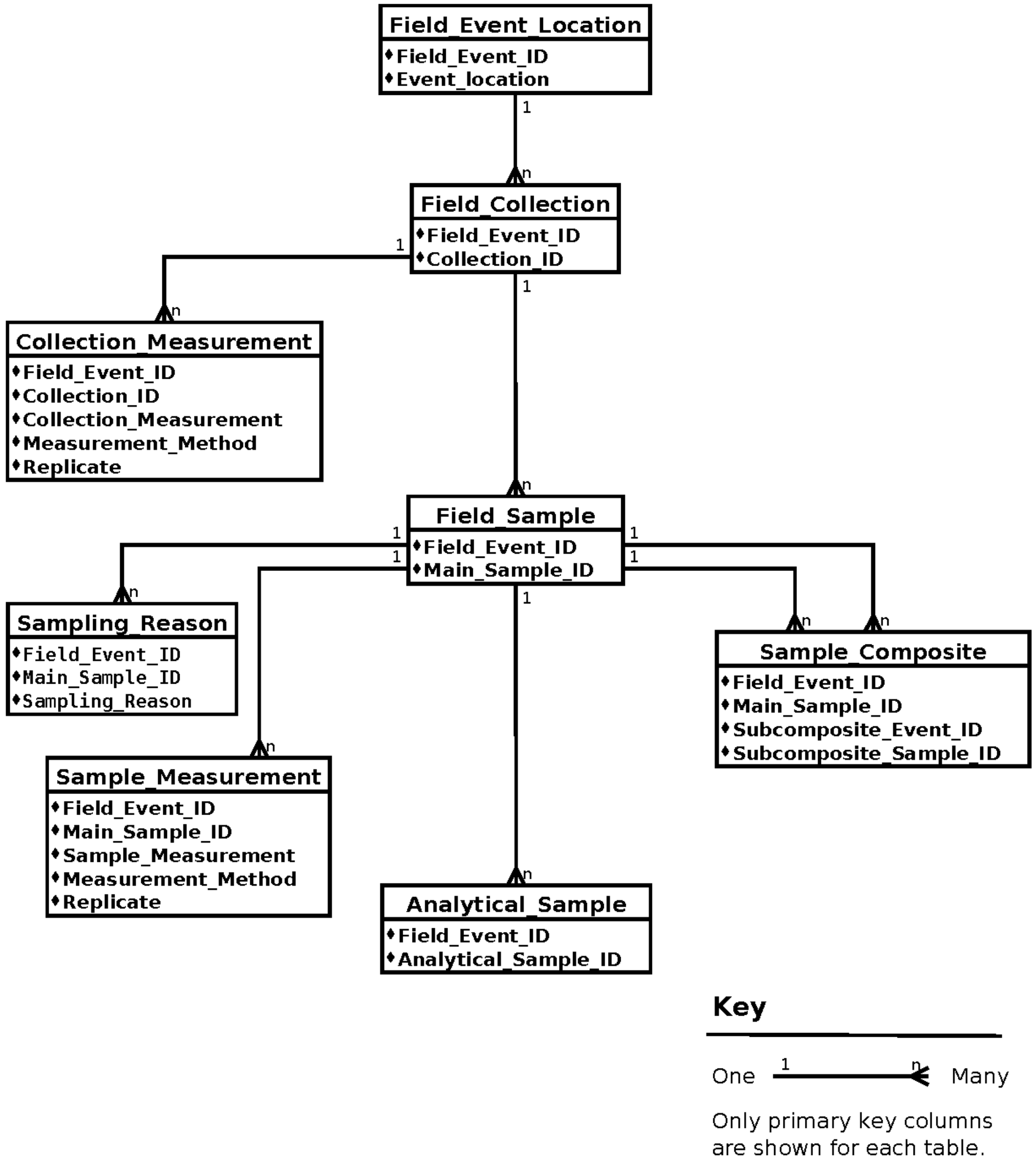


Figure 5
Overview of Sample Tables Structure

TABLES

Table 1 - ASAOC

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
ASAOC	ASAOC_ID	Unique ID for an Administrative Settlement Agreement and Order on Consent (AOC)	Text	50	Yes	No	**PK**	
	AOC_Effective_Date	Date AOC becomes effective	DateTime	0	No	No	No	
	AOC_Docket_Number	Associated AOC Docket Number	Text	50	No	No	No	
	Party_Group_Name	Group name of the parties associated with this ASAOC	Text	100	No	No	No	Party_Group

Table 2 - Site

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Site	Site_No	Site identifier	Text	12	Yes	No	**PK**	
	Site_Desc	Site description	Text	255	Yes	No	No	
	State_Code	Two-character state abbreviation	Text	2	Yes	Yes	No	State
	Spatial_Extent	A polygon or multipolygon representation of the spatial boundary of the site	Geometry	0	Yes	No	No	
	CERCLIS	The CERCLIS identifier used by the Superfund program	Text	20	Yes	No	No	

Table 3 - ASAOC Sites

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
ASAOC_Sites	ASAOC_ID	Unique ID for an Administrative Settlement Agreement and Order on Consent (AOC)	Text	50	Yes	No	**PK**	ASAOC
	Site_No	Site identifier	Text	12	Yes	No	**PK**	Site

Table 4 - Party

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Party	Party_Name	Unique name of the party	Text	100	Yes	No	**PK**	
	Party_Contact	Contact name for the party	Text	255	No	No	No	
	Organization	Organization of the party	Text	100	Yes	No	FK	Organization

Table 5 - Party Group

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Party_Group	Party_Group_Name	A name to identify a group of parties that are associated with one or more ASAOCs	Text	100	Yes	No	**PK**	
	Description	A description of the Group	Text	255	Yes	No	No	

Table 6 - Party Group Members

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Party_Group_Members	Party_Group_Name	A name to identify a group of parties that are associated with one or more ASAOCs	Text	100	Yes	No	**PK**	Party_Group
	Party_Name	Name of the party	Text	100	Yes	No	**PK**	Party

Table 7 - Organization

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Organization	Organization	Concise identifier for the organization	Text	100	Yes	Yes	**PK**	
	Description	Detailed description of the organization	Text	255				

Table 8 - Field Event

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Field_Event	Field_Event_ID	Concise name or identifier for the field sampling effort. For example: 2012_LBCR Predesign Invest	Text	50	Yes	No	**PK**	
	ASAOC_ID	Identifier for the ASAOC that this field event is conducted to support	Text	50	Yes	No	No	ASAOC
	Field_Activity	The type of activity to be conducted (e.g., remedial design sampling, confirmation sampling, post-closure monitoring)	Text	35	Yes	Yes	No	Field_Activity
	Event_Description	Description of the event. For example: Lower Burke Canyon Repository Predesign Investigation	Text	255	Yes	No	No	
	Event_Status	Event completion status (e.g., pending, underway, completed)	Text	24	No	Yes	No	Event_Status
	Spatial_Extent	A point, multipoint, line, multiline, polygon, or multipolygon representation of the spatial extent of the event. Although not strictly required, this should be provided if the sampling locations are constrained or defined by a spatial boundary such as a remedial design area or a removal area.	Geometry	0	No	No	No	
	Sponsor	Name of the organization on whose behalf the field event is undertaken	Text	100	Yes	No	No	Organization
	Sponsor_Contact	Contact information for the field event sponsor	Text	255	No	No	No	
	Contractor	Full name of the company contracted by the sponsor to conduct or manage the sampling	Text	100	Yes	No	No	Organization
	QAPP_Approved	Indicates if the QAPP has been approved for the sampling effort	Boolean	0	Yes	No	No	
	QAPP_Approving_Org	Identifies the agency or other organization that approved the QAPP for the sampling effort	Text	100	No	No	No	
	QAPP_Approved_Other	Other information about approval of the QAPP	Text	150	No	No	No	

Table 9 - Event Element

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Event_Element	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	Yes	**PK**	Field Event
	Event_Element	Identifier for a subset of the sampling effort (e.g., Phase 1, Phase 2, background sampling, random grid sampling)	Text	50	Yes	No	**PK**	
	Description	Description of the event	Text	255	Yes	No	No	

Table 10 - Admin Event

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Admin_Event	Admin_Event_ID	Concise name or identifier for the administrative event	Text	50	Yes	No	**PK**	
	Admin_Activity	The type of activity to be conducted (e.g., remedial design, remedial action, emergency response)	Text	35	Yes	Yes	No	Admin_Activity
	Event_Description	Description of the event. For example: Lower Burke Canyon Repository Predesign Investigation	Text	255	Yes	No	No	
	ASAOC_ID	Unique ID for an Administrative Settlement Agreement and Order on Consent (ASAOC). Although not required by the database design, this value should be provided if the event is controlled or specified by an ASAOC.	Text	50	No	No	No	ASAOC
	Event_Status	Event completion status (e.g., pending, underway, completed)	Text	24	No	Yes	No	Event_Status

Table 11 - Location

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Location	Location	Target sampling location identifier	Text	50	Yes	No	**PK**	
	Location_Description	Narrative description of the location	Text	255	Yes	No	No	
	Country_Code	Country code	Text	5	No	Yes	No	Country_Code
	County_FIPS_Code	County code	Text	5	No	Yes	No	Fips_Code
	Loc_Geom	Nominal (target) location, as a point, multipoint, polygon, or multipolygon geometry. Coordinates are stored as decimal degrees in WGS84 (SRID=4326). This value must be unique on every row (i.e., this is a candidate key).	Geometry	0	Yes	No	No	
	Elev_Datum	Datum used to determine the elevation measurement. (e.g., NAVD88; NGVD29)	Text	50	No	Yes	No	Elev_Datum
	Elev_Method	Method used to determine the elevation measurement. (e.g., Altimetry; GPS; Interpolation; Survey)	Text	24	No	Yes	No	Elev_Method
	Geo_Method	Geopositioning method used to establish latitude and longitude coordinates (e.g., GPS; Interpolation; Survey)	Text	24	Yes	Yes	No	Position_Method
	Huc_Eight_Digit_Code	Eight digit USGS HUC code. Equals Sub_Basin	Text	20	No	No	No	
	Location_Zone	Location categorization (e.g., subtidal, intertidal, upland, facility)	Text	24	No	Yes	No	Zone_Type
	River_Mile	River mile, to at least tenths of a mile	Numeric	0	No	No	No	
	River_Bank	Code to specify whether the location is near a river bank or in the navigation channel	Text	24	No	Yes	No	Riverbank
	State_Code	State code - 2 character state abbreviation	Text	2	No	Yes	No	State
	Stream_or_Creek_Name	Name of the waterbody represented by this location	Text	200	No	Yes	No	
	Surf_Elev	The ground elevation of a geographic point where samples or field measurements are collected	Double precision	0	No	No	No	
	Surf_Elev_Units	Surface elevation units (e.g., feet; meters)	Text	20	No	Yes	No	Unit
	USGS_Station_ID	USGS location identifier	Text	100	No	Yes	No	
	USGS_Station_Name	USGS location name	Text	255	No	Yes	No	

Table 12 - Field Event Location

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Field_Event_Location	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Event
	Event_location	Event-specific location code	Text	50	Yes	No	**PK**	
	Location	Sampling location code/monitoring location code	Text	50	Yes	No	No	Location

Table 13 - Collection Observation

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Collection_Observation	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	
	Collection_ID	Unique collection identifier for each core, grab, or group of related field samples	Text	50	Yes	No	**PK**	Field_Collection
	Observation_Type	The type of collection observation made (e.g., tide stage, sediment class)	Text	24	Yes	Yes	**PK**	Observation
	Observation	The observation made during collection (e.g., flood, silt)	Text	24	Yes	Yes	**PK**	Observation
	Collection_Observation_Comments	Comments on the collection observation	Text	255	No	No	No	

Table 14 - Field Collection

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Field_Collection	Field_Event_ID	Concise identifier for the field sampling effort.	Text	50	Yes	No	**PK**	Field_Event_Location
	Collection_ID	Unique collection identifier for each core, grab, or group of related field samples.	Text	50	Yes	No	**PK**	
	Event_Location	Event-specific location code	Text	50	Yes	No	No	Field_Event_Location
	Sub_Location	Narrative description of any systematic deviation or difference from the nominal location	Text	255	No	No	No	
	Collection_Date	Date and time of acquisition of the collection material. This is often the GPS time recorded.	DateTime	0	Yes	No	No	
	Collection_Method	Sample collection method (e.g., grab, core)	Text	24	Yes	Yes	No	Coll_Method
	Collection_Design	Sample collection design or scheme (e.g., single-point grab, spatial composite, temporal composite, MIS)	Text	24	Yes	Yes	No	Coll_Design
	Collection_Depth_Upper	Upper depth of the collection	Double precision	0	No	No	No	
	Collection_Depth_Lower	Lower depth of the collection	Double precision	0	No	No	No	
	Collection_Depth_Units	Units for the upper and lower collection depths	Text	20	No	Yes	No	Unit
	Composite_Type	The compositing method used for the collection (e.g., single, spatial, depth, temporal)	Text	24	Yes	Yes	No	Composite_Type
	Composite_Count	The number of other collections that were composited to create this collection	Integer	0	No	No	No	
	Composite_Period	The length of time over which a temporal composite was conducted	Double precision	0	No	No	No	
	Composite_Period_Units	The units for the composite period	Text	20	No	Yes	No	Unit
	Vertical_Reference_Point	The surface from which depths are measured (e.g., water surface, sediment surface, soil surface, well benchmark)	Text	24	Yes	Yes	No	Vert_Ref_Pt
	Vertical_Ref_Pt_Elev	Elevation of the reference point	Double precision	0	No	No	No	
	Vertical_Ref_Pt_Elev_Units	Units of the reference point elevation (e.g. feet, meters)	Text	20	No	No	No	
	Elev_Datum	Datum used to determine the elevation measurement. (e.g., NAVD88; NGVD29)	Text	50	No	Yes	No	Elev_Datum
	Elev_Method	Method used to determine the elevation measurement. (e.g., Altimetry; GPS; Interpolation; Survey)	Text	24	No	Yes	No	
	Coll_Coords	The actual coordinates at which the collection was made, as a point, polygon, multipoint, or multipolygon geometry. Coordinates are stored as decimal degrees in WGS84 (SRID=4326).	Geometry	0		No	No	
	Geo_Method	Geopositioning method used to establish latitude and longitude coordinates.	Text	30	Yes	Yes	No	Position_Method
	Horiz_Accuracy_Measure	Horizontal accuracy measurement	Double precision	0	No	No	No	
	Horiz_Accuracy_Measure_Units	Horizontal accuracy measurement units	Text	20	No	Yes	No	Unit
	Collection_Material	The type of material collected	Text	24	Yes	Yes	No	Sample_Material
	Taxon	Taxon code for organisms	Text	24	No	Yes	No	Taxon
	Event_Element	Identifier for a subset of the sampling effort (e.g., Phase 1, Phase 2, background sampling, random grid sampling)	Text	50	No	No	No	
	Sampler_Org	Name of organization that collected the sample	Text	100	Yes	Yes	No	
	Sampler_Person	Name of the person who collected the sample	Text	64	No	No	No	
	Remarks	Comments on sample collection	Text	255	No	No	No	

Table 15 - Collection Measurement

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Collection_Measurement	Field_Event_ID	Concise identifier for the field sampling effort.	Text	50	Yes	No	**PK**	Field_Collection
	Collection_ID	Unique collection identifier for each core, grab, or group of related field samples.	Text	50	Yes	No	No	Field_Collection
	Collection_Measurement	The type of measurement made (e.g., tide height, cloud cover, number of grabs)	Text	24	Yes	Yes	**PK**	Measurement
	Measurement_Method	The method used to make the sample measurement	Text	24	Yes	Yes	**PK**	Meas_Method
	Replicate	Replicate identifier to distinguish multiple measurements (e.g., “1”, “2”)	Text	10	Yes	No	**PK**	
	Measured_Value	The numerical result of the measurement	Double precision	0	Yes	No	No	
	Sig_Figs	Significant digits of the result	Integer	2	Yes	No	No	
	Collection_Measurement_Units	Result unit of measurement	Text	20	Yes	Yes	No	Unit
	QA_Level	Level of data quality review used	Text	24	Yes	Yes	No	QA_Level
	Reportable	Flag to distinguish reportable and non-reportable results based on data quality review	Boolean	0	Yes	No	No	
	Collection_Measurement_Comments	Comments on the measured value	Text	255	No	No	No	

Table 16 - Field Sample

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Field_Sample	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Collection
	Collection_ID	Unique collection identifier for each core, grab, or group of related field samples	Text	50	Yes	No	No	Field_Collection
	Main_Sample_ID	Unique sample identifier for this interpretive sample	Text	50	Yes	No	**PK**	
	Sample_Material	The type of material sampled	Text	24	No	Yes	No	Sample_Material
	Collection_Part	The fraction of the collection represented by this sample (e.g., entire, vertical horizon, filtered fraction)	Text	24	Yes	Yes	No	Subsample_Type
	Sample_Date	Date and time of creation of the sample for a composite or single sample. The sample date may be the same as the collection date if the collection is not subdivided into multiple samples.						
	Sample_Depth_Upper	Sample upper depth relative to the zero point of the collection	Double precision	0	No	No	No	
	Sample_Depth_Lower	Sample lower depth relative to the zero point of the collection	Double precision	0	No	No	No	
	Sample_Depth_Units	Sample depth units	Text	20	No	Yes	No	Unit
	Taxon	Taxon code for organisms	Text	24	No	Yes	No	Taxon
	Original_Sample_ID	Original sample ID, if the Sample_ID is a corrected value	Text	50	No	No	No	
	Sampler_Person	Name of person who collected the sample	Text	64	Yes	Yes	No	
	Field_Prep_Method	Sample preparation or treatment method carried out at the time of collection (filtering, fractionation)	Text	24	No	Yes	No	Prep_Method
	Sample_Mass	The mass of this sample	Double precision	0	No	No	No	
	Sample_Mass_Units	The units for the sample mass	Text	20	No	Yes	No	Unit
	Sample_Mass_Basis	"Wet" or "Dry" for sample masses	Text	10	No	Yes	No	Meas_Basis
	Sample_Volume	The volume of this sample	Double precision	0	No	No	No	
	Sample_Volume_Units	The units for the sample volume	Text	20	No	Yes	No	Unit
	Sample_Color	The color of the sample upon collection	Text	24	No	Yes	No	Color
	Sample_Odor	The odor of the sample upon collection	Text	24	No	Yes	No	Odor
	Remarks	Sample remarks	Text	255	No	No	No	

Table 17 - Sample Observation

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Sample_Observation	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	Yes	**PK**	Field Sample
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50			**PK**	Field Sample
	Observation_Type	The type of collection observation made (e.g., tide stage, sediment class)	Text	24			**PK**	Observation
	Observation	The observation made during collection (e.g., flood, silt)	Text	24			**PK**	Observation
	Sample_Observation_Comments	Description of the event	Text	255	Yes	No	No	

Table 18 - Sample Composite

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Sample_Composite	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for the interpretive sample that is a composite	Text	50	Yes	No	**PK**	Field_Sample
	Subcomposite_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Subcomposite_Sample_ID	Unique sample identifier for one of the interpretive samples that makes up the composite	Text	50	Yes	No	**PK**	Field_Sample
	Subcomposite_Mass	The mass of this subcomposite sample used in the composite	Double precision	0	No	No	No	
	Subcomposite_Mass_Units	The units for the subcomposite mass	Text	20	No	Yes	No	Unit
	Subcomposite_Volume	The volume of this subcomposite used in the composite	Double precision	0	No	No	No	
	Subcomposite_Volume_Units	The units for the subcomposite volume	Text	20	No	Yes	No	Unit

Table 19 - Sampling Reason

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Sampling_Reason	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for this interpretive sample	Text	50	Yes	No	**PK**	Field_Sample
	Sampling_Reason	Description of the purpose for collection of this sample	Text	24	Yes	Yes	**PK**	Sample_Reason

Table 20 - Sample Measurement

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Species_Measurement	Field_Event_ID	Concise identifier for the field sampling effort.	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50	Yes	No	**PK**	Field_Sample
	Sample_Measurement	The type of measurement made (e.g., length)	Text	24	Yes	Yes	**PK**	Measurement
	Measurement_Method	The method used to make the sample measurement	Text	24	Yes	Yes	**PK**	Meas_Method
	Replicate	Replicate identifier to distinguish multiple measurements (e.g., “1”, “2”)	Text	10	Yes	No	**PK**	
	Measured_Value	The numerical result of the measurement	Double precision	0	Yes	No	No	
	Sig_Figs	Significant digits of the result	Integer	2	Yes	No	No	
	Sample_Measurement_Units	Result unit of measurement	Text	20	Yes	Yes	No	Unit
	QA_Level	Level of data quality review used	Text	24	Yes	Yes	No	QA_Level
	Reportable	Flag to distinguish reportable and non-reportable results based on data quality review	Boolean	0	Yes	No	No	
	Comments	Comments on the measured value	Text	255	No	No	No	

Table 21 - Analytical Samples

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Analytical_Sample	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Analytical_Sample_ID	Unique sample ID for an analytical sample (split or duplicate) that is part or all of an interpretive sample	Text	50	Yes	No	**PK**	
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50	Yes	No	No	Field_Sample
	Original_Sample_ID	Original identifier for the analytical sample	Text	50	Yes	No	No	

Table 22 - Lab Results

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Lab_Result	Field_Event_ID	Concise identifier for the field sampling effort.	Text	50	Yes	No	**PK**	Analytical_Sample
	Analytical_Sample_ID	Unique sample identifier for each analytical sample	Text	50	Yes	No	**PK**	Analytical_Sample
	Lab_SDG	Laboratory Sample Delivery Group (batch) ID	Text	50	Yes	No	**PK**	
	Analysis	Lab Analysis (e.g., VOCs)	Text	24	Yes	Yes	**PK**	Lab_Analysis
	Analyte	Analyte/Parameter name (e.g., Lead; Arsenic; etc.)	Text	28	Yes	Yes	**PK**	Analyte
	Preparation_Method	Lab preparation method	Text	24	Yes	Yes	**PK**	Prep_Method
	Extraction_Method	Lab extraction method	Text	24	Yes	Yes	No	Extract_Method
	Analytical_Method	Lab Analytical Method (e.g., 8270M)	Text	24	Yes	Yes	**PK**	Anal_Method
	Material_Analyzed	Material analyzed	Text	24	Yes	Yes	**PK**	Sample_Material
	Fraction_Analyzed	Indicator of what fraction of the sample was analyzed (e.g., total, dissolved, leachate, sieved size interval)	Text	24	Yes	Yes	**PK**	Fraction
	Lab_Replicate	Laboratory replicate identifier	Text	20	Yes	No	**PK**	
	Result	Result (concentration or equivalent) reported by the lab	Double precision	0	Yes	No	No	
	Sig_Figs	Significant digits of the result	Integer	0	Yes	No	No	
	Result_Units	Result unit of measurement	Text	24	Yes	Yes	No	Unit
	Basis	"Wet" for wet_weight basis reporting; "Dry" for dry_weight reporting.	Text	10	Yes	Yes	No	Meas_Basis
	Lab_Qualifiers	Qualifiers and flags assigned by the laboratory	Text	12	No[a]	Yes	No	
	TIC	Is this a tentatively identified compound (TIC)?	Boolean	0	Yes	No	No	
	Validated	Has this result been validated?	Boolean	0	Yes	No	No	
	Validation_Level	Stage of validation - electronic and manual	Text	24	Yes	Yes	No	Validation_Level
	Result_Qualifiers	Final validated result qualifiers/flags (e.g., J;U;ND;<;>)	Text	12	No[a]	Yes	No	
	Undetected	Flag to distinguish detected and undetected results, based on lab or validation qualifiers	Boolean	0	Yes	No	No	
	Estimated	Flag to distinguish estimated and non-estimated results, based on lab or validation qualifiers	Boolean	0	Yes	No	No	
	Rejected	Flag to distinguish rejected and non-rejected results, based on lab or validation qualifiers	Boolean	0	Yes	No	No	
	Reportable	Flag to distinguish reportable and non-reportable results, based on validation or data quality review	Boolean	0	Yes	No	No	
	Validator	Validation company name	Text	100	No	No	No	Organization
	Validator_Comments	Comment on the validation assessment for this result	Text	255	No	No	No	
	Comments	Comments on the results that do not pertain to either the validation results or data quality review results	Text	255	No	No	No	
	QA_Comments	QA comment resulting from any data quality review conducted in addition to data validation	Text	255	No	No	No	
	Date_Received	Date and time that the analytical sample was received by the lab	DateTime	0	No	No	No	
	Date_Extracted	Date and time that the sample was extracted by the lab	DateTime	0	No	No	No	
	Date_Analyzed	Date and time that the analysis was performed by the lab	DateTime	0	No	No	No	
	Dilution_Factor	Effective test dilution factor	Double precision	0	Yes	No	No	
	Lab_Name	Laboratory that performed the analysis	Text	24	Yes	Yes	No	Lab
	Lab_Sample_ID	Sample identifier assigned by the laboratory	Text	50	No	No	No	
	Original_Lab_Result	Original result reported by the laboratory, if data validation resulted in restatement of the value	Double precision	0	No	No	No	
	Lab_Cal_Batch	The laboratory's calibration batch identifier	Text	50	No	No	No	
	Lab_QC_Batch	The laboratory's quality control batch identifier	Text	50	No	No	No	
	IDL	Instrument Detection Limit (IDL)	Double precision	0	No	No	No	
	IDL_Units	IDL units	Text	20	No	Yes	No	Unit
	MDL	Method Detection Limit (MDL)	Double precision	0	No	No	No	
	MDL_Units	MDL units	Text	20	No	Yes	No	Unit
	Quantitation_Limit	Quantitation limit as determined by the lab	Double precision	0	No	No	No	
	Quantitation_Limit_Units	Quantitation limit units	Text	20	No	Yes	No	Unit
	Reporting_Limit	Reporting limit as determined by the lab	Double precision	0	No	No	No	
	Reporting_Limit_Units	Reporting limit units	Text	20	No	Yes	No	Unit
	Value_Type	Water Quality Exchange (WQX) result value type (e.g., actual; estimated; calculated)	Text	24	No	Yes	No	Result_Value_Type

Notes

a. Laboratory and final result qualifiers are not required by the database design because they may not exist. However, they must be provided if they exist.

Table 23 - Calculated Chemistry

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Calculated_Chemistry	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for this interpretive sample	Text	50	Yes	No	**PK**	Field_Sample
	Material_Analyzed	Material analyzed	Text	24	Yes	Yes	**PK**	Sample_Material
	Fraction_Analyzed	Indicator of what fraction of the sample was analyzed (e.g., total, dissolved, leachate, sieved size interval)	Text	24	Yes	Yes	**PK**	Fraction
	Analyte	Analyte/Parameter name (e.g., Lead; Arsenic; etc.)	Text	28	Yes	Yes	**PK**	Analyte
	Calculation_Method	Standardized description of the calculation method used (data selection and summarization)	Text	24	Yes	Yes	**PK**	Calc_Method
	Result	Calculation result (concentration or equivalent)	Double precision	0	Yes	No	No	
	Undetected	Flag to distinguish detected and undetected results, propagated through the calculation process	Boolean	0	Yes	No	No	
	Estimated	Flag to distinguish estimated and non-estimated results, propagated through the calculation process	Boolean	0	Yes	No	No	
	Rejected	Flag to distinguish rejected and non-rejected results, propagated through the calculation process	Boolean	0	Yes	No	No	
	Limited_Analytes	Flag to distinguish if the number of analytes reported is "limited" and therefore below the expected number of analytes for analyte group totals	Boolean	0	Yes	No	No	
	Result_Units	Result unit of measurement	Text	20	Yes	Yes	No	Unit
	Basis	"Wet" for wet_weight basis reporting; "Dry" for dry_weight reporting.	Text	10	Yes	Yes	No	Meas_Basis

Table 24 - Tox Test Batch

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Tox_Test_Batch	Lab_Name	Toxicity test laboratory	Text	24	Yes	No	**PK**	Lab
	Tox_test_batch	Laboratory-specific toxicity test batch identifier	Text	50	Yes	No	**PK**	
	Tox_test_type	Type of toxicity test conducted	Text	24	Yes	Yes	No	Tox_Test_Type
	Taxon	Taxon code for organism used for the toxicity test	Text	24	Yes	Yes	No	Taxon
	Life_Stage	Life stage of organisms used for the toxicity test	Text	24	Yes	Yes	No	Life_Stage
	Start_Date	Starting date of the toxicity test	DateTime	0	No	No	No	
	End_Date	Ending date of the toxicity test	DateTime	0	No	No	No	
	QA_Level	Level of data quality review used	Text	24	Yes	Yes	No	QA_Level

Table 25 - Tox Test Result

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Tox_Test_Result	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50	Yes	No	**PK**	Field_Sample
	Lab_Name	Toxicity test laboratory	Text	24	Yes	Yes	**PK**	Tox_Test_Batch
	Tox_Test_Batch	Laboratory-specific toxicity test batch identifier	Text	50	Yes	No	**PK**	Tox_Test_Batch
	Tox_Test_Variable	The variable in which results are expressed (e.g., survival, growth, reproduction)	Text	24	Yes	Yes	**PK**	Tox_Test_Var
	Tox_Test_Measurement	The type of measurement made (e.g., count of survivors, percent normal, change in mass)	Text	24	Yes	Yes	**PK**	Tox_Test_Meas
	Replicate	Replicate identifier	Text	20	Yes	No	**PK**	
	Tox_Test_Result	The numeric result of the toxicity test	Numeric	0	Yes	No	No	
	Sig_Figs	Significant digits of the tox test result	Integer	2	Yes	No	No	
	Tox_Test_Units	The units for the toxicity test result	Text	20	Yes	Yes	No	Unit
	QA_Level	Level of data quality review used	Text	24	Yes	Yes	No	QA_Level
	Tox_Test_Results_Comments	Comments on this toxicity test result	Text	255	No	No	No	

Table 26 - Species Abundance

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Species_Abundance	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50	Yes	No	**PK**	Field_Sample
	Taxon	Taxon code for organisms	Text	24	Yes	Yes	**PK**	Taxon
	Sex	Sex of organisms	Text	1	Yes	Yes	**PK**	Sex
	Life_Stage	Life stage of organisms	Text	24	Yes	Yes	**PK**	Life_Stage
	Abundance_measurement	Type of abundance measurement (count, concentration, density, spatial coverage)	Text	24	Yes	Yes	**PK**	Abund_Meas
	Replicate	Replicate identifier	Text	20	Yes	No	**PK**	
	Abundance	Abundance measurement	Double precision	0	Yes	No	No	
	Sig_Figs	Significant digits of the Abundance	Integer	2	Yes	No	No	
	Abundance_units	Abundance measurement units	Text	20	Yes	Yes	No	Unit
	Lab_Name	Laboratory that measured the abundance	Text	24	No	Yes	No	Lab
	Species_Abundance_Comments	Comments on the species abundance measurement	Text	255	No	No	No	

Table 27 - Species Measurement

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Species_Measurement	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	Yes	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50	Yes	Yes	**PK**	Field_Sample
	Taxon	The type of collection observation made (e.g., tide stage, sediment class)	Text	24	Yes	Yes	**PK**	Taxon
	Sex	The sex of the species measured	Text	1		Yes	**PK**	Sex
	Life_Stage	The life stage of the species measured	Text	24		Yes	**PK**	Life_Stage
	Species_Measurement	The species measured	Text	24	Yes	Yes	**PK**	Measurement
	Measurement_Method	The method used to make the measurement	Text	24		Yes	**PK**	Meas_Method
	Replicate	Replicate identifier to distinguish multiple measurements (e.g., “1”, “2”)	Text	10	Yes	Yes	**PK**	
	Measured_Value	The numerical result of the measurement	Double precision		Yes			
	Sig_Figs	Significant digits of the result	Integer		Yes			
	Units	Result unit of measurement	Text	20	Yes	Yes		Units
	QA_Level	Level of data quality review used	Text	24	Yes	Yes		QA_level
	Reportable	Flag to distinguish reportable and non-reportable results based on data quality review	Boolean	0	Yes			
	Species_Measurement_Comments	Description of the species measurement	Text	255	No	No	No	

Table 28 - Wells

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Well	Well_ID	Well identifier	Text	50	Yes	No	**PK**	
	Location	Sampling Location Code/Monitoring Location Code	Text	50	Yes	No	No	Location
	Vertical_Reference_Point	The surface or reference point from which depths are measured (e.g., water, sediment, soil, well benchmark)	Text	24	Yes	Yes	No	Vert_Ref_Pt
	Well_Hole_Depth_Measure	Depth below the vertical reference point to the bottom of the hole on completion of drilling	Double precision	0	No	No	No	
	Well_Hole_Depth_Measure_Unit	Units for Well_Hole_Depth_Measure field	Text	20	No	Yes	No	Unit
	Completion_Date	Date that well construction was completed	DateTime	0	No	No	No	
	Screen_Upper_Depth	Upper depth of the screened interval	Double precision	0	No	No	No	
	Screen_Lower_Depth	Lower depth of the screened interval	Double precision	0	No	No	No	
	Well_Use	Usage of well (e.g., private water supply, public water supply, monitoring, stock watering, injection)	Text	24	No	Yes	No	Well_Use
	Well_Status	Status of well usage (e.g., active, abandoned)	Text	24	No	Yes	No	Well_Status
	Well_Status_Date	Effective date of the latest well status assessment	DateTime	0	No	Yes	No	

Table 29 - Criteria

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Criteria	Critcode	Concise identifier for the set of criteria	Text	50	Yes	No	**PK**	Critiadef
	Analyte	Analyte or parameter name (e.g., Lead; Arsenic; etc.)	Text	28	Yes	Yes	**PK**	Analyte
	Crit_Value	Numeric value of the criterion	Double precision	0	Yes	Yes	No	
	Unit	Unit for the criterion	Text	10	Yes	Yes	No	Unit
	Meas_Basis	"Wet" or "Dry"	Text	10	Yes	Yes	**PK**	Meas_Basis
	Range_Limit	Limit of the criteria range (i.e., upper or lower)	Text	5	Yes	Yes	**PK**	
	Criteria_Comments	Comments pertaining to this criterion	Text	255		No		
	Hardness_Dependent	Is the criterion hardness dependent?	Boolean	0		No		
	Hd_m	Parameter "mA" for criteria maximum concentration or "mC" for criterion continuous concentration	Double precision	0		No		
	Hd_b	Parameter "bA" for CMC or "bC" for criterion continuous concentration	Double precision	0		No		
	Cf_b	Intercept for Conversion Factor calculation	Double precision	0		No		
	Cf_m	Slope for Conversion Factor calculation	Double precision	0		No		
	Ph_dependent	Is the criterion Ph dependent?	Double precision	0		No		
	Ph_m	Slope for pH-dependent calculation	Double precision	0		No		
	Ph_b	Intercept for pH-dependent calculation	Double precision	0	No	No	No	

Table 30 - Criteriadef

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Criteriadef	Critcode	Concise identifier for the set of criteria	Text	50	Yes	No	**PK**	
	Critdescrip	Description of this set of criteria	Text	255	Yes	No	No	
	Material	The material relevant to the criteria (e.g., surface water, sediment)	Text	24	No	Yes	No	Material
	Resource	Resource to which the criteria apply (e.g., sediment invertebrates, humans)	Text	24	No	Yes	No	Resource
	Guideline_Type	Guideline type for the criteria (e.g., chronic, continuous)	Text	24	No	Yes	No	Guidelinetype
	Crityear	Year that the criteria were published or promulgated	Text	4	No	No	No	
	Citation	Citation of the criteria (e.g., Van den Berg et al., 2006)	Text	24	Yes	Yes	No	
	Criteria_Def_Comments	Comments on the criteria	Text	255	No	No	No	

Table 31 - Data Table Summary

Count	Data Table Name	Description
1	ASAOC	Description of Administrative Settlement Agreement and Order on Consent (ASAOC)
2	Site	Identity and geographic boundary of an area used for assessment (e.g., SDU)
3	ASAOC_Sites	Sites related to each ASAOC
4	Party	Performing party name
5	Party_Group	Identity of a group of multiple parties
6	Party_Group_Members	Parties in each party group
7	Organization	Organization on whose behalf the field event is undertaken
8	Field_Event	Description of a field sampling investigation (e.g., remedial design sampling)
9	Event_Element	Identifier for a subset of the sampling effort (e.g., Phase 1, background sampling, random grid sampling)
10	Admin_Event	Administrative event related to an ASAOC (e.g., remedial oversight)
11	Location	Geographic location and elevation related to sample collection that is not event specific
12	Field_Event_Location	Event specific name for a geographic location used for sample collection
13	Collection_Observation	Ad hoc observations made on field collections
14	Field_Collection	Description of the collection of environmental material that may be subsampled
15	Collection_Measurement	Ad hoc numerical measurements made on field collections
16	Field_Sample	Identification and description of interpretive samples used for data evaluation
17	Sample_Observation	Ad hoc observations made on samples
18	Sample_Composite	Lists of field samples that are composited to create other field samples
19	Sampling_Reason	Purpose(s) for collection of each sample (e.g., site characterization, ecological risk assessment)
20	Sample_Measurement	Ad hoc numerical measurements made on field samples
21	Analytical_Sample	Identifiers for splits of field samples
22	Lab_Result	Analytical laboratory results and qualifiers
23	Calculated_Chemistry	Chemical sums computed from laboratory analytical results
24	Tox_Test_Batch	Toxicity test batch and type
25	Tox_Test_Result	Toxicity test results for each batch and field sample
26	Species_Abundance	Biota abundance measurements
27	Species_Measurement	Ad hoc measurements made on samples
28	Well	Well description for locations used for groundwater sampling
29	Criteria	Stores chemical criteria (e.g., RALs)
30	Criteriadef	Stores chemical criteria citations

Table 32 - Lookup Tables

Count	Lookup Table Name	Description
1	Abund_Meas	Type of abundance measurement (e.g., count, concentration, density, spatial coverage)
2	Admin_Activity	The type of non-field administrative activity (e.g., remedial oversight, remedial action, emergency response)
3	Anal_Method	Lab analytical method (e.g., 8270M)
4	Analyte	Analyte/parameter name (e.g., lead, arsenic)
5	Calc_Method	Method used to calculate analyte totals and handling of non-detects
6	Chemclass	Group of analytes (e.g., semivolatiles, pesticides/PCBs, metals) How is this different than lab_analysis?
7	Coll_Design	Sample collection design or scheme (e.g., single-point grab, spatial composite, temporal composite, MIS)
8	Coll_Method	Sample collection gear or method (e.g., grab, core)
9	Color	Description of sample color
10	Composite_Type	Composite method (e.g., single [not composited], spatial composite, temporal composite)
11	County_Code	Three-character county abbreviation
12	Detection_Limit	Laboratory limits (instrument detection limit, method detection limit, quantitation limit, and reporting detection limit)
13	Dimension	Dimensions expressed in terms such as mass/mass (e.g., for units of milligrams per kilogram [mg/kg])
14	Elev_Datum	Datum used to determine the elevation measurement. (e.g., NAVD88)
15	Elev_Method	Method used to determine the elevation measurement. (e.g., Altimetry; GPS; Interpolation; Survey)
16	Event_Status	Completion status of a field or administrative event (e.g., pending, underway, completed)
17	Extract_Method	The extraction method used in the field or lab
18	Field_Activity	The type of field activity to be conducted (e.g., remedial design sampling, confirmation sampling, post-closure monitoring)
19	Fips_Code	Federal Information Processing Standard (FIPS) Publication. US states are a 2-digit number and US counties are identified by a 3-digit number
20	Fraction	Indicator of the fraction of the sample material that was analyzed (e.g., total, dissolved, leachate, sieved size interval)
21	Fraction_Matrix	Contains both the fraction and matrix combination
22	Guideline_Type	The guideline for a specific regulatory criteria
23	Lab	Analytical laboratory company name (e.g., A4 Scientific, Inc.)
24	Lab_Analysis	Group of analytes (e.g., semivolatiles, pesticides/PCBs, metals)
25	Life_Stage	Life stage of organism (e.g., egg, smolt, juvenile, adult)
26	Material	The material sampled (e.g., air, sediment, tissue)
27	Matrix	The matrix of the sampled material (e.g., liquid, solid, vapor)
28	Meas_Basis	Basis for measurement (e.g., "Wet" for wet-weight basis reporting; "Dry" for dry-weight reporting)
29	Meas_Method	Method used to make quantitative measurements on field samples or collections
30	Measbasis matrix	The basis of the lab measurements (e.g., dry weight, wet weight)
31	Measurement	Types of quantitative measurements made on field collections or samples
32	ND_All_Treatment	The handling of various detection limits in calculated chemistry (e.g., min, max, sum)
33	ND_Treatment	The detection limit summation method in calculated chemistry (e.g., full, half, zero)
34	Observation	Descriptions of the observations (e.g., silt, C horizon)
35	Observation_Type	The type of observations (e.g., sediment, tidal stage, soil horizons)
36	Odor	Description of sample odor
37	Position_Method	Geopositioning method used to establish latitude and longitude coordinates (e.g., GPS; Interpolation; Survey)
38	Prep_Method	Method to prepare samples for analysis (e.g., filtering, fractionation)
39	QA_Level	QA level codes
40	Resource	The resource related to the criteria definition
41	Result_Value_Type	Process used to determine the result value (e.g., actual, estimated)
42	Riverbank	Location within river (e.g., east bank, west bank, navigation channel)
43	Sample_Material	The most specific description of the type of material sampled
44	Sample_Reason	Purpose(s) for collection of each sample (e.g., site characterization, ecological risk assessment)
45	Sex	Sex of organism (e.g., female, male, hermaphrodite, intersex)
46	State	Two-character state abbreviation
47	Subsample_Type	Description of the method used to form interpretive samples (field samples) from a collection
48	Taxon	Taxon code for organisms
49	Tox_Test_Meas	Toxicity test measurement (e.g., count of survivors, percent normal, change in mass)
50	Tox_Test_Type	Toxicity test type (e.g., amphipod 10-day bioassay, echinoderm 72-hour bioassay)
51	Tox_Test_Var	Toxicity test variable (e.g., survival, growth, reproduction)
52	Unit	Unit used for any types of quantitative measurement
53	Validation_Level	Type of validation applied (e.g., Level II, Level IV)
54	Vert_Ref_Pt	Description of the zero point for depth measurements (e.g., water surface, sediment surface top of well casing)
55	Well_Status	Status of well usage (e.g., active, abandoned)
56	Well_Use	Usage of well (e.g., private water supply, public water supply, monitoring, stock watering, injection)
57	Zone_Type	Location categorization (e.g., subtidal, intertidal, upland, facility)

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
ASAOC	ASAOC_ID	Unique ID for an Administrative Settlement Agreement and Order on Consent (AOC)	Text	50	Yes	No	**PK**	
	AOC_Effective_Date	Date AOC becomes effective	DateTime	0	No	No	No	
	AOC_Docket_Number	Associated AOC Docket Number	Text	50	No	No	No	
	Party_Group_Name	Group name of the parties associated with this ASAOC	Text	100	No	No	No	Party_Group
Site	Site_No	Site identifier	Text	12	Yes	No	**PK**	
	Site_Desc	Site description	Text	255	Yes	No	No	
	State_Code	Two-character state abbreviation	Text	2	Yes	Yes	No	State
	Spatial_Extent	A polygon or multipolygon representation of the spatial boundary of the site	Geometry	0	Yes	No	No	
	CERCLIS	The CERCLIS identifier used by the Superfund program	Text	20	Yes	No	No	
ASAOC_Sites	ASAOC_ID	Unique ID for an Administrative Settlement Agreement and Order on Consent (AOC)	Text	50	Yes	No	**PK**	ASAOC
	Site_No	Site identifier	Text	12	Yes	No	**PK**	Site
Party	Party_Name	Unique name of the party	Text	100	Yes	No	**PK**	
	Party_Contact	Contact name for the party	Text	255	No	No	No	
	Organization	Organization of the party	Text	100	Yes	No	FK	Organization
Party_Group	Party_Group_Name	A name to identify a group of parties that are associated with one or more ASAOCs	Text	100	Yes	No	**PK**	
	Description	A description of the Group	Text	255	Yes	No	No	
Party_Group_Name	Party_Group_Name	A name to identify a group of parties that are associated with one or more ASAOCs	Text	100	Yes	No	**PK**	Party_Group
	Party_Name	Name of the party	Text	100	Yes	No	**PK**	Party
Organization	Organization	Concise identifier for the organization	Text	100	Yes	Yes	**PK**	
	Description	Detailed description of the organization	Text	255				
Field_Event	Field_Event_ID	Concise name or identifier for the field sampling effort. For example: 2012_LBCR Predesign Invest	Text	50	Yes	No	**PK**	
	ASAOC_ID	Identifier for the ASAOC that this field event is conducted to support	Text	50	Yes	No	No	ASAOC
	Field_Activity	The type of activity to be conducted (e.g., remedial design sampling, confirmation sampling, post-closure monitoring)	Text	35	Yes	Yes	No	Field_Activity
	Event_Description	Description of the event. For example: Lower Burke Canyon Repository Predesign Investigation	Text	255	Yes	No	No	
	Event_Status	Event completion status (e.g., pending, underway, completed)	Text	24	No	Yes	No	Event_Status
	Spatial_Extent	A point, multipoint, line, multiline, polygon, or multipolygon representation of the spatial extent of the event. Although not strictly required, this should be provided if the sampling locations are constrained or defined by a spatial boundary such as a remedial design area or a removal area.	Geometry	0	No	No	No	
	Sponsor	Name of the organization on whose behalf the field event is undertaken	Text	100	Yes	No	No	Organization
	Sponsor_Contact	Contact information for the field event sponsor	Text	255	No	No	No	
	Contractor	Full name of the company contracted by the sponsor to conduct or manage the sampling	Text	100	Yes	No	No	Organization
	QAPP_Approved	Indicates if the QAPP has been approved for the sampling effort	Boolean	0	Yes	No	No	
	QAPP_Approving_Org	Identifies the agency or other organization that approved the QAPP for the sampling effort	Text	100	No	No	No	
	QAPP_Approved_Other	Other information about approval of the QAPP	Text	150	No	No	No	
Event_Element	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	Yes	**PK**	Field Event
	Event_Element	Identifier for a subset of the sampling effort (e.g., Phase 1, Phase 2, background sampling, random grid sampling)	Text	50	Yes	No	**PK**	
	Description	Description of the event	Text	255	Yes	No	No	
Admin_Event	Admin_Event_ID	Concise name or identifier for the administrative event	Text	50	Yes	No	**PK**	
	Admin_Activity	The type of activity to be conducted (e.g., remedial design, remedial action, emergency response)	Text	35	Yes	Yes	No	Admin_Activity
	Event_Description	Description of the event. For example: Lower Burke Canyon Repository Predesign Investigation	Text	255	Yes	No	No	
	ASAOC_ID	Unique ID for an Administrative Settlement Agreement and Order on Consent (ASAOC). Although not required by the database design, this value should be provided if the event is controlled or specified by an ASAOC.	Text	50	No	No	No	ASAOC
	Event_Status	Event completion status (e.g., pending, underway, completed)	Text	24	No	Yes	No	Event_Status

Table 33 - Complete PHIDB Data Dictionary Version 2.0

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Location	Location	Target sampling location identifier	Text	50	Yes	No	**PK**	
	Location_Description	Narrative description of the location	Text	255	Yes	No	No	
	Country_Code	Country code	Text	5	No	Yes	No	Country_Code
	County_FIPS_Code	County code	Text	5	No	Yes	No	Fips_Code
	Loc_Geom	Nominal (target) location, as a point, multipoint, polygon, or multipolygon geometry. Coordinates are stored as decimal degrees in WGS84 (SRID=4326). This value must be unique on every row (i.e., this is a candidate key).	Geometry	0	Yes	No	No	
	Elev_Datum	Datum used to determine the elevation measurement. (e.g., NAVD88; NGVD29)	Text	50	No	Yes	No	Elev_Datum
	Elev_Method	Method used to determine the elevation measurement. (e.g., Altimetry; GPS; Interpolation; Survey)	Text	24	No	Yes	No	Elev_Method
	Geo_Method	Geopositioning method used to establish latitude and longitude coordinates (e.g., GPS; Interpolation; Survey)	Text	24	Yes	Yes	No	Position_Method
	Huc_Eight_Digit_Code	Eight digit USGS HUC code. Equals Sub_Basin	Text	20	No	No	No	
	Location_Zone	Location categorization (e.g., subtidal, intertidal, upland, facility)	Text	24	No	Yes	No	Zone_Type
	River_Mile	River mile, to at least tenths of a mile	Numeric	0	No	No	No	
	River_Bank	Code to specify whether the location is near a river bank or in the navigation channel	Text	24	No	Yes	No	Riverbank
	State_Code	State code - 2 character state abbreviation	Text	2	No	Yes	No	State
	Stream_or_Creek_Name	Name of the waterbody represented by this location	Text	200	No	Yes	No	
	Surf_Elev	The ground elevation of a geographic point where samples or field measurements are collected	Double precision	0	No	No	No	
	Surf_Elev_Units	Surface elevation units (e.g., feet; meters)	Text	20	No	Yes	No	Unit
	USGS_Station_ID	USGS location identifier	Text	100	No	Yes	No	
	USGS_Station_Name	USGS location name	Text	255	No	Yes	No	
Field_Event_Location	Field_Event_ID	Concise identifier for the field sampling effort.	Text	50	Yes	No	**PK**	Field_Event
	Event_location	Event-specific location code	Text	50	Yes	No	**PK**	
	Location	Sampling location code/monitoring location code	Text	50	Yes	No	No	Location
Collection_Observation	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	
	Collection_ID	Unique collection identifier for each core, grab, or group of related field samples	Text	50	Yes	No	**PK**	Field_Collection
	Observation_Type	The type of collection observation made (e.g., tide stage, sediment class)	Text	24	Yes	Yes	**PK**	Observation
	Observation	The observation made during collection (e.g., flood, silt)	Text	24	Yes	Yes	**PK**	Observation
	Collection_Observation_Comments	Comments on the collection observation	Text	255	No	No	No	
Field_Collection	Field_Event_ID	Concise identifier for the field sampling effort.	Text	50	Yes	No	**PK**	Field_Event_Location
	Collection_ID	Unique collection identifier for each core, grab, or group of related field samples.	Text	50	Yes	No	**PK**	
	Event_Location	Event-specific location code	Text	50	Yes	No	No	Field_Event_Location
	Sub_Location	Narrative description of any systematic deviation or difference from the nominal location	Text	255	No	No	No	
	Collection_Date	Date and time of acquisition of the collection material. This is often the GPS time recorded.	DateTime	0	Yes	No	No	
	Collection_Method	Sample collection method (e.g., grab, core)	Text	24	Yes	Yes	No	Coll_Method
	Collection_Design	Sample collection design or scheme (e.g., single-point grab, spatial composite, temporal composite, MIS)	Text	24	Yes	Yes	No	Coll_Design
	Collection_Depth_Upper	Upper depth of the collection	Double precision	0	No	No	No	
	Collection_Depth_Lower	Lower depth of the collection	Double precision	0	No	No	No	
	Collection_Depth_Units	Units for the upper and lower collection depths	Text	20	No	Yes	No	Unit
	Composite_Type	The compositing method used for the collection (e.g., single, spatial, depth, temporal)	Text	24	Yes	Yes	No	Composite_Type
	Composite_Count	The number of other collections that were composited to create this collection	Integer	0	No	No	No	
	Composite_Period	The length of time over which a temporal composite was conducted	Double precision	0	No	No	No	
	Composite_Period_Units	The units for the composite period	Text	20	No	Yes	No	Unit
	Vertical_Reference_Point	The surface from which depths are measured (e.g., water surface, sediment surface, soil surface, well benchmark)	Text	24	Yes	Yes	No	Vert_Ref_Pt
	Vertical_Ref_Pt_Elev	Elevation of the reference point	Double precision	0	No	No	No	
	Vertical_Ref_Pt_Elev_Units	Units of the reference point elevation (e.g. feet, meters)	Text	20	No	No	No	
	Elev_Datum	Datum used to determine the elevation measurement. (e.g., NAVD88; NGVD29)	Text	50	No	Yes	No	Elev_Datum
	Elev_Method	Method used to determine the elevation measurement. (e.g., Altimetry; GPS; Interpolation; Survey)	Text	24	No	Yes	No	

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Field_Collection (continued)	Coll_Coords	The actual coordinates at which the collection was made, as a point, polygon, multipoint, or multipolygon geometry. Coordinates are stored as decimal degrees in WGS84 (SRID=4326).	Geometry	0		No	No	
	Geo_Method	Geopositioning method used to establish latitude and longitude coordinates.	Text	30	Yes	Yes	No	Position_Method
	Horiz_Accuracy_Measure	Horizontal accuracy measurement	Double precision	0	No	No	No	
	Horiz_Accuracy_Measure_Units	Horizontal accuracy measurement units	Text	20	No	Yes	No	Unit
	Collection_Material	The type of material collected	Text	24	Yes	Yes	No	Sample_Material
	Taxon	Taxon code for organisms	Text	24	No	Yes	No	Taxon
	Event_Element	Identifier for a subset of the sampling effort (e.g., Phase 1, Phase 2, background sampling, random grid sampling)	Text	50	No	No	No	
	Sampler_Org	Name of organization that collected the sample	Text	100	Yes	Yes	No	
	Sampler_Person	Name of the person who collected the sample	Text	64	No	No	No	
	Remarks	Comments on sample collection	Text	255	No	No	No	
Collection_Measurement	Field_Event_ID	Concise identifier for the field sampling effort.	Text	50	Yes	No	**PK**	Field_Collection
	Collection_ID	Unique collection identifier for each core, grab, or group of related field samples.	Text	50	Yes	No	No	Field_Collection
	Collection_Measurement	The type of measurement made (e.g., tide height, cloud cover, number of grabs)	Text	24	Yes	Yes	**PK**	Measurement
	Measurement_Method	The method used to make the sample measurement	Text	24	Yes	Yes	**PK**	Meas_Method
	Replicate	Replicate identifier to distinguish multiple measurements (e.g., "1", "2")	Text	10	Yes	No	**PK**	
	Measured_Value	The numerical result of the measurement	Double precision	0	Yes	No	No	
	Sig_Figs	Significant digits of the result	Integer	2	Yes	No	No	
	Collection_Measurement_Units	Result unit of measurement	Text	20	Yes	Yes	No	Unit
	QA_Level	Level of data quality review used	Text	24	Yes	Yes	No	QA_Level
	Reportable	Flag to distinguish reportable and non-reportable results based on data quality review	Boolean	0	Yes	No	No	
	Collection_Measurement_Comments	Comments on the measured value	Text	255	No	No	No	
Field_Sample	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Collection
	Collection_ID	Unique collection identifier for each core, grab, or group of related field samples	Text	50	Yes	No	No	Field_Collection
	Main_Sample_ID	Unique sample identifier for this interpretive sample	Text	50	Yes	No	**PK**	
	Sample_Material	The type of material sampled	Text	24	No	Yes	No	Sample_Material
	Collection_Part	The fraction of the collection represented by this sample (e.g., entire, vertical horizon, filtered fraction)	Text	24	Yes	Yes	No	Subsample_Type
	Sample_Date	Date and time of creation of the sample for a composite or single sample. The sample date may be the same as the collection date if the collection is not subdivided into multiple samples.	DateTime	0	Yes	No	No	
	Sample_Depth_Upper	Sample upper depth relative to the zero point of the collection	Double precision	0	No	No	No	
	Sample_Depth_Lower	Sample lower depth relative to the zero point of the collection	Double precision	0	No	No	No	
	Sample_Depth_Units	Sample depth units	Text	20	No	Yes	No	Unit
	Taxon	Taxon code for organisms	Text	24	No	Yes	No	Taxon
	Original_Sample_ID	Original sample ID, if the Sample_ID is a corrected value	Text	50	No	No	No	
	Sampler_Person	Name of person who collected the sample	Text	64	Yes	Yes	No	
	Field_Prep_Method	Sample preparation or treatment method carried out at the time of collection (filtering, fractionation)	Text	24	No	Yes	No	Prep_Method
	Sample_Mass	The mass of this sample	Double precision	0	No	No	No	
	Sample_Mass_Units	The units for the sample mass	Text	20	No	Yes	No	Unit
	Sample_Mass_Basis	"Wet" or "Dry" for sample masses	Text	10	No	Yes	No	Meas_Basis
	Sample_Volume	The volume of this sample	Double precision	0	No	No	No	
	Sample_Volume_Units	The units for the sample volume	Text	20	No	Yes	No	Unit

Table 33 - Complete PHIDB Data Dictionary Version 2.0

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Field_Sample (continued)	Sample_Color	The color of the sample upon collection	Text	24	No	Yes	No	Color
	Sample_Odor	The odor of the sample upon collection	Text	24	No	Yes	No	Odor
	Remarks	Sample remarks	Text	255	No	No	No	
Sample_Observation	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	Yes	**PK**	Field Sample
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50			**PK**	Field Sample
	Observation_Type	The type of collection observation made (e.g., tide stage, sediment class)	Text	24			**PK**	Observation
	Observation	The observation made during collection (e.g., flood, silt)	Text	24			**PK**	Observation
	Sample_Observation_Comments	Description of the event	Text	255	Yes	No	No	
Sample_Composite	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for the interpretive sample that is a composite	Text	50	Yes	No	**PK**	Field_Sample
	Subcomposite_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Subcomposite_Sample_ID	Unique sample identifier for one of the interpretive samples that makes up the composite	Text	50	Yes	No	**PK**	Field_Sample
	Subcomposite_Mass	The mass of this subcomposite sample used in the composite	Double precision	0	No	No	No	
	Subcomposite_Mass_Units	The units for the subcomposite mass	Text	20	No	Yes	No	Unit
	Subcomposite_Volume	The volume of this subcomposite used in the composite	Double precision	0	No	No	No	
	Subcomposite_Volume_Units	The units for the subcomposite volume	Text	20	No	Yes	No	Unit
Sample_Reason	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for this interpretive sample	Text	50	Yes	No	**PK**	Field_Sample
	Sampling_Reason	Description of the purpose for collection of this sample	Text	24	Yes	Yes	**PK**	Sample_Reason
Species_Measurement	Field_Event_ID	Concise identifier for the field sampling effort.	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50	Yes	No	**PK**	Field_Sample
	Sample_Measurement	The type of measurement made (e.g., length)	Text	24	Yes	Yes	**PK**	Measurement
	Measurement_Method	The method used to make the sample measurement	Text	24	Yes	Yes	**PK**	Meas_Method
	Replicate	Replicate identifier to distinguish multiple measurements (e.g., “1”, “2”)	Text	10	Yes	No	**PK**	
	Measured_Value	The numerical result of the measurement	Double precision	0	Yes	No	No	
	Sig_Figs	Significant digits of the result	Integer	2	Yes	No	No	
	Sample_Measurement_Units	Result unit of measurement	Text	20	Yes	Yes	No	Unit
	QA_Level	Level of data quality review used	Text	24	Yes	Yes	No	QA_Level
	Reportable	Flag to distinguish reportable and non-reportable results based on data quality review	Boolean	0	Yes	No	No	
	Comments	Comments on the measured value	Text	255	No	No	No	
Analytical_Sample	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Analytical_Sample_ID	Unique sample ID for an analytical sample (split or duplicate) that is part or all of an interpretive sample	Text	50	Yes	No	**PK**	
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50	Yes	No	No	Field_Sample
	Original_Sample_ID	Original identifier for the analytical sample	Text	50	Yes	No	No	
Lab_Result	Field_Event_ID	Concise identifier for the field sampling effort.	Text	50	Yes	No	**PK**	Analytical_Sample
	Analytical_Sample_ID	Unique sample identifier for each analytical sample	Text	50	Yes	No	**PK**	Analytical_Sample
	Lab_SDG	Laboratory Sample Delivery Group (batch) ID	Text	50	Yes	No	**PK**	
	Analysis	Lab Analysis (e.g., VOCs)	Text	24	Yes	Yes	**PK**	Lab_Analysis
	Analyte	Analyte/Parameter name (e.g., Lead; Arsenic; etc.)	Text	28	Yes	Yes	**PK**	Analyte
	Preparation_Method	Lab preparation method	Text	24	Yes	Yes	**PK**	Prep_Method
	Extraction_Method	Lab extraction method	Text	24	Yes	Yes	No	Extract_Method
	Analytical_Method	Lab Analytical Method (e.g., 8270M)	Text	24	Yes	Yes	**PK**	Anal_Method
	Material_Analyzed	Material analyzed	Text	24	Yes	Yes	**PK**	Sample_Material
	Fraction_Analyzed	Indicator of what fraction of the sample was analyzed (e.g., total, dissolved, leachate, sieved size interval)	Text	24	Yes	Yes	**PK**	Fraction
	Lab_Replicate	Laboratory replicate identifier	Text	20	Yes	No	**PK**	
	Result	Result (concentration or equivalent) reported by the lab	Double precision	0	Yes	No	No	
	Sig_Figs	Significant digits of the result	Integer	0	Yes	No	No	
	Result_Units	Result unit of measurement	Text	24	Yes	Yes	No	Unit
	Basis	"Wet" for wet_weight basis reporting; "Dry" for dry_weight reporting.	Text	10	Yes	Yes	No	Meas_Basis
Lab_Result	Lab_Qualifiers	Qualifiers and flags assigned by the laboratory	Text	12	Noa	Yes	No	

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
(continued)	TIC	Is this a tentatively identified compound (TIC)?	Boolean	0	Yes	No	No	
Lab_Result (continued)	Validated	Has this result been validated?	Boolean	0	Yes	No	No	
	Validation_Level	Stage of validation - electronic and manual	Text	24	Yes	Yes	No	Validation_Level
	Result_Qualifiers	Final validated result qualifiers/flags (e.g., J;U;ND;<;>)	Text	12	No	Yes	No	
	Undetected	Flag to distinguish detected and undetected results, based on lab or validation qualifiers	Boolean	0	Yes	No	No	
	Estimated	Flag to distinguish estimated and non-estimated results, based on lab or validation qualifiers	Boolean	0	Yes	No	No	
	Rejected	Flag to distinguish rejected and non-rejected results, based on lab or validation qualifiers	Boolean	0	Yes	No	No	
	Reportable	Flag to distinguish reportable and non-reportable results, based on validation or data quality review	Boolean	0	Yes	No	No	
	Validator	Validation company name	Text	100	No	No	No	Organization
	Validator_Comments	Comment on the validation assessment for this result	Text	255	No	No	No	
	Comments	Comments on the results that do not pertain to either the validation results or data quality review results	Text	255	No	No	No	
	QA_Comments	QA comment resulting from any data quality review conducted in addition to data validation	Text	255	No	No	No	
	Date_Received	Date and time that the analytical sample was received by the lab	DateTime	0	No	No	No	
	Date_Extracted	Date and time that the sample was extracted by the lab	DateTime	0	No	No	No	
	Date_Analyzed	Date and time that the analysis was performed by the lab	DateTime	0	No	No	No	
	Dilution_Factor	Effective test dilution factor	Double precision	0	Yes	No	No	
	Lab_Name	Laboratory that performed the analysis	Text	24	Yes	Yes	No	Lab
	Lab_Sample_ID	Sample identifier assigned by the laboratory	Text	50	No	No	No	
	Original_Lab_Result	Original result reported by the laboratory, if data validation resulted in restatement of the value	Double precision	0	No	No	No	
	Lab_Cal_Batch	The laboratory's calibration batch identifier	Text	50	No	No	No	
	Lab_QC_Batch	The laboratory's quality control batch identifier	Text	50	No	No	No	
	IDL	Instrument Detection Limit (IDL)	Double precision	0	No	No	No	
	IDL_Units	IDL units	Text	20	No	Yes	No	Unit
	MDL	Method Detection Limit (MDL)	Double precision	0	No	No	No	
	MDL_Units	MDL units	Text	20	No	Yes	No	Unit
	Quantitation_Limit	Quantitation limit as determined by the lab	Double precision	0	No	No	No	
	Quantitation_Limit_Units	Quantitation limit units	Text	20	No	Yes	No	Unit
	Reporting_Limit	Reporting limit as determined by the lab	Double precision	0	No	No	No	
	Reporting_Limit_Units	Reporting limit units	Text	20	No	Yes	No	Unit
	Value_Type	result value type (e.g.., actual; estimated; calculated)	Text	24	No	Yes	No	Result_Value_Type
Calculated_Chemistry	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for this interpretive sample	Text	50	Yes	No	**PK**	Field_Sample
	Material_Analyzed	Material analyzed	Text	24	Yes	Yes	**PK**	Sample_Material
	Fraction_Analyzed	Indicator of what fraction of the sample was analyzed (e.g., total, dissolved, leachate, sieved size interval)	Text	24	Yes	Yes	**PK**	Fraction
	Analyte	Analyte/Parameter name (e.g., Lead; Arsenic; etc.)	Text	28	Yes	Yes	**PK**	Analyte
	Calculation_Method	Standardized description of the calculation method used (data selection and summarization)	Text	24	Yes	Yes	**PK**	Calc_Method
	Result	Calculation result (concentration or equivalent)	Double precision	0	Yes	No	No	
	Undetected	Flag to distinguish detected and undetected results, propagated through the calculation process	Boolean	0	Yes	No	No	
	Estimated	Flag to distinguish estimated and non-estimated results, propagated through the calculation process	Boolean	0	Yes	No	No	
	Rejected	Flag to distinguish rejected and non-rejected results, propagated through the calculation process	Boolean	0	Yes	No	No	
	Limited_Analytes	Flag to distinguish if the number of analytes reported is "limited" and therefore below the expected number of analytes for analyte group totals	Boolean	0	Yes	No	No	
	Result_Units	Result unit of measurement	Text	20	Yes	Yes	No	Unit
	Basis	"Wet" for wet_weight basis reporting; "Dry" for dry_weight reporting.	Text	10	Yes	Yes	No	Meas_Basis

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Tox_Test_Batch	Lab_Name	Toxicity test laboratory	Text	24	Yes	No	**PK**	Lab
	Tox_test_batch	Laboratory-specific toxicity test batch identifier	Text	50	Yes	No	**PK**	
	Tox_test_type	Type of toxicity test conducted	Text	24	Yes	Yes	No	Tox_Test_Type
	Taxon	Taxon code for organism used for the toxicity test	Text	24	Yes	Yes	No	Taxon
	Life_Stage	Life stage of organisms used for the toxicity test	Text	24	Yes	Yes	No	Life_Stage
	Start_Date	Starting date of the toxicity test	DateTime	0	No	No	No	
	End_Date	Ending date of the toxicity test	DateTime	0	No	No	No	
	QA_Level	Level of data quality review used	Text	24	Yes	Yes	No	QA_Level
Tox_Test_Result	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50	Yes	No	**PK**	Field_Sample
	Lab_Name	Toxicity test laboratory	Text	24	Yes	Yes	**PK**	Tox_Test_Batch
	Tox_Test_Batch	Laboratory-specific toxicity test batch identifier	Text	50	Yes	No	**PK**	Tox_Test_Batch
	Tox_Test_Variable	The variable in which results are expressed (e.g., survival, growth, reproduction)	Text	24	Yes	Yes	**PK**	Tox_Test_Var
	Tox_Test_Measurement	The type of measurement made (e.g., count of survivors, percent normal, change in mass)	Text	24	Yes	Yes	**PK**	Tox_Test_Meas
	Replicate	Replicate identifier	Text	20	Yes	No	**PK**	
	Tox_Test_Result	The numeric result of the toxicity test	Numeric	0	Yes	No	No	
	Sig_Figs	Significant digits of the tox test result	Integer	2	Yes	No	No	
	Tox_Test_Units	The units for the toxicity test result	Text	20	Yes	Yes	No	Unit
	QA_Level	Level of data quality review used	Text	24	Yes	Yes	No	QA_Level
	Tox_Test_Results_Comments	Comments on this toxicity test result	Text	255	No	No	No	
Species_Abundance	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	No	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50	Yes	No	**PK**	Field_Sample
	Taxon	Taxon code for organisms	Text	24	Yes	Yes	**PK**	Taxon
	Sex	Sex of organisms	Text	1	Yes	Yes	**PK**	Sex
	Life_Stage	Life stage of organisms	Text	24	Yes	Yes	**PK**	Life_Stage
	Abundance_measurement	Type of abundance measurement (count, concentration, density, spatial coverage)	Text	24	Yes	Yes	**PK**	Abund_Meas
	Replicate	Replicate identifier	Text	20	Yes	No	**PK**	
	Abundance	Abundance measurement	Double precision	0	Yes	No	No	
	Sig_Figs	Significant digits of the Abundance	Integer	2	Yes	No	No	
	Abundance_units	Abundance measurement units	Text	20	Yes	Yes	No	Unit
	Lab_Name	Laboratory that measured the abundance	Text	24	No	Yes	No	Lab
	Species_Abundance_Comments	Comments on the species abundance measurement	Text	255	No	No	No	
Species_Measurement	Field_Event_ID	Concise identifier for the field sampling effort	Text	50	Yes	Yes	**PK**	Field_Sample
	Main_Sample_ID	Unique sample identifier for an interpretive sample	Text	50	Yes	Yes	**PK**	Field_Sample
	Taxon	The type of collection observation made (e.g., tide stage, sediment class)	Text	24	Yes	Yes	**PK**	Taxon
	Sex	The sex of the species measured	Text	1		Yes	**PK**	Sex
	Life_Stage	The life stage of the species measured	Text	24		Yes	**PK**	Life_Stage
	Species_Measurement	The species measured	Text	24	Yes	Yes	**PK**	Measurement
	Measurement_Method	The method used to make the measurement	Text	24		Yes	**PK**	Meas_Method
	Replicate	Replicate identifier to distinguish multiple measurements (e.g., “1”, “2”)	Text	10	Yes	Yes	**PK**	
	Measured_Value	The numerical result of the measurement	Double precision		Yes			
	Sig_Figs	Significant digits of the result	Integer		Yes			
	Units	Result unit of measurement	Text	20	Yes	Yes		Units
	QA_Level	Level of data quality review used	Text	24	Yes	Yes		QA_level
	Reportable	Flag to distinguish reportable and non-reportable results based on data quality review	Boolean	0	Yes			
	Species_Measurement_Comments	Description of the species measurement	Text	255	No	No	No	

Category (Database Table)	Data Element (Database Field)	Description	Field Type	Field Size	Required	Valid Values	Primary Key	Foreign key to
Well	Well_ID	Well identifier	Text	50	Yes	No	**PK**	
	Location	Sampling Location Code/Monitoring Location Code	Text	50	Yes	No	No	Location
	Vertical_Reference_Point	The surface or reference point from which depths are measured (e.g., water, sediment, soil, well benchmark)	Text	24	Yes	Yes	No	Vert_Ref_Pt
	Well_Hole_Depth_Measure	Depth below the vertical reference point to the bottom of the hole on completion of drilling	Double precision	0	No	No	No	
	Well_Hole_Depth_Measure_Unit	Units for Well_Hole_Depth_Measure field	Text	20	No	Yes	No	Unit
	Completion_Date	Date that well construction was completed	DateTime	0	No	No	No	
	Screen_Upper_Depth	Upper depth of the screened interval	Double precision	0	No	No	No	
	Screen_Lower_Depth	Lower depth of the screened interval	Double precision	0	No	No	No	
	Well_Use	Usage of well (e.g., private water supply, public water supply, monitoring, stock watering, injection)	Text	24	No	Yes	No	Well_Use
	Well_Status	Status of well usage (e.g., active, abandoned)	Text	24	No	Yes	No	Well_Status
	Well_Status_Date	Effective date of the latest well status assessment	DateTime	0	No	Yes	No	
Criteria	Critcode	Concise identifier for the set of criteria	Text	50	Yes	No	**PK**	Critiadef
	Analyte	Analyte or parameter name (e.g., Lead; Arsenic; etc.)	Text	28	Yes	Yes	**PK**	Analyte
	Crit_Value	Numeric value of the criterion	Double precision	0	Yes	Yes	No	
	Unit	Unit for the criterion	Text	10	Yes	Yes	No	Unit
	Meas_Basis	"Wet" or "Dry"	Text	10	Yes	Yes	**PK**	Meas_Basis
	Range_Limit	Limit of the criteria range (i.e., upper or lower)	Text	5	Yes	Yes	**PK**	
	Criteria_Comments	Comments pertaining to this criterion	Text	255		No		
	Hardness_Dependent	Is the criterion hardness dependent?	Boolean	0		No		
	Hd_m	Parameter "mA" for criteria maximum concentration or "mC" for criterion continuous concentration	Double precision	0		No		
	Hd_b	Parameter "bA" for CMC or "bC" for criterion continuous concentration	Double precision	0		No		
	Cf_b	Intercept for Conversion Factor calculation	Double precision	0		No		
	Cf_m	Slope for Conversion Factor calculation	Double precision	0		No		
	Ph_dependent	Is the criterion Ph dependent?	Double precision	0		No		
	Ph_m	Slope for pH-dependent calculation	Double precision	0		No		
	Ph_b	Intercept for pH-dependent calculation	Double precision	0	No	No	No	
Criteriadef	Critcode	Concise identifier for the set of criteria	Text	50	Yes	No	**PK**	
	Critdescrip	Description of this set of criteria	Text	255	Yes	No	No	
	Material	The material relevant to the criteria (e.g., surface water, sediment)	Text	24	No	Yes	No	Material
	Resource	Resource to which the criteria apply (e.g., sediment invertebrates, humans)	Text	24	No	Yes	No	Resource
	Guideline_Type	Guideline type for the criteria (e.g., chronic, continuous)	Text	24	No	Yes	No	Guidelinetype
	Crityear	Year that the criteria were published or promulgated	Text	4	No	No	No	
	Citation	Citation of the criteria (e.g., Van den Berg et al., 2006)	Text	24	Yes	Yes	No	
	Criteria_Def_Comments	Comments on the criteria	Text	255	No	No	No	

ATTACHMENT A

RESPONSE TO COMMENTS

	NW Natural Comments	Response to Comment and Actions Taken
1.	• ***EPA Question #1: Is there anything you'd like to track in the database that is not there?*** – *Including calculated totals in comprehensive databases can be cumbersome and confusing because of the variations in how calculations are made. These variations include the choice of constituents, the treatment of non-detected constituents, and the choice of toxicity equivalency factors. In addition, calculation methods may vary depending on the purpose (risk assessment versus nature and extent or remedial design). If stored in the database, documentation of calculation rules will need to be developed along with a list of unique calculation identifiers (equivalent to CAS RNs) and chemical names. If specific sums will be part of the database, it is very important that a process be established that ensures the correct summing rules are applied consistently with the Record of Decision.*	Agreed. The Portland Harbor Interim Database (PHIDB) has been designed to address this comment. Calculated results will be stored in a separate table from laboratory analytical results in a table named "Calculated_Chemistry," and calculated results will contain a description of the calculation method in a column named "*calculation_method.*" The valid values for the calculation method will be controlled to ensure that they are clear and complete, and an expanded definition will be provided in corresponding documentation.
2.	• ***EPA Question #2: Do you have any issues with the look up- values – like analytes….?*** – *It is very important that the look-up values are standardized and non-duplicated.* – *Our databases have thousands of look-up values. Based on the material provided, it appears that a small fraction of potential look-up values has been provided (this was noted in EPA's spreadsheet). What will be the process for developing and sharing a comprehensive list of valid values?* – *Some of the values shown in the spreadsheet have duplicate entries (different codes pointing to the same item). This needs to be corrected and future valid values lists will need to be checked to remove duplicated values. Who will be responsible for ensuring the look-up values are correct and valid?* – *Valid values should consist of short codes and longer descriptions that are defined in look-up or reference tables. EPA Region 10's list and data dictionary are not consistent in the use of short codes (good) and longer descriptions (undesirable) in valid values fields of data tables*	The PHIDB (and EDD) will use sets of valid values that are more extensive than those in the original draft. Those lookup values will be standardized and defined so that they are neither ambiguous nor duplicated. Once the list is completed in draft, they will be provided to Portland Harbor data users for review. [Note to EPA team—we haven't discussed the next step of public review—this is a placeholder to address the NWN comment]. The State and EPA technical teams are working together to ensure that there are not duplicate entrees. Valid values will be defined as relatively short codes, each with a lengthier description.
3.	• ***EPA Question #3: Are there fields that do not have valid values, but should?*** – *Due to limited time and incomplete database, we cannot assess this question at this time. See preceding comment regarding long, descriptive values.*	Understood.

	NW Natural Comments (cont.)	Response to Comment and Actions Taken
4.	• ***EPA Question #4: Are there places we should improve the data definitions?*** – *USEPA's Environmental Response Team's (ERT) Scribe web page includes a Region 10 database template in Microsoft Access format that is comprehensive and that defines the relationships between database tables (Region 10 DMP Auditor Rules Database, "R10 DMP RuleSet Project.mdb"). This was not provided in EPA's email. This Access template or something similar would provide a better framework for discussing elements of the database model.* – *Alternatively, as identified above, consider using EarthSoft EQuIS.*	The ERT Scribe template has been reviewed and elements of that data structure have been incorporated into the PHIDB.
5.	• ***EPA Question #5: Other feedback or potential concerns?*** – *Consider indicating the Region 10 Scribe database(s) that are in use and provide the data models for these databases.*	A customized version of Scribe will be developed that is based on the default Scribe template but customized for compatibility with the PHIDB.

	Oregon DEQ Comments	Response to Comment and Actions Taken
6.	**General** • *In addition to development of an electronic data template, a separate guidance document or "database rules" should accompany the template to provide information on default parameters, such as preferred units for each analyte group, elevation datum, summation rules, and other relevant inputs consistent with, and perhaps as an Appendix to the Remedial Design Guidelines and Considerations document.*	In addition to the EDD template and lists of valid values, a guidance document is being prepared and will provide additional information on data submission requirements.
7.	• *One of the goals of the ICIAP/DMP ASAOC between EPA, the State, and the City of Portland is to evaluate the needs of Portland Harbor data users to help inform development of a long-term information management system. Early coordination between this work and development of the Scribe data management tool is important to ensure that these two efforts complement rather than duplicate one another. DEQ also recognizes that there are many types of information that will be collected beyond analytical data and we support a comprehensive data management approach that integrates all of this information into a single platform.*	The PHIDB is planned to export data to a compatible Scribe format that will hold priority data. The PHIDB will be designed to accommodate more than analytical data: specifically including toxicity test data, species abundance data, and other types of measurements made on samples. One of the goals of the PHIDB is to allow a seamless integration of data into the final data management platform.
8.	**Questions:** • *Is the LabAnalysis tab intended to be an all-inclusive list or is this just a starting point for analyses? For example, organochlorine pesticide analytical method is shown as 8081B. It would be useful to allow use of (or evaluate whether to require) Method 1699M to avoid co-elution and get more reliable results*	The lists of valid values will include most analytes that have been, or may be, measured in the Portland Harbor Superfund Site. The database will be inclusive of differing EPA Methods to accommodate specific methods and procedures in EPA-approved QAPPs for PHSS remedial design and remedial action investigations.
9.	• *The result qualifier data element in the dictionary appears to have a different meaning in the Valid Values tab. Is this correct?*	The PHIDB has been revised to address this comment. The representation of result qualifiers will be standardized as Boolean (True/False) value, while also accommodating the text representations provided by analytical laboratories and data validators. The PHIDB contains a table entitled "Lab Result" with Boolean flag columns for *undetected*, *estimated*, *rejected*, and *reportable* qualifiers.
10.	***The following are some useful data fields that should be considered based on DEQ's experience querying the LWG's SCRA database*** • *A field for river mile to the nearest tenth of a mile and side of the river (i.e. E/W).*	Agreed. The PHIDB is designed to address this comment: columns for *river mile* and *river zone* (e.g., East, West and Navigation Channel) will be included.
11.	• *In addition to the fish tissue matrix, a field for specifying the fish species.*	The PHIDB design includes a column in the "Species Abundance" table named "*taxon*" to identify the species or higher-level taxon as appropriate.
12.	• *A way to identify the parent sample for duplicates.*	The PHIDB design is structured so that duplicate or split samples are related to one another through the use of a common "parent" sample ID and the structure of the "Field Collection," "Field Sample," and "Analytical Sample" tables.

	Oregon DEQ Comments (cont.)	Response to Comment and Actions Taken
13.	• *A field that identifies the calculation method for sums, TEQs, etc (i.e. how were non-detects handled or which subset of dioxin/furan congeners are included in a dioxin/furan TEQ calculation).*	The PHIDB design includes a column in the "Calculated Chemistry" table named "c*alculation method*" to describe the calculation method used for calculated results. This set of values will be controlled by a valid value list, and the set of valid values will be controlled to ensure that they describe the method used.
14.	• *Columns showing the MDLs and MRLs.*	The PHIDB design includes a column in the "Lab_Result" table named "*mdl*" to identify MDLs and reporting limits in the table of laboratory results.

	Port of Portland Comments	Response to Comment and Actions Taken
15.	***Is there anything you'd like to track in the database that's not there?*** • *Generally calculated totals are not tracked in comprehensive databases. The reason for this is that there are variations in how calculations are made (for example: in how non-detected constituents are handled) and calculation methods may vary depending on the purpose (risk assessment versus nature and extent or remedial design). This necessitates numerous cas_rns and chemical names. If specific sums will be part of the database, consider setting up a process to ensure the correct summing rules are used for the intended purpose.*	The PHIDB will store calculated results separate from laboratory analytical results in a table named "Calculated_Chemistry." Calculated results must contain a standardized description of the calculation method in a column named "*calculation_method.*" The valid values for the calculation method will be controlled to ensure that they are clear and complete, and an expanded definition will be provided in corresponding documentation.
16.	***Do you have any issues with the look up- values – like analytes….?*** • *Typically, databases we use have thousands of possible look-up values. Based on the attached material provided, it appears there is a small fraction of look-up values have been provided (as noted in EPA's spreadsheet). What is the process for developing and sharing a comprehensive list?* • *Is there specific uses for this database that would limit the look-up values? Consider identifying those uses now to better identify look-up values needed.* • *We observed that some of the values shown in the spreadsheet have duplicate entries (different codes pointing to the same thing).* • *Valid values should consist of short codes and longer descriptions defined in look-up or reference tables. EPA R10's list and data dictionary are not consistent in the use of codes (good) and descriptions (undesirable) in valid fields of data tables.*	The PHIDB (and EDD) will use sets of valid values that are more extensive than those in the original draft. Those lookup values will be standardized and defined so that they are neither ambiguous nor duplicated. Once the list is completed in draft, they will be provided to Portland Harbor data users for review. We agree that considering uses of the database at this time will better identify the look-up values needed and have incorporated this process into the database structure and in development of the valid values list. The State and EPA technical teams are working together to ensure that there are not duplicate entries. Valid values will be defined as relatively short codes, each with a lengthier description.
17.	***Are there fields that don't have valid values, but should?*** • *Unable to fully assess at this time, but see preceding comment regarding long, descriptive values.*	Understood.
18.	***Are there places we should improve the data definitions?*** • *Cannot make a complete assessment based on the materials provided.*	Understood.
19.	***Other feedback or potential concerns?*** • *Consider indicating the current R10 Scribe database(s) and the data models for this database.*	The ERT Scribe template has been reviewed and elements of that data structure have been incorporated into the PHIDB.

	RM11E Comments	Response to Comment and Actions Taken
20.	***Attribute Tables (Additional Data Tracking)*** *We assume that the Site, Events, Location, Samples, and LabResults table structures provided in the recent EPA spreadsheet will be more extensive. We recommend these tables contain the full complement of attributes, as shown in Attachment A, with some attributes being not required. When possible we believe it is important to retain information such as the date that Site data was input, the task identification associated with Samples, the map scale when used to derive Location coordinates, if a particular LabResult was reportable, and other attributes found in Attachment A but not found in the EPA spreadsheet.*	The EPA Region 10 Scribe template has been reviewed and elements of that data structure have been incorporated into the PHIDB. Location coordinates will be primarily derived from GPS. Locations derived from geo-referenced map will be categorized as having a low accuracy range, therefore tacking the map scale is unnecessary. The PHIDB contains a binary reportable field in the "Lab Result" table named "*reportable.*"
21.	***Site*** • *ContractorEmail: the email of the contractor performing work on the site, for a complete set of contact information.* ***Required.*** • *SiteOwner: the owner of the site.* ***Optional.*** • *SiteMap: an image portraying a base map of the site.* ***Optional.*** • *SiteImage: an aerial of the site.* ***Optional.***	The organization conducting a field investigation is an attribute of the field event rather than the site, and therefore the PHIDB will not contain contractor information in the "Site" table. Instead, the sponsoring organization will be identified in the "Field Event" table, and both the organization and the individual responsible for collecting each sample will be identified in the sample table. For the Portland Harbor Superfund Site, a "site" will be defined by river mile segments to a tenth of a mile. Consequently, sites will not have distinct owners. Instead of relating a party to a site, the database will associate both sites and parties to ASAOCs. Web and desktop map applications can provide dynamic representation of location and aerial imagery for a site. The static map solutions proposed by *SiteMAp* and *Siteimage* columns is unnecessary.
22.	***Events*** • *EventStatus: the current status of the particular event with attribute values of "In Progress," "Pending," "Completed," "Scheduled."* ***Required.*** • *EventName: a text field providing a name for the event.* ***Optional.***	The PHIDB will include a "*Status*" attribute for both administrative and field events. The PHIDB table named "Event Location" contains a concise identifier for the field sampling effort in the column named "*event ID*" and event-specific location code named "*event location.*"

	RM11E Comments (cont.)	Response to Comment and Actions Taken
23.	***Location*** • *LocationType: the type of sampling location (e.g., pore water, or monitoring well). Locations could be of multiple types, and two LocationType fields could be beneficial.* ***Required.*** • *LocationGeology: the geologic unit at the station.* ***Optional.*** • *Basin: the basin in which the station occurs.* ***Optional.*** • *Directions: the driving directions to the location.* ***Optional.*** • *SampFreq: the frequency of sampling at the respective location.* ***Optional.*** • *CurrentStatus: the status of the sampling location (e.g., Abandoned, In Service, Etc.).* ***Optional.*** • *QCCode: any quality control associated with the sample location.* ***Optional.*** • *PermitNumber: the number of any permits associated with the sampling location.* ***Optional.*** • *LocationReach: a numeric value describing the location along a stream or other linear network.* ***Optional.***	Please see response by subheading: LocationType: Descriptions such as "porewater," "groundwater," and "sediment" apply to samples, not to locations. A location could be sampled for a variety of environmental media. The types of material collected from a location can be identified by a query that finds the appropriate attribute for all samples from that location. In addition, the PHIDB allows multiple sampling reasons (e.g., site characterization, risk assessment, boundary delineation) to be associated with each sample, and these also can be summarized by location. LocationGeology: Geologic units vary with depth, therefore the geologic unit is better suited to be managed in a boring log. Currently the PHIDB does not accommodate boring log information. Basin: The database will identify the basin through the use of the 8-digit HUC code. Directions: Web and desktop map applications can provide dynamic driving directions. SampFreq: The frequency with which sampling has been conducted at a location can be obtained by querying the samples at each location and aggregating by the desired time intervals. CurrentStatus: Because locations are characterized as geographic locations rather than as specific uses, the status of a location with respect to a specific use is not an attribute of the location. QCCode: QC codes are associated with the sample not the location. The QCCode is stored in the validation_level, which is required and controlled by a list of valid values. Any additional information can be included in the "qa_comments" column. PermitNumber: The PHIDB is not planning to store permit numbers. LocationReach: The database will contain columns to identify the *river mile* and *river zone* (e.g., East, West and Navigation Channel).
24.	***Samples*** • *Geology: the geology at the sample.* ***Optional***. • *Lithology: the lithology at the sample.* ***Optional.*** • *CoolerID: the number to identify the cooler in which primary samples and QC samples were shipped.* ***Optional.*** • *COCTrackingNumber: the chain-of-custody tracking number.* ***Optional.*** • *SampleFiltered: filter information at the sample level.* ***Required.*** • *SampleResult: the result of the sampling effort such as "dry" or "no access."* ***Optional.*** • *TisueTypeCode: identifies the tissue type of a sample. Required when tissue is collected.*	Please see response by subheading: Geology: See LocationGeology response in the response to comment 23, above. Lithology: See LocationGeology response in the response to comment 23, above. Cooler ID and COC Tracking Number: Samples may be placed in multiple containers, and containers may go into different coolers with different COC IDs. The scope of the PHIDB (at this time) does not include tracking of sample containers, coolers, or COC forms. SampleFiltered: The "Field Sample" table includes a "*field treatment*" column that can be used to record field filtering, and the "Lab Results" table contains a "*fraction*" column that can be used to record measurements made on filtered samples. TissueType: Valid values for the "*sample material*" column will include tissue types.

	RM11E Comments (cont.)	Response to Comment and Actions Taken
25.	***LabResults:*** • *LabMatrixCode: matrix used during lab analysis.* ***Required.*** • *LeachMethod: the method used to leach the sample.* ***Required.***	LabMatrix: The "*material analyzed*" column will identify the material used by the laboratory. LeachMethod: The "*preparation_method*" column in the "Lab_Results" table can be used to identify the leaching method used.
26.	***Valid Values, Lookup Tables, and Required Fields*** • *Valid values for sediment matrix types use different units and are not mutually exclusive.* • *Other should possibly be excluded as a valid value for Location.ElevDatum. If a datum is used besides those given in the valid values list, there should be a mechanism to add the new datum to the database.* • *Valid values are missing for Location.Stream_or_Creek_Name, Location.USGS_StationID, and Location.USGS_StationName.* • *The valid values for LabResults.Total_Or_Dissolved appear to be a combination of several things. For instance, Inorganic is a valid value, but inorganic analysis could be performed on both filtered or unfiltered samples. The Total_Or_Dissolved field usually refers to whether the analysis is performed on the total or dissolved fraction of the sample matrix.* • *Fields that are not required and we feel should, at times, be required are as follows:* • *ElevDatum: when an elevation datum is used, we feel this field should be a required field.* • *Samp_Depth_From: this should be a required field for soil and sediment samples.* • *Samp_Depth_To: this should be a required field for soil and sediment samples.* • *Samp_Depth_Untis: this should be a required field for soil and sediment samples.*	"*Total or dissolved*" column has been replaced by a "*fraction*" column and a "*measurement basis*" column, and consistent unambiguous codes will be defined for each. ElevDatum: Elevation values will require a datum reference in the "*elev datum*" column of the "Location" table. Sample depths and units: this is a constraint that will be checked by the data loading routine, not through check constraints in the database.
27.	***Data Definitions*** *The current Scribe data model contains definitions that appear to be relatively comprehensive. Some definitions, such as LinkSampleNo, should have expanded definitions, highlighting how the field can be used (e.g., LinkSampleNo can be used to store linked sample identifiers such as the parent sample of a duplicate sample. We recommend defining attributes that are not currently defined.*	The data definitions have been revised and updated in the current PHIDB—the reviewer is referred to the current PHIDB data dictionary.

	RM11E Comments (cont.)	Response to Comment and Actions Taken
28.	***Mapping Recommendations*** *We recommend a standard horizontal coordinate projection system for Portland Harbor, particularly if all data is going to be stored in a centralized database. Currently, the data on the EPA Interim portal contains different coordinate projection systems, and this does present confusion.*	The World Geodetic System of 1984 (WGS84) is chosen for the horizontal datum and coordinate system. WGS84 is the native Global Positioning System coordinate system and the modern horizontal datum for latitude and longitude (spherical) coordinates.
29.	***Additional Questions*** *We also have some additional questions for EPA regarding the Portland Harbor Scribe data structure as follows:* *1. Please confirm that EPA is not proposing a reduction in attributes or tables from the full Scribe data model (AttachmentA.pdf) for the Portland Harbor database.*	The PHIDB will incorporate elements of the Scribe data model, but will be customized to include other information and a more detailed representation of the relationships between samples.
30.	*2. Would this database be used to assess performance?* *a. Does the database achieve this purpose on a sediment management area (SMA) or site-wide basis?*	The PHIDB database will provide a source for data on performance. However, there is no application to specifically assess performance.
31.	*3. What are the procedures for deriving analyte summations, including procedures for the treatment of non-detects?*	Calculated results will be stored in a separate table from laboratory analytical results entitled "Calculated Chemistry," and calculated results will contain a description of the calculation method in a column entitled "*calculation_method.*" The valid values for the calculation method will be controlled to ensure that they are clear and complete, and an expanded definition will be provided in corresponded documentation.
32.	*4. Will the final Scribe data be downloadable or accessible as a single database?*	Yes, data contributors will be able to access the Scribe version of the PHIDB.
33.	*5. Will EPA be providing the values for EventID and Activity in the Events table?*	See the response to comment 22, above, regarding Event-Location.

	Yakama Nation Fisheries Comments	Response to Comment and Actions Taken
34.	***DataElements Dictionary - Analysis Section*** *In the Sample section, this data structure has a common problem of not dealing well with composite samples. While "composite sample" is a sample type, there needs to be a way of recording the various subsamples that are composited, including lat/long of collection and all the other sample attributes. This will be important for things like water samples collected across transects, fish tissue composites created from multiple fish collected in a variety of locations, etc.*	The PHIDB will contain a "Sample Composite" table that identifies the individual samples that make up a composite sample.
35.	*In the Analysis section, wet vs. dry weight should also be recorded for sediments, sediment trap samples (i.e., particulates), and tissue samples.*	All solid matrices will be required to have a measurement basis of either "Wet" or "Dry" specified.
36.	*In the Analysis section, there should be a separate field for the qualifier assigned by the laboratory and that assigned by the third-party data validator. Sometimes they are different.*	The PHIDB is designed to address this comment. The table of analytical results ("Lab Result") has separate columns for the lab qualifiers named "*lab qualifiers*" and for validation qualifiers named "*result qualifiers.*"
37.	*The following fields should be added to make it easier to query usable results:* • *Detect = Y/N* • *Reportable Result =Yes/No.* • *ValidatorComments to provide additional information from the validator that may not be present in a validation report.* • *Separate the result qualifier field into two fields: laboratory flag and validation qualifier. The data user may want to see the difference between what the laboratory flagged the results and how the validator qualified the result.*	The PHIDB is designed to address this comment. The table of analytical results ("Lab_Result") has separate columns for Boolean flag columns named "*undetected*," "*estimated*," "*rejected*," and "*reportable*" qualifiers. The "Lab Result" table also contains a column named "*validator comments.*" The list of valid values will address this comment.
38.	***Valid Values*** *We recommend defining valid values or providing a description, example, or guidance to make use of valid values clearer. For example, under LabResults for the Data Element Result_Qualifier, the qualifier flags of J, J-, and J+ would benefit from explanation or guidance on when to use the appropriate value.* *Under Events, there should be more categories for other studies not directly associated with the CERCLA cleanup but which may be relied on in various ways. For example NRDA studies, agency or university research, source control evaluations, dredged material evaluations, TMDL studies, etc.*	This comment will be addressed after preliminary data are loaded in the PHIDB and the list of valid values is completed.

	Yakama Nation Fisheries Comments (cont.)	Response to Comment and Actions Taken
39.	*Under Events, there should similarly be more options for who approves a QAPP, depending on what type of study it is. Perhaps an "other" write-in field.*	The PHIDB is designed to address this comment. The table defining field events (d_field_event) contains a "*QAPP_Approved_Other*" column to allow identification of other organizations that have reviewed and approved the QAPP.
40.	*Under Location, if you're going to allow distances to be measured in either feet or meters, ideally there should be a way for the database to convert so that all measurements are in the same units when output to the user. If not, I suggest requiring entry in one consistent unit, preferably metric. Similar comments wherever English vs. metric may occur as options.*	The PHIDB will store geographic coordinates in a single spatial reference system (e.g., WGS84 decimal degrees), and distances between points can be obtained in any units. Where units are explicitly specified (e.g., for depths), conversion to other units can be done during data summarization using conversion factors for each type of unit.
41.	*Under Samples, Activity – as noted above there are a variety of other types of sampling events that could be carried out within the large Portland Harbor site, that should be able to be stored in this database. The types of allowed activities needs to be expanded as above – or simply rely on the events field.*	As suggested in this comment, sampling activities will be represented in the "Event" table. Field events will have descriptive attributes specifically relevant to sample collection. In addition, the sample ("Field Collection") table will contain a column that can be used to provide additional detail about the sampling design.
42.	*Under Samples, Matrix Description:* • *There are more kinds of shellfish tissues than just clams. Suggest expanding at least to bivalves.* • *It is unclear where crustaceans such as crayfish or other benthic invertebrate tissue would be placed.* • *Floating product is listed but not DNAPL.* • *Tissue – unk. Origin is a poor descriptor. Sufficient taxonomic levels should be included to describe the sample. It's not really the case that the origin would be unknown.* • *Just having one category for fish seems insufficient to cover the various trophic levels and life stages that may be sampled (e.g., eggs, smolt, juvenile, adult, anadromous, resident, etc.).* • *There have been a fair number of eggshell studies on the Willamette.* • *Particulates (i.e., from water filtrate or sediment traps) should be included.*	The list of valid values for taxa will be expanded to include those that have been, or plausibly may be, collected in Portland Harbor. The lists of valid values for tissue (material sampled) will be as specific as possible. Two levels of characterization will be supported: one specific and one general. *Life stage* will be included as a key (required) column for all measurements of species abundance and organism characteristics. The list of valid values for sample materials will include particulates.
43.	*Under Samples, Depth Units, this should not only be in feet. Ideally meters or centimeters would be used rather than feet. Sediment sample depths are typically reported in the metric system in cm. Either the data entry user will need to convert or the database will, but all common units should be allowed OR only one metric unit for consistency.*	The PHIDB will accommodate a variety of units and allow standardization (conversion) of units during data summarization. Accommodating the units used by each investigation will best support QA checks.

	Yakama Nation Fisheries Comments (cont.)	Response to Comment and Actions Taken
44.	*Under Samples, Sample Collection – the volume or dimensions of the sampling device can be important for data interpretation purposes. Common samplers (such as van Veen grabs) should include dimensions.*	The dictionary of valid values for sampling gear will contain columns to characterize the volume of the gear. In addition, the PHIDB will accommodate values for the mass and volume of each individual sample.
45.	*Under Samples, Sampler – Please allow this to be a write-in field. It is important to know who conducted the sampling. "RP" is insufficient – the consultant actually doing the work in the field needs to be listed. This is also not inclusive of many other federal/state/local agencies, universities, and other parties that frequently collect data in Portland Harbor. There should be a "study sponsor" field that describes who is sponsoring the study separate from the actual entity doing the sampling. That field should list the specific RP if the sponsor is an RP.*	The PHIDB will contain a "*study sponsor*" column for each field sampling event and both a "*sampler organization*" and "*sampler person*" column for each sample. These columns will allow write-in values.
46.	*Under Samples, Type – all lab QC samples should be listed here along with field QC samples. There should be a way to enter sample composites and link them to individually collected subsamples in space and time. SampleType should also include Split as a valid value to indicate a sample that is split and analyzed by two different laboratories or by different entities.*	The PHIDB will contain only analytical results for field samples, not for laboratory QC samples.
47.	*Under LabResults, Result_Units – ensure that all samples for one chemical have an enforced unit so that different units are not entered for the sample chemical by different users. Make sure there is a ppt unit for dioxins/furans. Use the "micro" symbol rather than a ug/l. Also please use standard SI notation – e.g., L rather than l for liters. Same comments for all lab units fields.*	There are conditions under which results for an analyte may be reported in different units, such as mg/kg and micromoles/g for metals in sediment. The PHIDB will allow data to be reported and stored in the units used by the original investigator, and will allow unit standardization (conversion) during data summarization.
48.	*Under LabResults or somewhere, there needs to be a way to identify if data have been OC-normalized or lipid-normalized. Or any other form of normalization, such as grain size fractions.*	The table of laboratory results will include a "*fraction*" column to identify whether the data were measured on, or reported on, a specific fraction of the entire sample.
49.	*Under LabResults for the Data Element QA_Comment, multiple valid values might apply to a single entry. We suggest defining when a value is to be used. For example, data could be both final and validated, and without further definition, it would be difficult to apply the appropriate valid value consistently.*	The "*QA codes*" column will be a free-form text column, not controlled by a list of valid values. Other columns will store information about the validation status and validation level in a systematic form, controlled by valid values.
50.	**refAnalytes** The list should be sorted and numbered by analyte category as currently it is difficult to see whether all important analytes in a category are included with different analytes in a category separated in the list.	N/A
51.	COPC is an inappropriate term for the third column. Not all analytes may be COCs, many are included in standard analyses or are physical or chemical parameters used to interpret other variables.	The term COCs is not used in the PHIDB.

	Yakama Nation Fisheries Comments (cont.)	Response to Comment and Actions Taken
52.	*COPC analytes such as TBT and BEHP should be spelled out rather than an abbreviation. If a short name for an analyte is desirable perhaps a field called Short Name would be available in the refAnalytes table.* *Hopefully TBT is not the only butyltin being measured/tracked. All butyltins should be included. Also, why are TBT and tributyltin two different entries? Are TBT and tributyltin synonyms in the refAnalytes table? TBT is listed on row 19 and tributyltin is listed as row 98.*	The PHIDB analyte look up table contains a "*full name*" column. For example, BEHP is named "bis(2-Ethylhexyl)phthalate" in this column. The full suite of butyltins will be able to be stored in the PHIDB.
53.	*Additional information for "analytes" that require calculation (such as totals or TEQ calculations) should be identified to ensure calculation and the handling of non-detect values are consistent.*	Summing specifications will be documented, and SQL scripts will be developed to carry out these operations. The rules for data summarization will not necessarily all be stored in the database.
54.	*There needs to be a way of distinguishing metals species where important (e.g., Chromium III vs. VI, inorganic vs. organic mercury and arsenic). These are often different analytical results.*	The set of valid values for analytes will distinguish different metal species.
55.	*It should be possible to list more PAHs than these if more complex studies of PAHs are conducted (e.g., parent compounds, creosote and tar tracers, etc.).*	All analytes listed in QAPPS will be supported in PHIDB.
56.	*For consistency with other similar totals, cPAHs (BaP eq) should be listed as cPAHs TEQ. For TEQs, it should be noted whether they are fish, avian, or mammal/human.*	All analytes listed in QAPPS will be supported in PHIDB.
57.	*Similarly, the list of TPH and other petroleum fractions is quite incomplete*	All analytes listed in QAPPS will be supported in PHIDB.
58.	*All PCB Aroclors and congeners should be individually stored, not just totals. Methods of summing vary and the original data need to be available for review.*	All analytes listed in QAPPS will be supported in PHIDB.
59.	*The DDx listed are inconsistent and incomplete. Ensure that there is only one entry for each DDx and that both ortho- and para- congeners are included.*	All analytes listed in QAPPS will be supported in PHIDB.
60.	*BEHP and dibutyl phthalate are not the only two phthalates present in Portland Harbor or reported by laboratories.* *Conventionals are entirely missing from this list but are crucial to data interpretation. At a minimum, percent solids, grain size fractions, AVS/TVS, total sulfides and ammonia, lipids, organic carbon, pH, DO, etc. should be included.*	All analytes listed in QAPPS will be supported in PHIDB.
61.	*Overall, please review large Portland Harbor data sets and include ALL analytes reported by laboratories on this list, even if non-detected. Data should be stored as reported by the laboratories, and there need to be sufficient categories for all types of data.*	All analytes from prior investigations imported to PHIDB will be supported.

ATTACHMENT B

EDD Specifications for the PHIDB

EDD SPECIFICATIONS FOR THE PORTLAND HARBOR INTERIM DATABASE (PHIDB)

Prepared for

State of Oregon—Department of State Lands

775 Summer Street NE, #100, Salem, Oregon 97301

Prepared by

Integral Consulting, Inc. 719 2nd Avenue, Suite 700 Seattle, Washington 98104	**GeoEngineers, Inc.** 5820 S Kelly Avenue, Suite B Portland, Oregon 97239

December 7, 2021

CONTENTS

TABLES

FIGURES

APPENDICES

1.0 INTRODUCTION

The Portland Harbor Interim Database (PHIDB) is designed specifically to store environmental characterization data that are collected under a U.S. Environmental Protection Agency (EPA) consent order during remedial design (RD) sampling in the Portland Harbor Superfund Site (PHSS). Compilation of data from multiple RD studies into a single system will establish consistency in data structure and encoding, and thereby facilitate consistency in data interpretation and presentation.

Because data will be collected by numerous parties, all of whom may manage data in different systems and structures, a key step to establishing consistency within the PHIDB is for parties to submit their data in a standard digital format, or electronic data deliverable (EDD). This document describes the PHIDB EDD format for data submission. This document is organized by the following sections:

- **Data Structure (Section 2).** This section provides background on the typical scope, complexity, and organization of environmental investigations and how the PHIDB system integrates this complexity. The location identifier and coordinates, collections and samples, quality control (QC) samples and splits, and use of multiple sample identifiers are presented in Section 2.

- **Sample Specific Details (Section 3).** This section provides details related to the specific collection method in the EDD format.

- **EDD Specifications (Section 4).** The EDD Specifications describe the data dictionary for the EDD tables and the valid values and metadata for the PHIDB.

- **Data Submittal and Review (Section 5).** The data submittal section describes the process for data submittal and review.

An electronic template for the EDD tables is available to assist with the preparation and submission of RD data. It is recognized that the EDD templates and valid values will be reviewed and updated throughout the life of the PHSS cleanup. The current templates and valid value lists can be downloaded from the Portland Harbor Environmental Data Portal[1] as part of the Portland Harbor Remedial Design Programmatic Data Management Plan. In addition, detail on the design of the PHIDB can be found in the document *Portland Harbor Interim Database (PHIDB) Design Summary*[2] *(GeoEngineers/Integral, 2021)*, which can be downloaded from the Portland Harbor Environmental Data Portal.

[1] http://ph-public-data.com/

[2] GeoEngineers and Integral Consulting, 2021. Portland Harbor Interim Database Design Summary. Prepared for State of Oregon. Version 2.0, December 16, 2021.

2.0 DATA STRUCTURE

Environmental sampling programs often incorporate resampling, field replication, subsampling, compositing, and sample splitting for QC purposes. Different investigators and different investigations commonly use different conventions for naming locations and samples, and to represent the relationships between samples, subsamples, composites, and splits. Software used to manage environmental data may also enforce conventions or establish constraints on how information is represented. Because the PHIDB system integrates data from multiple investigations, those different conventions must be unified in a consistent manner.

Two elements of sampling complexity that are addressed by the PHIDB structure, and therefore are represented in the EDD, are:

- Subsampling of collected material. The most common example of this in sediment sampling programs is the collection of a core that is then subsampled at different depths. Each sample that represents a core horizon must be distinguished from all other such samples, but must also be unambiguously related to all other samples from the same core. Sampling programs frequently use sample identifier schemes in which one field within the sample identifier is common among all samples from a core. A database system such as the PHIDB cannot feasibly be designed to interpret fields within sample identifiers, particularly when sample identifier designs differ between investigations. These relationships are therefore represented by the data structure design rather than by the sample identifier design to enforce data integrity.
- Splitting of samples for QC purposes. Ordinarily 5 percent of samples collected for chemical analyses is homogenized and split in the field prior to submission to a laboratory. These splits, or duplicates, must be distinguished from one another, but must both be linked so that they are identifiable as representative of the same portion of the environment that will be used for characterization and assessment. As with subsamples, the relationships between splits are represented by data structure rather than by the content of sample identifiers.

The following sections describe the way that the PHIDB system represents some of the complexity of sampling data structure. The EDD format is a simplified representation of this structure.

2.1 LOCATION IDENTIFIERS AND COORDINATES

The EDD format includes a single table for location information in which event location identifiers and target coordinates[3] are required. Target coordinates represent intended sampling locations and are commonly included in Field Sampling Plans (Table 3—**location table**).

The actual sample location coordinates (e.g., collected with a Differential Global Positioning System) at which each sample is collected are also recorded in the EDD in Table 5 (**sample table**) with the sampling information. Both target and actual coordinates are required entries and can be the same when the sample is collected as planned. Whereas target coordinates are always represented as a single geographic point, actual sampling locations can be represented as the center of one or more points (e.g., composite sample), a line midpoint (e.g., transect), or area centroids (e.g., test pit). All coordinate data should be submitted in one coordinate system as decimal degrees in the WGS84 (World Geodetic System 1984) coordinate system and horizontal datum. Vertical datum elevations may be submitted in the North American Vertical Datum of 1988 (NAVD88) or the National Geodetic Vertical Datum of 1929 (NGVD29).

2.2 COLLECTIONS AND SAMPLES

When a sediment core is collected and subsampled into separate horizons, the core as a whole is referred to as a *collection.* The individual horizons are referred to as *samples* or *interpretive samples*[4]. Therefore, one core *collection* has a one-to-many relationship to the many *interpretive samples* within the core. The same terminology (*collection* and *sample*) is used for soil borings or sediment cores. The same terminology is also used in other cases where material is subsampled, such as when a fish is subdivided into fillet and carcass subsamples. A single fish *collection* has a one-to-many relationship to the subdivided fillet and/or carcass *interpretive samples*. For uniformity of data representation within the PHIDB, the same terminology is also used for single samples, such as surface sediment grab samples—in those cases the collection and the interpretive sample are the same.

Sampling programs may also include composite sampling, which involves collection of multiple individual samples (sub-composites) and compositing them into one sample (the composite sample) for chemical analysis; the data submittal process for composite sampling within the EDD structure is detailed in Section 3.1.

Interpretive samples can themselves be further subdivided into *analytical samples*. An interpretive sample may be split into multiple analytical samples as part of a QC program in which (typically) 5 percent of the interpretive samples is split and the two resulting analytical samples are submitted independently to the laboratory for "parent" and "duplicate" analysis. Interpretive samples may also be split into multiple analytical samples if material is to be sent to

---

[3] The terms "target coordinates," "canonical coordinates," and "design coordinates" all refer to the coordinates at which samples are intended to be collected. These are the coordinates that are ordinarily listed in a field sampling plan.

[4] The phrase "interpretive sample" indicates that this is the material that will be used for interpretation of environmental conditions.

different laboratories, or if different types of containers or preservatives must be used for different analyses.

This hierarchy, which includes three levels—collections, interpretive samples, and analytical samples—is used for samples in the PHIDB. The application of this hierarchy to both grab and core samples is shown on Figure 1 Common Structure for All Sampling Information. For surface sediment grab samples, the *collection* and the *interpretive sample* are the same thing (e.g., one collection to one interpretive sample with the same identifiers), and if split there will be one or more analytical samples per interpretive sample (e.g., one interpretive sample to one or more analytical samples with different identifiers). For sediment cores, the collection represents the entire core, and there will be multiple interpretive samples for that collection (e.g., one core collection to many interpretive sample[s] with different identifiers). And, again, there could be one or more analytical samples per interpretive sample.

The EDD **sample table** (Table 5) contains information to describe both collection identifiers (e.g., collection_id) and the sample identifiers (e.g., main_sample_id) derived from those collections. For collections that are subsampled, such as sediment cores, a single collection identifier should be assigned to all samples from that collection, and each of those samples must also be assigned its own unique identifier as shown on Figure 1.

When the collection and the sample are equivalent (e.g., for sediment grab samples), the collection and sample identifier should be the same (e.g., one-to-one relationship), as shown on Figure 1, and for convenience, only the collection identifier needs to be entered into the EDD sample table.

Separate upper and lower depths for collections and samples can be recorded in the EDD. The upper and lower depths for a core, for example, will not be the same as the upper and lower depths for any horizon from that core (i.e., the upper and lower depths of the interpretive sample[s] from the core). The EDD **sample table** (Table 5) contains an upper collection depth column and a lower collection depth column to store the minimum and maximum core depth, and it also includes an upper sample depth column and a lower sample depth column to store the sample-specific depths.

Separate sample materials for collections and samples can be recorded in the EDD. For example, if a water sample is filtered in the field, the material for the collection would be whole water, whereas the samples would have materials of filtered water and filterable particulates.

One approach the submitter may take is to prepare separate tables for the collection and sample information, and then join the tables using the main sample identifier to create the EDD sample table. Only the complete EDD sample table can be submitted.

2.3 QUALITY CONTROL SPLITS FOR CHEMISTRY ANALYSES

When a field interpretive sample is split to create duplicate samples that are to be analyzed separately as part of the QC program (i.e., *analytical samples*), the splits must have different analytical sample identifiers assigned but must also share the same main (interpretive) sample identifier. The PHIDB refers to field duplicates and splits as *analytical samples* where the main sample identifier contains the parent (or natural) sample identification. Field sampling programs commonly append a suffix such as "-D" to the main sample identifier to label one of the split or duplicate samples, using the main sample identifier for the other split sample, which is commonly referred to as a parent or natural sample. The PHIDB data structure enforces the duplicate or split relationship to parent sample by the data structure design rather than by the sample identifier design. Therefore, when a sample is split, each of the splits must be assigned both its own unique identifier and the identifier of the main sample, which is common between the two splits, as shown on Figure 1.

The EDD format includes separate tables for sampling information in the **sample table** (Table 5) and for analytical chemistry results for environmental samples in the **lab result table** (Table 11). The main sample identifier appears in both tables and serves to link analytical results to interpretive and analytical (e.g., duplicate) samples. The analytical sample identifier appears only in the table of analytical chemistry results, where it distinguishes between results for different splits of the same interpretive sample. Therefore, only the collection and interpretive sample identifiers are required in the **sample table**.

2.4 USE OF MULTIPLE SAMPLE IDENTIFIERS

As described above, up to three sample identifiers are used in the PHIDB EDD format: the collection identifier, the main sample identifier (which corresponds to the interpretive sample), and the analytical sample identifier. For a core horizon that has been split, these three identifiers may all be different. For a surface sediment grab sample that has not been split, these identifiers will ordinarily all be the same (Figure 1).

Many environmental investigations use fields within a sample identifier to encode information that can be used to distinguish collections, interpretive samples, and analytical samples. For example, each sediment core may not be explicitly assigned a unique identifier, but a core (or location) identifier may be embedded within the main sample ID, so that all interpretive samples from that core have the same value within that field of the sample identifier. When data are prepared in the PHIDB EDD format, the preparer may need to create distinct identifiers to properly represent the relationships between collections, interpretive samples, and sample splits (e.g., *analytical sample*). For example, the field within the sample identifier that identifies the core may be adopted as the core (*collection*) identifier. To minimize the effort required by data preparers, the EDD allows the following simplifications (Figure 1):

- The collection and main sample identifiers are the same when there is one collection to one interpretive sample (e.g., as for a sediment grab sample)—then only one identifier, the main sample identifier, needs to be provided.
- The main sample identifier and the analytical sample identifier are the same when there are no duplicates or splits; only the main sample identifier need be provided.

3.0 SAMPLE SPECIFIC DETAILS

This section provides recommendations for how to submit composite samples and how to submit collections and samples for different types of material and sampling gear.

These examples describe common sampling designs and materials.

3.1 COMPOSITE SAMPLES

There are two ways to submit the results of sample compositing:

1. Submit only the composite sample. The composite sample is required to be included in the sample table, including a value for the "composite_type" column and optionally for the "composite_count" column. If the sample is a temporal composite, consider populating the "composite_period" column.
2. Submit both composite and subcomposite samples. This option requires submitting both the **sample table** (Table 5) and **sample composite** table (Table 6). In addition to submitting the composite sample itself, all of the subcomposite samples that make up the composite sample need to be included in the **sample table**. The relationships between the composite and subcomposite samples are described by entries in the **sample composite** table.

Which of these methods is used may depend on directions from EPA or the discretion of the data submitter. For example, the first approach may be taken for a simple three-point composite around a target location, and the second approach may be taken for multi-increment sampling because the subcomposites are taken over a much larger area. However, the second approach can also be adopted for three-point composites at the discretion of the data submitter.

If the second approach is taken, each of the subcomposite samples must have target coordinates and other required information recorded for creation of a row in the location table for each subcomposite sample. Actual coordinates for the collection are recorded in the **sample table** (Table 5).

There are different sample scenarios based on the type of sampling:

A. Multipoint Composite:

   a. Both composite and subcomposite are part of the same collection
   b. Subcomposite coordinates are represented by "sample_lat_centroid_dd" and "sample_lon_centroid_dd" columns or by "coll_coords" column. The composite sample is a spatial average of the subcomposite coordinates. Subcomposite depths are recorded in the **sample table** (Table 5). The composite depth should be representative of all the subcomposites (e.g., an average of maximum value).

B. Multi Increment Sampling:

   a. Composite and subcomposite samples are part of different collections.

   b. Subcomposite coordinates are represented by “sample_lat_centroid_dd” and “sample_lon_centroid_dd” columns or by “coll_coords” column. The composite sample is a spatial average of the subcomposite coordinates. Subcomposite depths are recorded in the **sample table** (Table 5). The composite depth should be representative of all the subcomposites (e.g., an average of maximum value).

The “coll_coords” column in the **sample table** (Table 5) is formatted as “well-known text” and can be used to record multiple subcomposite points, a line (transect), or a polygon area to represent the extent of a spatial composite.[5]

The EDD specifications themselves do not require that the **sample composite** table (Table 6) be used, but other documents may impose that requirement. Similarly, the “composite_count” and “composite_period” values are not required by the EDD specification, but may be required by some other controlling document (e.g., the Sampling and Analysis Plan [SAP] or the Quality Assurance Project Plan [QAPP] for the sampling program).

3.2 COLLECTIONS AND SAMPLES FOR SPECIFIC DATA TYPES

For many types of material and sampling methods, a collection corresponds to a single gear deployment. This is not a firmly fixed relationship however, because sometimes multiple gear deployments may generate the material for a single collection. The following specific examples illustrate some of the relationships that may occur between gear deployments, collections, and samples. Note that gear deployments are not explicitly represented in the PHIDB, but they provide some context for the interpretation of collections and samples.

The collection method recorded in the **sample table** (Table 5) should identify the gear and/or sampling method used, not the material sampled.

Sediment Cores and Soil Borings

The entire core (or boring) should be considered a collection, and there should be multiple samples per collection. The collection upper and lower depths should represent the top and bottom of the core (or boring), respectively. The collection date should be the time that the core was retrieved. The vertical reference point should be the sediment or soil surface. The collection part should indicate that each sample is a vertical section. Upper and lower sample depths should be the depths of each interpretive sample below the vertical reference point.

[5] Well-known text is a text markup language for representing vector geometry objects.

Sediment or Soil Surface Grabs, Surface Water Samples, and Groundwater Samples

Ordinarily each grab sample should be both a collection and a sample. Upper and lower collection and sample depths should be the same. The vertical reference point for sediment and soil surface grabs should be the sediment or soil surface, respectively. The vertical reference point for water samples may be either the water surface or the sediment surface. In some cases multiple deployments of a grab sampler may be required to collect enough material to create a collection and sample. In those cases the collection upper and lower depths should be representative of the entire set of gear deployments.

Sediment Box Corers

The material collected in a single box corer deployment should be considered to be a single collection if subsamples from within the box core will not be vertically sectioned. In that case the subsamples will represent multiple samples from the same collection. If subsamples (sub-cores) will be vertically sectioned, then each sub-core should be considered to be a collection, and each vertical horizon from a sub-core should be considered to be a separate sample.

Multi Corers

Each core collected with a multi-corer should be considered to be a separate collection, and each vertical horizon from any of the cores should be considered to be a separate sample.

Sediment Traps

For sediment traps that collect a single quantity of material during each deployment (e.g., the Aquatec[6] single-bottle trap), all of the material collected from each deployment should be considered to be both a collection and a sample, similar to a grab sample.

For sediment traps that simultaneously collect several amounts of material during a single deployment (e.g., the OceanInstruments ST-20 sediment trap), all of the material collected during a single deployment of the instrument should be considered to be a collection, and the material from each individual trap should be considered to be a separate sample.

For sediment traps that sequentially collect a series of samples (e.g., the Hydrobios multi sediment trap), each sequentially-collected sample should be considered to be both a collection and a sample.

The sediment trap gear should be specified as part of the collection method. The collection date for sediment traps should be the date at which the sampler is retrieved, and the collection period should be provided in the "composite_period" column and should represent the duration of the deployment.

[6] Use of trade names does not constitute an endorsement of any manufacturer or particular equipment.

The vertical reference point for sediment trap samples may be either the water surface or the sediment surface. Collection and sample depths are required.

Porewater

A porewater collection is considered to be all of the material collected using a single method from the time of gear deployment to retrieval. For most collection methods, including pushpoint samplers, peepers, and semipermeable membranes, each collection will correspond to a single sample. Polydimethylsiloxane-coated fiber (PDMS) samplers may be sectioned after retrieval, and in such cases the entire fiber should be considered to be the collection and sections considered to be individual samples. The porewater collection period should be provided in the "composite_period" column and should represent the duration of the deployment.

3.3 DETECTION LIMITS

PHIDB contains method detection limit (MDL), quantitation limit (QL), instrument detection limit (IDL), and reporting limit (RL) columns.

The RL may be the MDL, QL or IDL, all of which would be supplied by the analytical laboratory, or another value supplied by the data validator. If the RL does not match the MDL, QL or IDL then the data submitter needs to populate the data validator comments to identify that the value was identified based on the data validator's quality assurance/quality control review of the data.

4.0 EDD SPECIFICATIONS

The EDD Specifications consist of a Data Dictionary, Valid Values, and Metadata. Each of these elements is described below.

4.1 DATA DICTIONARY

The PHIDB data dictionary is a set of informational tables describing the contents, format, and structure of the database and the relationship between the elements of the database. The PHIDB data dictionary consists of up to 13 data tables, not all of which may be needed for every study. The tables are listed below, and Table 1 provides a brief description of each table.

- **field_event table** (Table 2)
- **location table** (Table 3)
- **well table** (Table 4)
- **sample table** (Table 5)
- **sample_composite table** (Table 6)
- **collection_measurement table** (Table 7)
- **sample_measurement table** (Table 8)
- **collection_observation table** (Table 9)
- **sample_observation table** (Table 10)
- **lab_result table** (Table 11)
- **tox_test_batch table** (Table 12)
- **tox_test_result table** (Table 13)
- **species_abundance table** (Table 14)

Descriptions of each EDD table (i.e., the data dictionary) are presented in the attached Tables 2 through 14. The EDD table descriptions specify the columns that are to be included in each table, the type of information that each column represents, the data type of each column, and constraints on each column. Constraints include: (a) whether or not a value is required in the column; (b) whether or not each column makes up the primary key of a table, where the primary key must be unique across all rows; and (c) whether or not a column must be filled with one of a set of valid values.

Ordinarily, only four tables need to be populated for an environmental chemistry study, which are: the **field event table** (Table 2), **location table** (Table 3), **sample table** (Table 5), and **lab result table** (Table 11). An Entity Relationship Diagram (ERD) of these four EDD tables is shown on Figure 2 EDD Tables—Entity Relationship Diagram. Other tables allow additional information to be provided on field measurements, sample compositing, toxicity test data, and species abundance data.

4.2 VALID VALUES

Lists of valid values are included as Appendix A Valid Values. More detailed notes are provided for some entries to provide additional guidelines for populating the data columns. Data submitters are encouraged to translate related values when possible. For example, analyte names can be translated by matching Chemical Abstracts Service (CAS) numbers.

Please choose from the following two options if there is no corresponding or equivalent valid value.

1. Submitters may prepare new valid values in the same structure as Appendix A tables. Please provide new valid values tables prior to, or with, the data submittal. If that is the case, the submitter may propose a code and definition for the new valid value, ensuring that its meaning does not conflict or overlap with any established value.

2. Email the Data Coordinator for valid value suggestions and you will receive a response normally within one business day.

4.3 METADATA

A table of metadata should accompany each data submission. The metadata table should have two columns, titled “Metadata item” and “Description”; a template is provided in Table 15—Metadata items for each submission. The table should contain all of the following items:

- Submittal title: A name that uniquely identifies the data set that is being submitted. This may correspond to the name of a document such as a Work Plan or to an Administrative Settlement Agreement and Order on Consent (ASAOC) title. The submittal title should remain the same if subsequent versions of the data set are prepared and submitted.
- Submittal version: An integer that should be equal to 1 for the first submittal and that is incremented by 1 for each re-submittal of the same data set (if any).
- Data set date: The date and time, in ISO-8601 base or extended format, at which data were accessed or extracted for preparation of the EDD tables.
- Preparation date: The date and time, in ISO-8601 base or extended format, at which preparation of the EDD tables was completed. This must not be earlier than the data set date.
- Preparer: The name of the person who prepared the EDD tables, or to whom questions can be addressed.
- Preparer organization: The preparer’s employer or other affiliation.

- Preparer email: The email address of the preparer.
- Preparer phone: A phone number for the preparer.
- Version revisions: Text describing changes to the data set or the EDD tables since the prior version of the same submittal, if any. This is required if the value of the *Submittal version* item is greater than 1.
- PHDP document: The title of a document on the Portland Harbor Environmental Data Portal that most accurately describes the scope or content of the submitted data set. This item is optional and need not be included if no relevant document exists.
- Comments: Any other description of the data set or the EDD that the submitter considers to be pertinent and valuable. This item is optional.

The metadata should be submitted in a digital form with the EDD tables in a table named "Metadata."

5.0 DATA SUBMITTAL AND REVIEW

EDD tables should be prepared using the table and column names exactly as given in Tables 2 through 14. Valid values to be used in the EDD tables are listed in Appendix A. Please check the portal before each data submittal for the current version of the valid values. First time submitters are encouraged to contact the PHIDB Data Manager prior to initiating data upload to review the submittal process; the Data Manager can be reached at the following address: data-submittal@phssidb.com.

EDD tables can be prepared and submitted in any of the following formats:

- Microsoft Access database file. A template Access file is available at the Portland Harbor Environmental Data Portal. However, Access is limited in its ability to represent PHIDB relationships. Please contact the PHIDB Data Manager if issues are encountered with the Access database template.
- Microsoft Excel workbook. Each worksheet should represent a single EDD table, and the worksheet name should match the corresponding EDD table name. The first row in each worksheet should contain column names, and all succeeding rows should contain data as specified in Tables 2 through 14. Comments and formatting should not be used to explicitly or implicitly convey any additional information.
- LibreOffice Calc workbook. This has the same requirements as an Excel workbook.
- CSV files. Each CSV file should represent a single EDD table, and the filename should match the corresponding EDD table name. The first row in each file should contain the column names, and all following rows should contain data as specified in Tables 2 through 14.

File names should all be prefixed with the date of submission, in YYYY-MM-DD format. If multiple versions are submitted on the same date, version indicators should be added (e.g., v1).

After each data set is submitted, it will be subjected to a series of checks to verify conformance with the EDD structure and other constraints. A list of the current set of initial QA checks[7] is provided in Appendix B Quality Assurance Checks. These checks will include, but not necessarily be limited to:

- Submission of a complete set of tables;
- Use of defined or proposed valid values;
- Valid primary keys for each table;

[7] QA checks may be expanded as warranted.

- All required data values are provided based on the "Required" column in Tables 2 through 14;
- Valid relationships between tables (e.g., all main sample identifiers that are referenced in the **lab result table** [Table 11] are present in the **sample table** [Table 5]);
- Consistent identifiers within and between EDD tables (e.g., all sample identifiers contain consistent use of lower or upper case);
- Internal consistency checks (e.g., no collection has different depths on different lines of the **sample table** [Table 5]); and
- Conformance checks: requirements of the relevant ASAOC are met (e.g., with respect to sampling locations and analytes).

If these checks result in the identification of problems or ambiguities, a report of the issues will be provided to the submitter, and resubmittal of the data will be requested. Undefined valid values are the most frequent problem identified that requires a resubmittal.

After submitted data have passed these checks, the data will be loaded into a copy of PHIDB. Additional integrity and relational checks occur when the data are loaded into PHIDB that may identify issues beyond the QA checks. Once the data are loaded to a copy of PHIDB it is reviewed by EPA or its contractors. This phase of review will focus on technical content and compliance with any relevant order and approved SAP and QAPP documents. If deficiencies are found, the submitter will be asked to correct them by providing additional information or by correcting and re-submitting the data set. After completion of this technical review, the data will be loaded into the main PHIDB database and will be available for online users.

TABLES

Table 1. EDD Table Summary

Table name	Table No.	Purpose	Required
field_event	2	Describes the study (field event) that was carried out. This table should contain only a single row for each data submission.	Always
location	3	Describes target locations for sampling, including coordinates and event-specific location identifiers.	Always
well	4	Describes wells that are sampled for groundwater.	
sample	5	Describes the environmental samples collected in terms of both collections and main samples, as described in the text of the EDD specification document.	Always
sample_composite	6	Describes how individual environmental samples were composited.	As determined by SOW or work plan.
collection_measurement	7	Contains quantitative field measurements that were made on a collection or at the time that a collection was acquired.	Only when field measurements are made that are specific to a collection.
sample_measurement	8	Contains quantitative field measurements that were made on a sample--i.e., on a portion of a collection. If collections and samples are equivalent (e.g., for surface grab samples), then all field measurements should be recorded in Table 7 *collection_measurement* table.	Only when field measurements are made that are specific to a sample.
collection_observation	9	Contains categorical field observations that were made on a collection or at the time that a collection was acquired.	
sample_observation	10	Contains categorical field observations that were made on a sample. If collection and samples are equivalent, then all field observations should be recorded in Table 9 *collection_observation* table.	
lab_result	11	Contains analytical chemistry results for environmental samples. This includes both laboratory-reported results and calculated results such as sums. Results are reported only for environmental samples, including field splits (duplicates), but not any other field or laboratory quality control samples. Detail down to the level of laboratory replicate results should be included.	Only when analytical chemistry measurements have been made.
tox_test_batch	12	Describes the type(s) of toxicity tests run on environmental samples, and identifies each batch of tests that were conducted simultaneously.	Only when toxicity tests are conducted.
tox_test_result	13	Contains the results of toxicity tests on environmental samples.	Only when toxicity tests are conducted.
species_abundance	14	Contains measurements of species abundance (e.g., benthic infauna enumeration) made on each environmental sample.	Only when species abundance measurements have been made.

Table 2. The field_event table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
field_event	event_description	Description of the event, for example: Lower Burke Canyon Repository Predesign Investigation.	Text	255	Yes	No	Yes	
	field_activity	The type of activity to be conducted (e.g., remedial design sampling, confirmation sampling, post-closure monitoring).	Text	35	Yes	Yes	No	l_field_activity
	asaoc_id	Identifier for the ASAOC that this field event is conducted to support.	Text	50	Yes	No	No	
	event_status	Event completion status (e.g., pending, underway, completed).	Text	24	No	Yes	No	l_event_status
	spatial_extent	A polygon or multipolygon representation of the spatial extent of the event, in well-known text (WKT) format.	Text	Unlimited	No	No	No	
	sponsor	Name of the organization on whose behalf the field event is undertaken.	Text	200	No	No	No	
	sponsor_contact	Contact information for the field event sponsor.	Text	255	No	No	No	
	contractor	Full name of the company contracted by the sponsor to conduct or manage the sampling.	Text	200	Yes	No	No	
	qapp_approved	Indicates if the QAPP has been approved for the sampling effort.	Boolean	0	Yes	No	No	
	qapp_approving_org	Identifies the agency or other organization that approved the QAPP for the sampling effort.	Text	100	No	No	No	
	qapp_approved_other	Other information about approval of the QAPP.	Text	150	No	No	No	

Table 3. The location table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
location	event_location_id	Investigation-specific location identifier (Note 1). This is the target location identifier.	Text	50	Yes	No	**Yes**	
	location_description	Narrative description of the location.	Text	255	Yes	No	No	
	other_location_id	The universal PHIDB location identifier, if known (Note 1).	Text	50	No	No	No	
	target_latitude_dd	The target latitude for this sampling location, in decimal degrees.	Floating-point		Yes	No	Note 2	
	target_longitude_dd	The target longitude for this sampling location, in decimal degrees, with at least five decimal digits of precision.	Floating-point		Yes	No	Note 2	
	target_srid	The spatial reference ID for the coordinates. This should be 4326, representing WGS84.	Text	8	Yes	No	No	
	country_code	Country code. This should be "USA."	Text	75	No	Yes	No	l_country_code
	county_fips_code	County code.	Text	150	No	Yes	No	l_fips_code
	elev_datum	Datum used to determine the elevation measurement (e.g., NAVD88; NGVD29).	Text	50	Note 3	Yes	No	l_elev_datum
	elev_method	Method used to determine the elevation measurement (e.g., Altimetry, GPS, Interpolation, Survey).	Text	24	No	Yes	No	l_elev_method
	geo_method	Geopositioning method used to establish latitude and longitude coordinates (e.g., GPS, Interpolation, Survey).	Text	24	Yes	Yes	No	l_position_method
	huc_eight_digit_code	Eight digit USGS HUC code. equals Sub_Basin.	Text	20	No	No	No	
	location_zone	Location categorization (e.g., subtidal, intertidal, upland, facility).	Text	24	No	Yes	No	l_zone_type
	river_mile	River mile, to at least tenths of a mile.	Floating-point	0	No	No	No	
	river_bank	Code to specify whether the location is near a river bank or in the channel.	Text	24	No	Yes	No	l_riverbank
	state_code	State code—2 character state abbreviation.	Text	2	No	Yes	No	l_state
	stream_or_creek_name	Name of the water body represented by this location.	Text	200	No	No	No	
	surf_elev	The ground elevation of a geographic point where samples or field measurements are collected.	Floating-point	0	No	No	No	
	surf_units	Surface elevation units (e.g., feet, meters).	Text	20	Note 3	Yes	No	l_unit
	usgs_station_id	USGS location identifier.	Text	100	No	No	No	
	usgs_station_name	USGS location name.	Text	255	No	No	No	

Notes

1 Each investigation may assign a new identifier to a previously sampled location that has a different "universal" location identifier.
2 The combination of target latitude and longitude must be unique on every row.
3 Required if a value is provided for *surf_elev* .

Table 4. The well table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
well	well_id	Well identifier.	Text	50	Yes	No	Yes	
	event_location_id	Investigation-specific location identifier.	Text	50	Yes	No	No	location.event_location_id
	vertical_reference_point	The surface from which depths are measured (e.g., top of casing, well benchmark).	Text	24	Yes	Yes	No	l_vert_ref_pt
	well_hole_depth_measure	The maximum depth of the well.	Floating point		No	No	No	
	well_hole_depth_measure_unit	The units for the well depth and screen depth measurements.	Text	20	Yes	Yes	No	l_unit
	completion_date	The date on which the well was completed.	Date/time					
	screen_upper_depth	The upper depth of the screened interval, or the upper depth of the shallowest screened interval.	Floating point		No	No	No	
	screen_lower_depth	The lower depth of the screened interval, or the lower depth of the deepest screened interval.	Floating point		No	No	No	
	well_use	The primary usage of the well.	Text	24	No	Yes	No	l_well_use
	well_status	The current status of the well.	Text	24	No	Yes	No	l_well_status
	well_status_date	The date on which the well status became applicable.	Date/time		Note 1	No	No	

Table 5. The sample table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
sample	main_sample_id	Unique sample identifier for this interpretive sample.	Text	50	Yes	No	**Yes**	
	collection_id	Unique collection identifier for each core, grab, or group of related field samples.	Text	50	Note 1	No	No	
	event_location_id	Investigation-specific location identifier.	Text	50	Yes	No	No	location table
	sub_location	Narrative description of any systematic deviation or difference from the nominal location	Text	255	No	No	No	
	collection_date	Date and time of acquisition of the collection material, in ISO-8601 base or extended format. This is often the GPS time recorded. For temporal composites, use the collection time of the final sub-composite unless specified otherwise by the sampling plan.	DateTime	0	Yes	No	No	
	sample_date	Date and time of creation of the sample for a composite or single sample, in ISO-8601 base or extended format. The sample date may be the same as the collection date if the collection is not subdivided into multiple samples.	DateTime	0	Yes	No	No	
	event_element	Identifier for a subset of the sampling effort (e.g., Phase 1, Phase 2, background sampling, random grid sampling).	Text	50	No	No	No	
	collection_method	Sample collection method (e.g., grab, core, sediment trap, low purge, box corer, peeper, etc.).	Text	24	Yes	Yes	No	l_coll_method
	collection_design	Sample collection design or scheme (e.g., single-point grab, spatial composite, temporal composite, MIS).	Text	24	Yes	Yes	No	l_coll_design
	collection_depth_upper	Upper depth of the collection relative to the vertical reference point.	Floating-point	0	Note 2	No	No	
	collection_depth_lower	Lower depth of the collection relative to the vertical reference point.	Floating-point	0	Note 2	No	No	
	collection_depth_units	Units for the upper and lower collection depths.	Text	20	Note 2	No	No	l_unit
	vertical_reference_point	The surface from which elevations or depths are measured (e.g., water surface, sediment surface, soil surface, well benchmark).	Text	50	Yes	Yes	No	l_vert_ref_pt
	vert_ref_pt_elev	The elevation of the surface reference point in the specified elevation datum.	Floating-point		No	No	No	
	vert_ref_pt_elev_units	Units for the vertical reference point.	Text	10	Note 3	Yes	No	l_unit
	elev_datum	The vertical datum for the vertical reference point elevation measurement. This should be either NAVD88 or NGVD29.	Text	50	Note 3			
	composite_type	The compositing method used for the collection (e.g., single, spatial, depth, temporal).	Text	24	Yes	Yes	No	l_composite_type
	composite_count	The number of other collections that were composited to create this collection. For example, a triplicate multipoint has a value of 3. When individually identified samples are composited, those samples can be listed in the *sample_composite* table.	Integer	0	No	No	No	
	composite_period	The length of time over which a temporal composite was conducted.	Floating-point	0	No	No	No	
	composite_period_units	The units for the composite period, for temporal composites.	Text	20	Note 4	Yes	No	l_unit
	sample_lat_centroid_dd	The latitude of the sampling point or of the centroid of the sampling line or area, in decimal degrees.	Floating-point	0	Yes	Note 5	No	
	sample_lon_centroid_dd	The longitude of the sampling point or of the centroid of the sampling line or area, in decimal degrees.	Floating-point	0	Yes	Note 5	No	
	sample_srid	The spatial reference ID for the sample centroid coordinates and for the collection coordinates. This should be 4326, representing WGS84.	Text	8	Yes	Note 5	No	
	coll_coords	Coordinates for the collection in well-known-text (WKT) format. This may represent one or more points, lines, or areas. If a value is provided, the centroid must match the provided sample centroid values.	Text	Unlimited	No	Note 6	No	

Please refer to notes on last page of table.

Table 5. The sample table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
sample	geo_method	Geopositioning method used to establish coordinates for the collection.	Text	30	Yes	Yes	No	l_position_method
	horiz_accuracy_measure	Horizontal accuracy measurement—the radius of a circle around the measured point within which the true location occurs with a 95% probability.	Floating-point	0	No	No	No	
	horiz_accuracy_measure_units	Horizontal accuracy measurement units.	Text	20	Note 7	Yes	No	l_unit
	collection_material	The type of material collected.	Text	24	Note 8	Yes	No	l_sample_material
	sample_material	The type of material sampled—this may differ from the collection material if any fractionation of the material has been performed in the field.	Text	24	Yes	Yes	No	l_sample_material
	collection_part	The fraction of the collection represented by this sample (e.g., entire, vertical horizon, filtered fraction).	Text	24	Yes	Yes	No	l_subsample_type
	sample_depth_upper	Sample upper depth relative to the vertical reference point.	Floating-point	0	Note 9	No	No	
	sample_depth_lower	Sample lower depth relative to the vertical reference point.	Floating-point	0	Note 9	No	No	
	sample_depth_units	Sample depth units.	Text	20	Note 10	Yes	No	l_unit
	taxon	Taxon code for organisms.	Text	24	No	Yes	No	l_taxon
	field_prep_method	Sample preparation, fractionation, or treatment method carried out at the time of collection (e.g., filtering).	Text	24	No	Yes	No	l_prep_method
	sample_mass	The mass of this sample.	Floating-point	0	No	No	No	
	sample_mass_units	The units for the sample mass.	Text	20	Note 11	Yes	No	l_unit
	sample_mass_basis	"Wet" or "Dry" for sample masses.	Text	10	No	Yes	No	l_meas_basis
	sample_volume	The volume of this sample.	Floating-point	0	No	No	No	
	sample_volume_units	The units for the sample volume.	Text	20	Note 12	Yes	No	l_unit
	sample_color	The color of the sample upon collection.	Text	24	No	Yes	No	l_color
	sample_odor	The odor of the sample upon collection.	Text	24	No	Yes	No	l_odor
	sampling_reasons	Description of the purpose for collection of this sample.	Text	255	No	Yes	No	l_sample_reason—Note 13
	sampler_org	Name of organization that collected the sample.	Text	100	Yes	No	No	
	sampler_person	Name of the person who collected the sample.	Text	64	No	No	No	
	remarks	Comments on the sample.	Text	255	No	No	No	

Notes

1 The collection ID must be provided if it is different from the main sample ID. If it is not provided, it will be set identical to the main sample ID during import of the EDD.

2 Collection depths are required if the collection and the sample are not identical, and the material collected is sediment, surface water, soil, or groundwater. If the collection and the sample are identical, collection depths should be omitted or equal to the sample depths. If depths are required, both depths must be provided.

3 Required if *vert_ref_pt_elev* is provided.

4 Required if *composite_period* is provided.

5 Sample-specific actual coordinates are required. If there are multiple coordinates for a sample (as for a spatial composite), then those may be provided as WKT in the *coll_coords* column. However, the required centroid of the locations must be provided in the *sample_lat_centroid_dd* and s*ample_lon_centroid_dd* columns.

6 See https://www.ogc.org/standards/wkt-crs

7 Required if *horiz_accuracy_measure* is provided.

8 Required if the collection material is different than the sample material.

9 Required if the sample material is sediment, surface water, soil, or groundwater. Both depths must be provided.

10 Required if *sample_depth_upper* and *sample_depth_lower* are provided.

11 Required if *sample_mass* is provided.

12 Required if *sample_volume* is provided.

13 The entry must be a comma-separated list of valid values.

Table 6. The sample_composite table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
sample_composite	main_sample_id	Unique sample identifier for the interpretive sample that is a composite.	Text	50	Yes	No	**Yes**	sample table
	subcomposite_sample_id	Unique sample identifier for one of the interpretive samples that makes up the composite. A separate row should be added for each subcomposite sample.	Text	50	Yes	No	**Yes**	sample table
	subcomposite_mass	The mass of this subcomposite sample used in the composite.	Floating-point	0	No	No	No	
	subcomposite_mass_units	The units for the subcomposite mass.	Text	20	Note 1	Yes	No	l_unit
	subcomposite_volume	The volume of this subcomposite used in the composite.	Floating-point	0	No	No	No	
	subcomposite_volume_units	The units for the subcomposite volume.	Text	20	Note 2	Yes	No	l_unit

Notes

1 Required if *subcomposite_mass* is provided.

2 Required if *subcomposite_volume* is provided.

Table 7. The collection_measurement table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
Collection_Measurement	collection_id	Unique collection identifier for each core, grab, or group of related field samples.	Text	50	Yes	No	**Yes**	sample.collection_id Note 1
	collection_measurement	The type of measurement made (e.g., Salinity, pH, Water Temperature).	Text	24	Yes	Yes	**Yes**	l_measurement
	measurement_method	The method used to make the sample measurement.	Text	24	Yes	Yes	**Yes**	l_meas_method
	replicate	Replicate identifier to distinguish multiple measurements (e.g., "1", "2").	Text	10	Yes	No	**Yes**	
	measured_value	The numerical result of the measurement.	Floating-point	0	Yes	No	No	
	collection_measurement_units	Result unit of measurement.	Text	20	Yes	Yes	No	l_unit
	qa_level	Level of data quality review used.	Text	24	Yes	Yes	No	l_qa_level
	reportable	Flag to distinguish reportable and non-reportable results based on data quality review.	Boolean	0	Yes	No	No	
	collection_measurement_comments	Comments on the measured value.	Text	255	No	No	No	

Note

1 The sample table (Table 5) does not have a unique key on the *collection_id* column; however, the values used here must appear in that column.

Table 8. The sample_measurement table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
sample_measurement	main_sample_id	Unique sample identifier for this interpretive sample.	Text	50	Yes	No	**Yes**	d_sample
	sample_measurement	The type of measurement made (e.g., mass, volume, temperature).	Text	24	Yes	Yes	**Yes**	l_measurement
	measurement_method	The method used to make the sample measurement.	Text	24	Yes	Yes	**Yes**	l_meas_method
	replicate	Replicate identifier to distinguish multiple measurements (e.g., "1", "2").	Text	10	Yes	No	**Yes**	
	measured_value	The numerical result of the measurement.	Floating-point	0	Yes	No	No	
	sample_measurement_units	Result unit of measurement.	Text	20	Yes	Yes	No	l_unit
	qa_level	Level of data quality review used.	Text	24	Yes	Yes	No	l_qa_level
	reportable	Flag to distinguish reportable and non-reportable results based on data quality review.	Boolean	0	Yes	No	No	
	sample-measurement_comments	Comments on the measured value.	Text	255	No	No	No	

Table 9. The collection_observation table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
Collection_Observation	collection_id	Unique collection identifier for each core, grab, or group of related field samples.	Text	50	Yes	No	Yes	sample.collection_id
	observation_type	The class of observation made (e.g., Tide Stage) using the *observation_type* column.	Text	24	Yes	Yes	Yes	l_observation
	observation	The observation made; one of the categorical values using the observation class column of the observation lookup table. For example "Ebb" for tide stage.	Text	24	Yes	Yes	Yes	l_observation
	collection_observation_comments	Comments on the observation.	Text	255	No	No	No	

Table 10. The sample_observation table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
Sample_Observation	main_sample_id	Unique sample identifier for this interpretive sample.	Text	50	Yes	No	Yes	sample.main_sample_id
	observation_type	The class of observation made (e.g., Sediment Class) using the *observation_type* column.	Text	24	Yes	Yes	Yes	l_observation
	observation	The observation made; one of the categorical values using the observation class column of the observation lookup table. For example "Silty gravel" for Sediment Class.	Text	24	Yes	Yes	Yes	l_observation
	sample_observation_comments	Comments on the observation.	Text	255	No	No	No	

Table 11. The lab_result table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
lab_result	main_sample_id	Unique sample identifier for the interpretive sample.	Text	50	Yes	No	**Yes**	sample
	analytical_sample_id	Unique sample identifier for each analytical sample.	Text	50	Note 1	No	**Yes**; Note 1	
	lab_sdg	Laboratory Sample Delivery Group (batch) ID.	Text	50	Yes, Note 2	Note 3	**Yes**	
	analysis	Lab analysis (e.g., metals, VOCs).	Text	24	Yes, Note 2	Yes	**Yes**	l_lab_analysis
	analyte	Analyte/parameter name (e.g., lead, arsenic, etc.).	Text	24	Yes	Yes	**Yes**	l_analyte
	preparation_method	Lab preparation method (e.g., extraction method).	Text	24	Yes, Note 2	Yes	**Yes**	l_prep_method
	analytical_method	Lab analytical method (e.g., 8270M).	Text	24	Yes, Note 2	Yes	**Yes**	l_anal_method
	material_analyzed	Material analyzed.	Text	24	Yes	Yes	**Yes**	l_sample_material
	fraction_analyzed	Indicator of what fraction of the sample was analyzed (e.g., total, dissolved, leachate, sieved size interval).	Text	24	Yes	Yes	**Yes**	l_fractions
	lab_replicate	Laboratory replicate identifier.	Text	20	Yes	No	**Yes**; Note 4	
	calculated	Is this a calculated result rather than a value reported by the laboratory?	Boolean	0	Yes	No	No	
	calculation_method	Standardized description of the calculation method used.	Text	24	Note 5	Yes	No	l_calc_method
	result	Result (concentration or equivalent) reported by the lab. If the analyte was not detected, the chosen reporting limit should be used.	Floating-point	0	Yes	No	No	
	sig_figs	Significant digits of the result.	Integer	0	Yes, Note 2	No	No	
	result_units	Result unit of measurement.	Text	20	Yes	Yes	No	l_unit
	basis	"Wet" for *wet_weight* basis reporting; "Dry" for *dry_weight* reporting.	Text	10	Yes	Yes	No	l_meas_basis
	lab_qualifiers	Qualifiers and flags assigned by the laboratory.	Text	16	No	Yes	No	

Please refer to notes on last page of table.

Table 11. The lab_result table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
lab_result	tic	Is this a tentatively identified compound (TIC)?	Boolean	0	Yes, Note 2	No	No	
	validated	Has this result been validated?	Boolean	0	Yes, Note 2	No	No	
	validation_level	Stage of validation: electronic and manual.	Text	24	Yes, Note 2	Yes	No	l_validation_level
	result_qualifiers	Final validated result qualifiers/flags (e.g., J;U;ND;<;>).	Text	12	Note 6	Yes	No	
	undetected	Flag to distinguish detected and undetected results, based on lab or validation qualifiers.	Boolean	0	Yes	No	No	
	estimated	Flag to distinguish estimated and non-estimated results, based on lab or validation qualifiers.	Boolean	0	Yes	No	No	
	rejected	Flag to distinguish rejected and non-rejected results, based on lab or validation qualifiers.	Boolean	0	Yes	No	No	
	reportable	Flag to distinguish reportable and non-reportable results, based on validation or data quality review.	Boolean	0	Yes, Note 2	No	No	
	limited_analytes	Flag to distinguish if the number of analytes reported is "limited" and therefore below the expected number of analytes for analyte group totals	Boolean	0	Note 7	No	No	
	validator	Validation company name.	Text	100	No	No	No	
	validator_comments	Comment on the validation assessment for this result.	Text	255	No	No	No	
	lab_result_comments	Comments on the results that do not pertain to either the validation results or data quality review results.	Text	255	No	No	No	
	qa_comments	QA comment resulting from any data quality review conducted in addition to data validation. The name(s) of the data quality reviewer(s) should be included.	Text	255	No	No	No	

Please refer to notes on last page of table.

Table 11. The lab_result table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
lab_result	date_received	Date and time that the analysis was received by the lab in ISO-8601 basic or extended format	DateTime	0	No	No	No	
	date_extracted	Date and timethat the sample was extracted by the lab in ISO-8601 basic or extended format	DateTime	0	No	No	No	
	date_analyzed	Date and time that the analysis was performed by the lab in ISO-8601 basic or extended format	DateTime	0	No	No	No	
	dilution_factor	Effective test dilution factor.	Floating-point	0	Yes, Note 2	No	No	
	lab_name	Laboratory that performed the analysis	Text	24	Yes, Note 2	Yes	No	l_lab
	lab_sample_id	Sample identifier assigned by the laboratory	Text	50	No	No	No	
	original_lab_result	Original result reported by the laboratory, if data validation resulted in restatement of the value.	Numeric	0	No	No	No	
	lab_cal_batch	The laboratory's calibration batch identifier.	Text	50	No	No	No	
	lab_qc_batch	The laboratory's quality control batch identifier.	Text	50	No	No	No	
	idl	Instrument Detection Limit (IDL) as determinded by the instrument.	Floating-point	0	No	No	No	
	idl_units	IDL units.	Text	20	Note 8	Yes	No	l_unit
	mdl	Sample-specific Method Detection Limit (MDL) as provided by the lab.	Floating-point	0	No	No	No	
	mdl_units	MDL units.	Text	20	Note 9	Yes	No	l_unit
	quantitation_limit	Quantitation limit as determined by the lab.	Floating-point	0	No	No	No	

Please refer to notes on last page of table.

Table 11. The lab_result table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
lab_result	quantitation_limit_units	Quantitation limit units.	Text	20	Note 10	Yes	No	l_unit
	reporting_limit	Reporting limit as determined by the lab or validator.	Floating-point	0	Note 11	No	No	
	reporting_limit_units	Reporting limit units.	Text	20	Note 12	Yes	No	l_unit
	value_type	WQX result value type (e.g., actual, estimated, calculated).	Text	24	No	Yes	No	l_result_value_type

Notes

1 The analytical sample ID must be provided if it is different from the main sample ID. If it is not provided, the analytical sample ID will be set identical to the main sample ID during import of the EDD.

2 Required for laboratory results but not for calculated sums.

3 If there are duplicate SDG IDs from different laboratories, prefix them with the laboratory name to eliminate these conflicts.

4 When samples are analyzed in duplicate, both results should be reported and distinguished by the laboratory replicate ID.

5 The *calculation_method* is required if the *calculated* value is set to True.

6 Final qualifiers should be reported if they have been assigned. This should include any laboratory qualifiers that are retained through the data validation process.

7 Required if the number of analytes used to calculate a total value is limited as shown in Portland Harbor Data Management Plan Table 3.

8 Required if *idl* is provided.

9 Required if *mdl* is provided.

10 Required if *quantitation_limit* is provided.

11 The reporting limit may be the MDL, QL or IDL or another value supplied by the data validator. If the RL does not match the MDL, QL or IDL then populate the data validator comments with an explanation of the value.

12 Required if *reporting_limit* is provided.

Table 12. The tox_test_batch table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
tox_test_batch	lab_name	Toxicity test laboratory.	Text	24	Yes	No	**Yes**	l_lab
	tox_test_batch	Laboratory-specific toxicity test batch identifier.	Text	50	Yes	No	**Yes**	
	tox_test_type	Type of toxicity test conducted (e.g., amphipod 10-day bioassay, echinoderm 72-hour bioassay).	Text	24	Yes	Yes	No	l_tox_test_type
	taxon	Taxon code for organism used for the toxicity test.	Text	24	Yes	Yes	No	l_taxon
	life_stage	Life stage of organisms used for the toxicity test.	Text	20	Yes	Yes	No	l_life_stage
	start_date	Starting date of the toxicity test, in ISO-8601 basic or extended format.	DateTime	0	No	No	No	
	end_date	Ending date of the toxicity test, in ISO-8601 basic or extended format.	DateTime	0	No	No	No	
	qa_level	Level of data quality review used.	Text	24	Yes	Yes	No	l_qa_level

Table 13. The tox_test_result table

Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
tox_test_result	main_sample_id	Unique sample identifier for an interpretive sample.	Text	50	Yes	No	Yes	sample
	lab_name	Toxicity test laboratory.	Text	50	Yes	No	Yes	l_lab
	tox_test_batch	Laboratory-specific toxicity test batch identifier.	Text	50	Yes	No	Yes	tox_test_batch
	tox_test_variable	The variable in which results of the test are expressed (e.g., survival, growth, reproduction).	Text	24	Yes	Yes	Yes	l_tox_test_var
	tox_test_measurement	The type of measurement made (e.g., count of survivors, percent normal, change in mass).	Text	24	Yes	Yes	Yes	l_tox_test_meas
	replicate	Replicate identifier.	Text	20	Yes	No	Yes	
	tox_test_result	The numeric result of the toxicity test.	Numeric	0	Yes	No	No	
	tox_test_units	The units for the toxicity test result.	Text	20	Yes	Yes	No	l_unit
	qa_level	Level of data quality review used.	Text	24	Yes	Yes	No	l_qa_level
	tox_test_result-comments	Comments on this toxicity test result.	Text	255	No	No	No	

Table 14. The species_abundance table

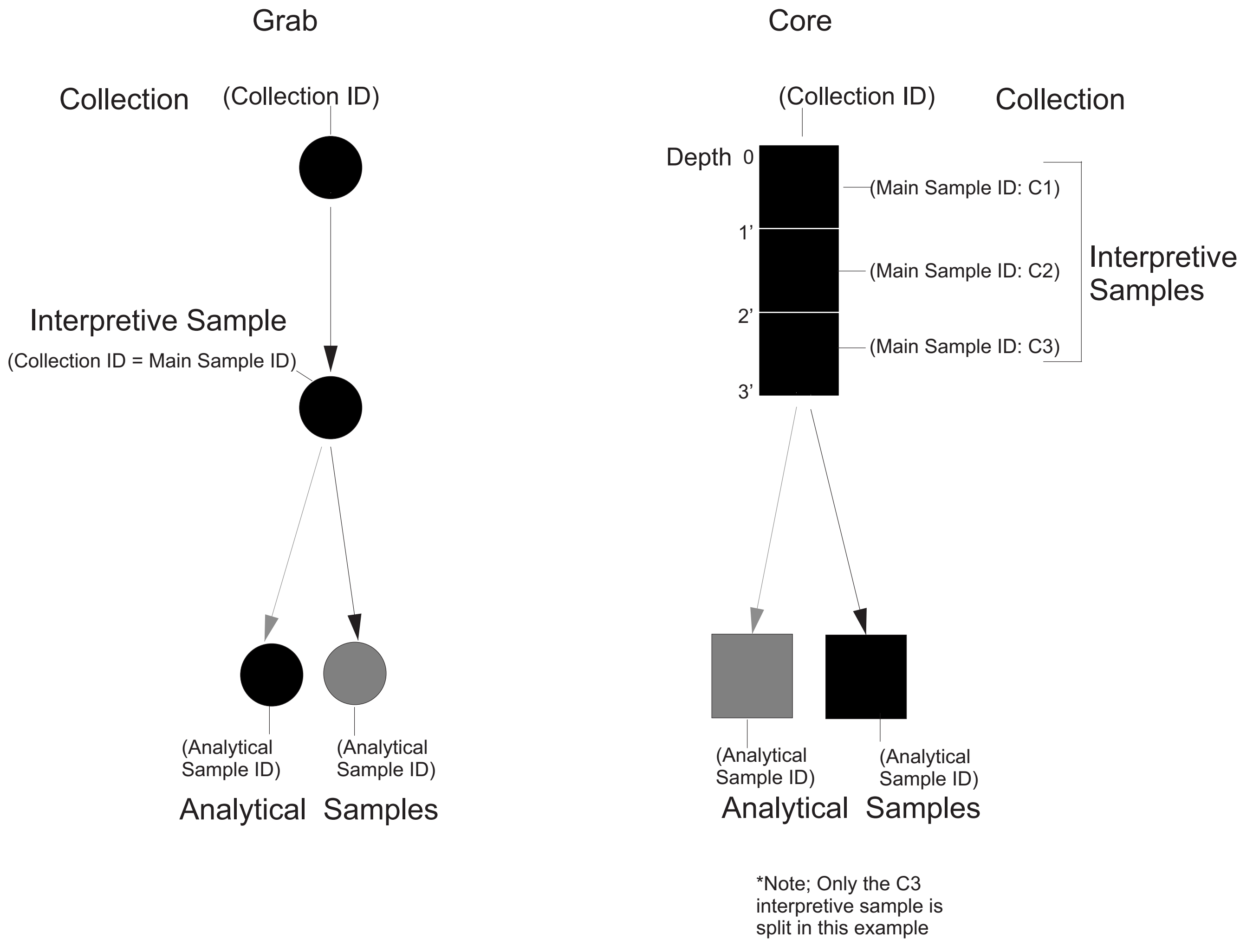
Table	Column	Description	Data Type	Size	Required	Valid Values	Primary Key	Foreign key to
species_abundance	main_sample_id	Unique sample identifier for an interpretive sample.	Text	50	Yes	No	Yes	sample
	taxon	Taxon code for the organism for which abundance was measured.	Text	24	Yes	Yes	Yes	l_taxon
	sex	Sex of organism.	Text	1	Yes	Yes	Yes	l_sex
	life_stage	Life stage of organism.	Text	24	Yes	Yes	Yes	l_life_stage
	abundance_measurement	Type of abundance measurement (e.g., count, concentration, density, spatial coverage).	Text	24	Yes	Yes	Yes	l_abund_meas
	replicate	Replicate identifier.	Text	20	Yes	No	Yes	
	abundance	Abundance measurement.	Floating-point	0	Yes	No	No	
	abundance_units	Abundance measurement units.	Text	20	Yes	Yes	No	l_unit
	lab_name	Laboratory that measured the abundance.	Text	24	No	Yes	No	l_lab
	species_abundance_comments	Comments on the species abundance measurement.	Text	255	No	No	No	

Table 15. Metadata items for each submission

Metadata item	Description
Submittal title	
Submittal version	
Data set date	
Preparation date	
Preparer	
Preparer organization	
Preparer email	
Preparer phone	
Version revisions	
PHDP document	

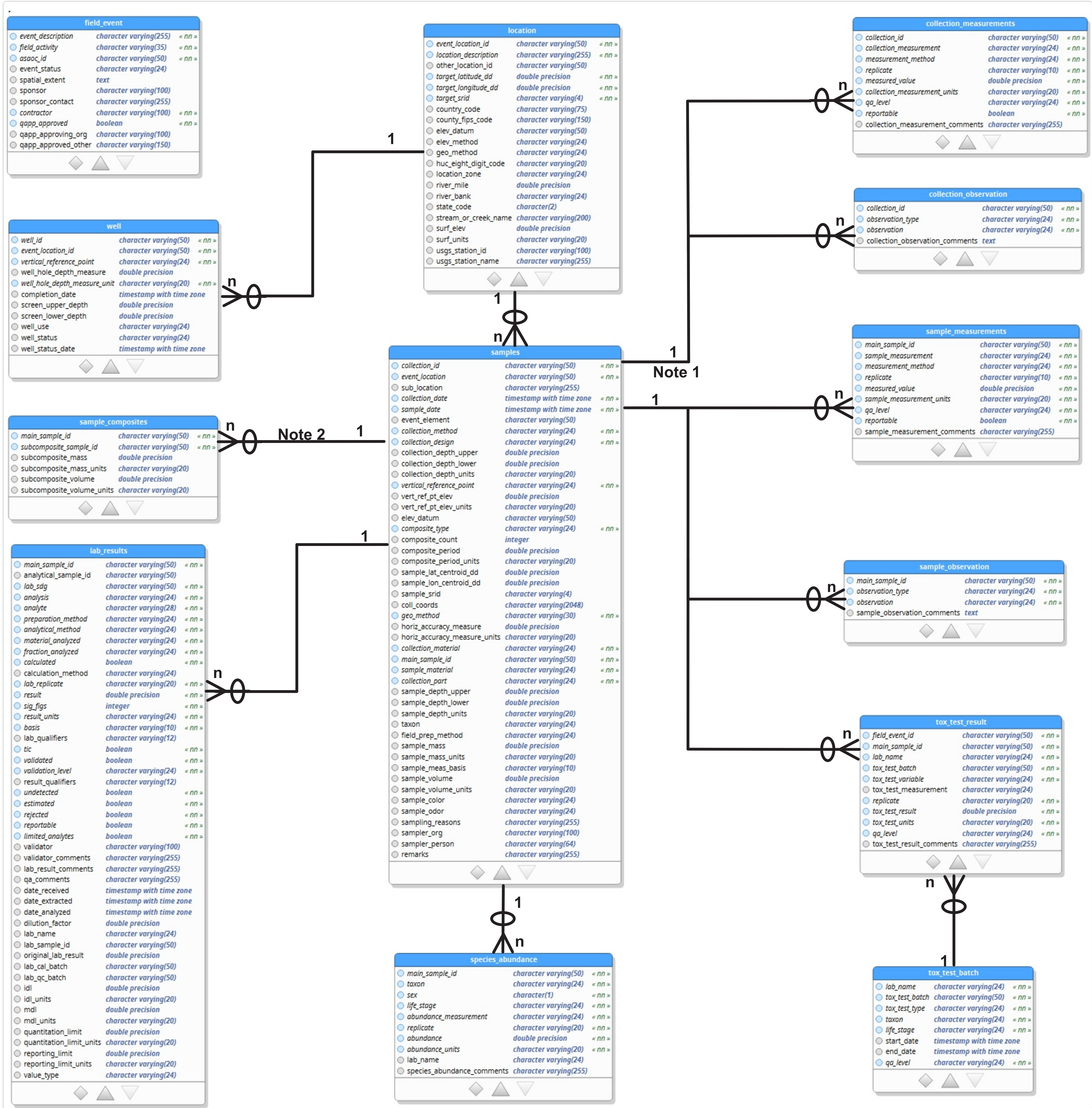
Comments

FIGURES

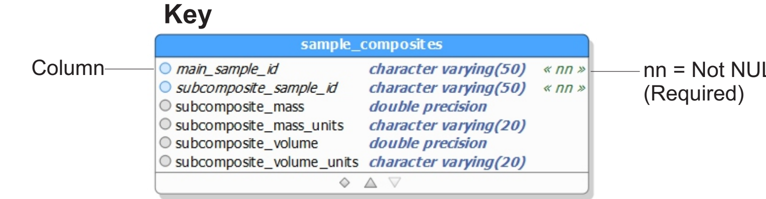


Common Structure for All Sampling Information

Figure **1**



Key



Notes
1. The collection_id is not a primary key of the sample table, but the sample table must contain all collection_id values used in the collection_measurement or collection_observation tables.
2. Both main sample IDs and subcomposite sample IDs that appear in the sample_composite table must match main_sample_id values in the sample table.

EDD Tables - Entity Relationship Diagram

Figure **2**

APPENDIX A

Valid values provided in separate Excel file on Portland Harbor Environmental Data Portal

APPENDIX B

Quality Assurance Checks

QA ID	Message	EDD Table	Severity[1]	Comments
1	Missing (null) elevation datum (elev_datum) values when the surface elevation is specified.	location	Fatal	
2	Missing (null) elev_datum or vert_ref_point values when the vertical reference point elevation is specified.	location	Fatal	
3	Missing (null) location identifiers.	location	Fatal	
4	Missing (null) location description values.	location	Fatal	
5	Missing (null) main_sample_id values.	sample	Fatal	
6	Missing (null) event_location values.	sample	Fatal	
7	Missing (null) collection_date values.	sample	Fatal	
8	Missing (null) sample_date values.	sample	Fatal	
9	Missing (null) collection_method values.	sample	Fatal	
10	Missing (null) collection_design values.	sample	Fatal	
11	Missing (null) vertical_reference_point values.	sample	Fatal	
12	Missing (null) sample_material values.	sample	Fatal	
13	Missing (null) lab_sdg identifiers.	lab_result	Fatal	
14	Missing (null) preparation_method values.	lab_result	Fatal	
15	Missing (null) analytical_method values.	lab_result	Fatal	
16	Missing (null) material_analyzed values.	lab_result	Fatal	
17	Missing (null) result (e.g., concentration) values.	lab_result	Fatal	
18	Missing (null) result_units values.	lab_result	Fatal	
19	Missing (null) basis values.	lab_result	Fatal	
20	Missing (null) tic values.	lab_result	Fatal	
21	Missing (null) validated (flag) values.	lab_result	Fatal	
22	Missing (null) validation_level values.	lab_result	Fatal	
23	Missing (null) undetected flags.	lab_result	Fatal	
24	Missing (null) estimated flags.	lab_result	Fatal	
25	Missing (null) rejected flags.	lab_result	Fatal	
26	Missing (null) reportable flags.	lab_result	Fatal	
27	Missing (null) lab_name identifiers.	lab_result	Fatal	
28	Different location identifiers with the same target coordinates. A single identifier should be used or may be created during data loading to ensure uniqueness.	location	Warning	A single location identifier is expected to be assigned to each unique pair of coordinates. Exceptions indicate a potential error or a special circumstance.
29	Collections with multiple (different) upper and lower depths or units for collections that may not be cores or borings.	sample	Fatal	Multiple samples with different upper and lower depths are expected for cores and borings but may be an issue for other types of samples.
30	Collections with multiple (different) collection dates.	sample	Fatal	If a collection is a temporal composite, the latest date should be used.
31	Collections with multiple (different) location identifiers.	sample	Fatal	If a collection is a spatial composite, the centroid or other representative location should be used.
32	Collection depths are missing (null) for a sediment, groundwater, or surface water sample.	sample	Fatal	
33	Missing (null) horiz_accuracy_measure_units where horiz_accuracy_measure is provided.	sample	Fatal	

QA ID	Message	EDD Table	Severity[1]	Comments
34	Missing (null) sample_depth_upper or sample_depth_lower for surface water, groundwater, or sediment (not including sediment traps).	sample	Fatal	
35	Missing (null) sample depths units where sample depths are provided.	sample	Fatal	
36	Missing (null) sample_mass_units where sample_mass is provided.	sample	Fatal	
37	Missing (null) sample_volume_units where sample_volume is provided.	sample	Fatal	
38	The same SDG identifier is used for different laboratories.	lab_result	Fatal	
39	Missing (null) calculation_method for calculated results.	lab_result	Fatal	
40	Missing (null) idl_units where an IDL value is provided.	lab_result	Fatal	
41	Missing (null) mdl_units where an MDL value is provided.	lab_result	Fatal	
42	Missing (null) quantitation_limit_units where a quantitation limit is provided.	lab_result	Fatal	
43	Missing (null) reporting_limit_units where a reporting limit is provided.	lab_result	Fatal	
44	Unrecognized geo_method (positioning method) value in location table.	location	Fatal	
45	Unrecognized river bank value.	location	Fatal	
46	Unrecognized country_code value.	location	Fatal	
47	Unrecognized elev datum value.	location	Fatal	
48	Unrecognized surf_units value.	location	Fatal	
49	Unrecognized elev_method value.	location	Fatal	
50	Unrecognized location_zone value.	location	Fatal	
51	Unrecognized composite_type value.	sample	Fatal	
52	Unrecognized collection_method value.	sample	Fatal	
53	Unrecognized elev_datum value.	sample	Fatal	
54	Unrecognized collection_depth_units value.	sample	Fatal	
55	Unrecognized taxon value.	sample	Fatal	
56	Unrecognized horiz_accuracy_measure_units value.	sample	Fatal	
57	Unrecognized collection_design value.	sample	Fatal	
58	Unrecognized collection_material value.	sample	Fatal	
59	Unrecognized composite_period_units value.	sample	Fatal	
60	Unrecognized vertical_reference_point value.	sample	Fatal	
61	Unrecognized geo_method (positioning method) value in samples table.	sample	Fatal	
62	Unrecognized sample_volume_units value.	sample	Fatal	
63	Unrecognized sample_mass_units value.	sample	Fatal	
64	Unrecognized sample_odor value.	sample	Fatal	
65	Unrecognized sample_depth_units value.	sample	Fatal	
66	Unrecognized field_prep_method value.	sample	Fatal	
67	Unrecognized sample_color value.	sample	Fatal	
68	Unrecognized sampling_reason value.	sample	Fatal	
69	Unrecognized sample_mass_basis value.	sample	Fatal	
70	Unrecognized idl_units value.	lab_result	Fatal	
71	Unrecognized validator value.	lab_result	Fatal	
72	Unrecognized result_units value.	lab_result	Fatal	
73	Unrecognized basis value.	lab_result	Fatal	
74	Unrecognized lab_name value.	lab_result	Fatal	

QA ID	Message	EDD Table	Severity[1]	Comments
75	Unrecognized analytical_method value.	lab_result	Fatal	
76	Unrecognized analysis value.	lab_result	Fatal	
77	Unrecognized material_analyzed value.	lab_result	Fatal	
78	Unrecognized quantitation_limit_units value.	lab_result	Fatal	
79	Unrecognized validation_level value.	lab_result	Fatal	
80	Unrecognized analyte value.	lab_result	Fatal	
81	Unrecognized mdl_units value.	lab_result	Fatal	
82	Unrecognized reporting_limit_units value.	lab_result	Fatal	
83	Unrecognized preparation_method value.	lab_result	Fatal	
84	Unrecognized value_type value.	lab_result	Fatal	
85	Unrecognized fraction_analyzed alue.	lab_result	Fatal	
86	Unrecognized calculation_method value.	lab_result	Fatal	
87	There are multiple collection_id values for a group of samples that appear to be a single collection (e.g., core).	sample	Warning	This check can only identify possible issues; those need to be evaluated individually.
88	Invalid (non-numeric) result value(s).	lab_result	Fatal	

Notes

1. Fatal errors must be corrected before the data can be loaded. Warnings indicate a potential issue that requires attention by the PHIDB Data Manager and Submitter.

ATTACHMENT C

PHIDB User Guide

THE PORTLAND HARBOR INTERIM DATABASE (PHIDB) USER GUIDE

Prepared for

State of Oregon—Department of State Lands

775 Summer Street NE # 100, Salem, Oregon 97301

Prepared by

Integral Consulting, Inc. 719 2nd Avenue, Suite 1450 Seattle, Washington 98104	**GeoEngineers, Inc.** 5820 S Kelly Avenue, Suite B Portland, Oregon 97239

December 8, 2021

CONTENTS

1 INTRODUCTION

The Portland Harbor Interim Database (PHIDB) is an internet-accessible client-server database that is designed to store environmental characterization data collected during remedial design activities at the Portland Harbor Superfund Site (PHSS). This User's Guide contains technical documentation to support PHIDB users. The intended audience is data managers of the U.S. Environmental Protection Agency (EPA), performing parties, and contractors. Users are expected to have prior experience managing environmental data and using relational databases.

1.1 Purpose of the Portland Harbor Interim Database

The PHIDB is intended to provide a centralized and standardized repository for the environmental characterization data, and related administrative data, relevant to remedial activities at the PHSS. Essential capabilities of the database, and related tools, standards, and processes, are:

- Ensure that the data collected by various parties are consistent in structure, unambiguous in representation of information, and meet basic standards for data integrity. Data integrity is supported by the use of primary keys, foreign keys, and check constraints within the database.
- Ensure that data are available to EPA and performing parties in a single system that allows integration and summarization of information across different sites or study areas within the overall PHSS.
- Ensure that a standard set of rules is applied to summarize data and calculate derived quantities such as total polychlorinated biphenyls (PCBs) and total polynuclear aromatic hydrocarbons (PAHs).
- Allow export of data to a customized version of a Scribe database (Scribe is an EPA data management application).

The PHIDB is intended to include information for making decisions about conditions and the result of actual or potential remedial actions within the PHSS. Environmental investigations typically generate copious supporting documentation in the form of plans, reports, photographs, field logs, geologic boring logs, and laboratory deliverables. The PHIDB is not designed to store these types of accessory documentation, though some of that documentation may be made available on the Portland Harbor Environmental Data Portal (http://ph-public-data.com/). The PHIDB is also not designed to store all of the types of quality control information generated by laboratories (e.g., calibration information and the results of analyses of spiked samples, blanks, and certified reference materials). The PHIDB does store the results of replicated analyses of environmental samples; although replicate analyses are used for quality control evaluations, they also provide additional information that allows more precise estimation of true environmental concentrations.

2 DATABASE MANAGEMENT SYSTEM

A complete description of the structure of the database is contained in the *Portland Harbor Interim Database (PHIDB) Design Summary* document ("Design Document"; GeoEngineers/Integral, December 16, 2021) [1]. The PHIDB database design document can be downloaded from the Portland Harbor Environmental Data Portal located here: http://ph-public-data.com/. It is recommended that PHIDB users familiarize themselves with the Design Document prior to initiating access to the database. Users should refer to the design document data dictionary, valid values, and entity diagram relationship figures to efficiently query the database. Effective querying of the PHIDB requires some familiarity with the main data tables and the relationships between them.

The PHIDB database uses the PostgreSQL database management system (Postgres; https://www.postgresql.org/). Postgres is used because:

- Postgres supports cascading updates and deletions through multiple pathways between tables, unlike other common client-server database management systems.
- Postgres allows the creation of custom data types and aggregate functions that simplify the propagation of qualifiers and significant digits while averaging and summing data.
- The PostGIS extension (http://postgis.org/) provides the most comprehensive set of features available for storage and manipulation of spatial data in a database management system.

The database is implemented on a Digital Ocean (https://www.digitalocean.com/) cloud server.

[1] GeoEngineers and Integral Consulting, 2021, *Portland Harbor Interim Database Design Summary*, prepared for State of Oregon. Version 2.0, December 16, 2021.

3 USING THE DATABASE

Using the database involves four separate steps:

1. Obtaining access credentials—a username and password
2. Selecting client software[2] to use
3. Establishing a connection from the client software to the database
4. Writing and running Structured Query Language (SQL) queries to select, summarize, and export data.

The first three of these comprise the initiation and setup steps, and the fourth of these steps is carried out throughout operational use of the database.

The following sections provide directions or guidance for carrying out these steps. The first step is well-defined, but later steps depend on each user's own preferences and goals, and there is considerable scope for variation, specifically in the choice of client software and the types of SQL queries that are used.

3.1 Obtaining User Credentials

Access credentials should be requested from the EPA Remedial Project Manager. Upon approval, a username and initial PHIDB default password, plus the IP address for the database, will be provided by a database administrator.

3.2 Selecting Client Software

Client software such as an interface or application will be needed to connect to the PHIDB server. Many different programs are available and can be used , including a graphical user interface (GUI), a command-line or text-based interface, desktop applications that can connect to and use Postgres databases (discussed in Section 3.2.1), or custom software that may be developed using programming languages such as Python. A partial list of Postgres client programs can be found at https://wiki.postgresql.org/wiki/PostgreSQL, and additional client programs can be found by searching online.

The following sections provide additional discussion of a few client programs that can be used with the PHIDB. Several GUI and command-line interfaces are described; use of the database with application programs such as R and QGIS, or within programming languages, is not

[2] Client software runs on the user's computer, whereas server software runs on the remote computer that hosts the database.

described. The descriptions of software in the following sections are intended to be useful but are not prescriptive. Users should select client software that matches their needs and skills.

3.2.1 GUI Interfaces

Client programs that use a GUI have an appearance and method of operation that is familiar to users of desktop software. Most functionality is available via menus, ribbons, or toolbar buttons. pgAdmin4 (https://www.pgadmin.org/download) is a free and open-source GUI client program that is designed specifically for Postgres. The examples proved in the User Guide are exclusively from pgAdmin. No support is provided for troubleshooting specific client software.

Other software is designed to work with multiple database management systems, including Postgres among others. Some GUI interfaces provide visual query builders, which can simplify the creation of simple queries and do not require as much knowledge of PHIDB tables and relationships. GUI interfaces with visual query builders also provide a relatively easy transition for users who have only previously used Microsoft Access. Some GUI tools that include visual query builders are:

- LibreOffice Base (https://www.libreoffice.org/discover/base/)
- DBeaver Enterprise Edition (https://dbeaver.com/)
- Navicat (https://www.navicat.com/en/)
- FlySpeed SQL Query (https://www.activedbsoft.com/overview-querytool.html)
- Quest Toad Edge (https://www.quest.com/products/toad-edge/).

3.2.2 Command-Line and Scripting Interfaces

GUIs allow easy interactive querying of the database, but executing queries using SQL scripts can provide greater efficiency for operations that will be conducted repeatedly. SQL scripts are also more amenable to quality assurance review. Two different command-line tools that can be used to run SQL scripts for the PHIDB are:

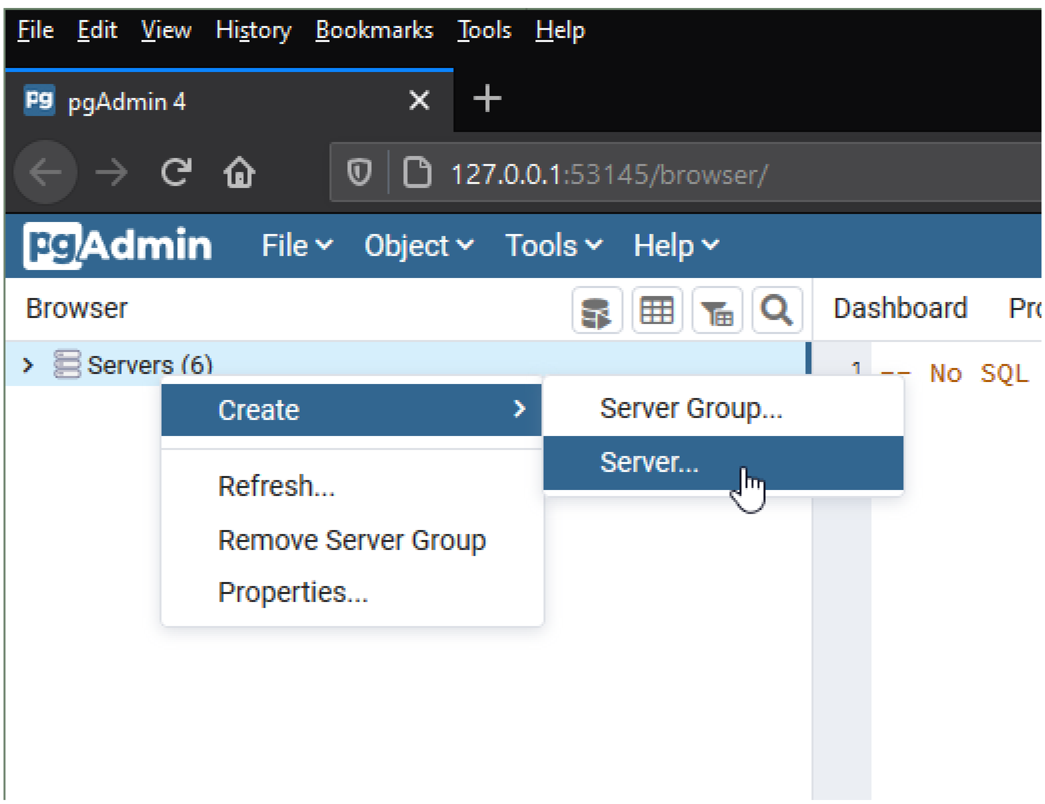
- psql. psql is a terminal-based (i.e., non-graphical) interface that can be used both interactively and to run SQL scripts. Psql is installed when pgAdmin 4 is installed and can be found in a directory path like "C:\Program Files\pgAdmin 4\v4\runtime" – the version number may differ. Documentation for psql is available at https://www.postgresql.org/docs/12/app-psql.html. To use psql to execute SQL scripts, the "-f" option must be used on the command line, followed by the name of the script file to run.

- execsql.py. execsql.py is a command-line Python program that will run a SQL script and that can export query results in several different formats. It also provides several types of conditional tests in addition to those supported by Postgres' PL/pgSQL language. Execsql.py can be installed from the Python Package Index (https://pypi.org/project/execsql/). Documentation is available at http://execsql.osdn.io/.

3.3 Connecting to the Database

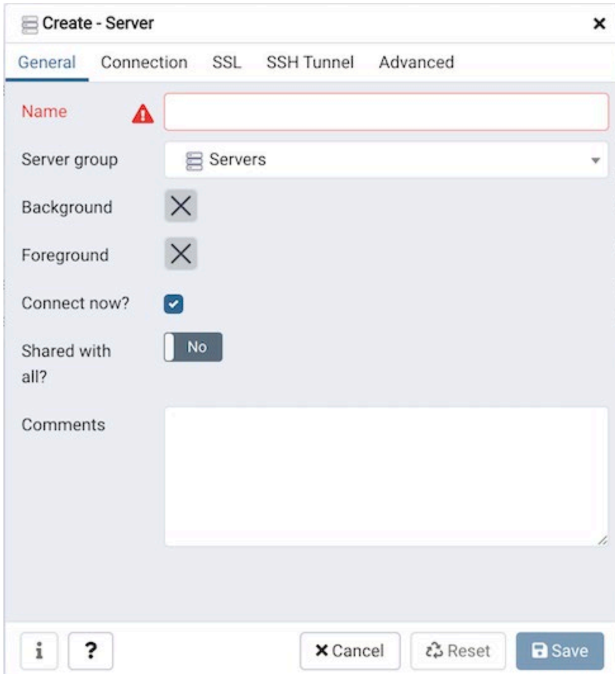
After the software has been selected to connect to the PHIDB server, the next step is to establish the database connection. The following example illustrates the creation of a database connection using the GUI pgAdmin4. The process is similar for other GUI interfaces. Users should consult the documentation for their chosen software for details.

After pgAdmin4 is installed, launch the program, then right-click on "Servers" in the left-hand pane and select the *Create/Server…* option.



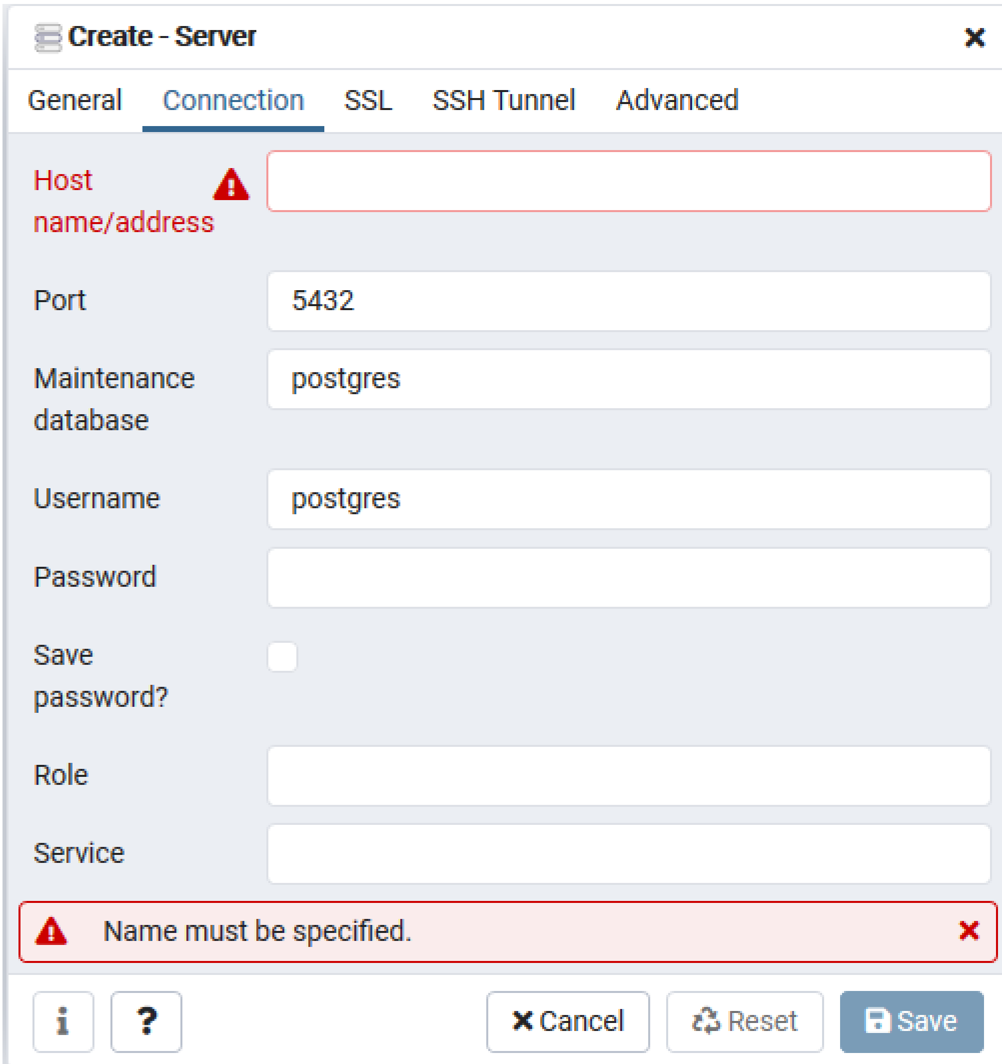
On the "General" tab of the "Create-Server" dialog, enter the following required information:

- Name: Choose a name that will distinguish this connection to the PHIDB from any other server connections that you may make in pgAdmin.



Enter the following information on the "Connection" tab

- Host name/address: Enter the IP address that was provided with your user credentials.
- Port: 5432
- Maintenance database: phidb
- Username: The username that was provided
- Password: The password that was provided.



If the "Save password?" box is checked, a copy of your password will be saved in a file on your computer; this is a potential security vulnerability and is not recommended.

Click the "Save" button to create the connection to the PHIDB.

4 SQL FOR SELECTING AND SUMMARIZING DATA

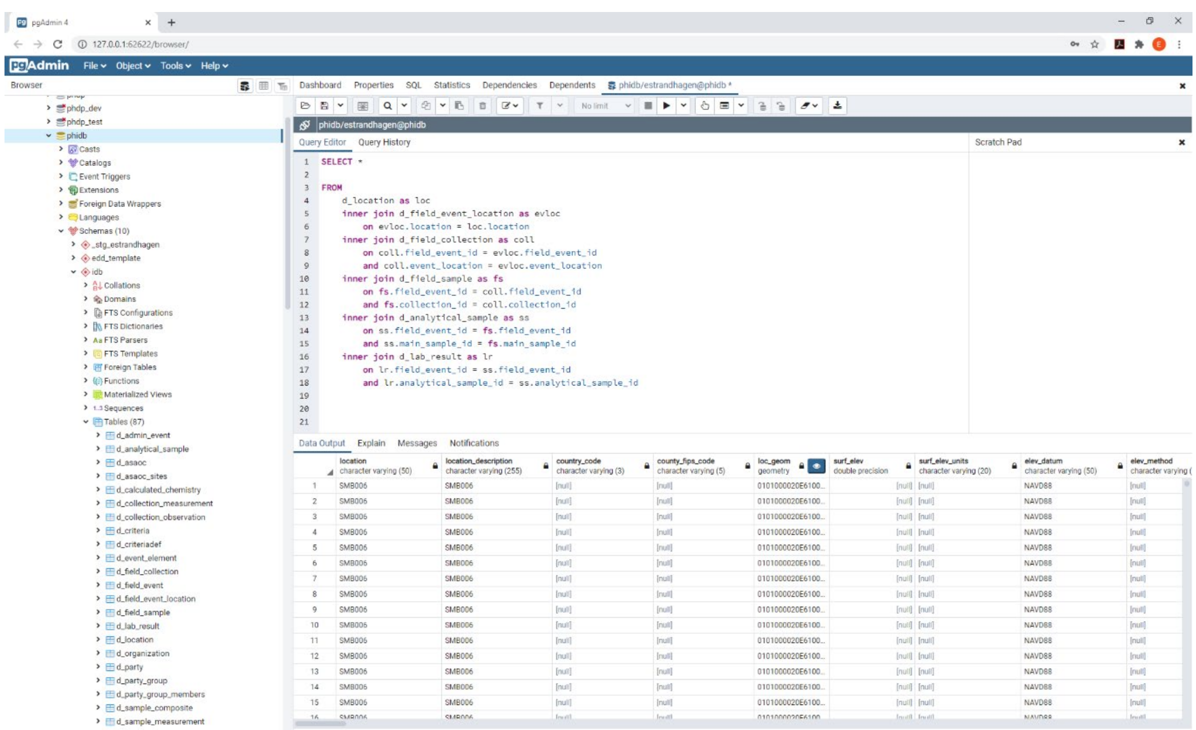
Database client software that has graphical query builders allows easy and simple querying of the database, but more complex queries require the use of SQL. This section summarizes some potentially useful applications of SQL when querying data from PHIDB. This material assumes basic familiarity with SQL; many online and print resources are available for those who are new to SQL. The SQL code presented in this User Guide is executable in the pgAdmin as described in the following section.

Running SQL Queries using the pgadmin Query Tool

The pgadmin Query Tool (shown below) executes SQL commands and displays the result in the data output window. The Query Tool is available via the *Query Tool* menu option on the *Tools* menu

The Query Tool allows you to:

- Run SQL statements provided in the User Guide
- Displays current connection and transaction status as configured by the user.
- Save the data displayed in the output panel to a CSV file.
- View analytical information about a SQL statement.



4.1 PostgreSQL SQL Syntax and Features

The SQL language supported by Postgres is highly conformant to the latest version of the ISO standard for SQL (ISO/IEC 9075:2011). Postgres also implements a number of extensions to the standard, but effective querying of the PHIDB does not require learning any Postgres-specific variations of, or additions to, standard SQL. Online documentation for Postgres is comprehensive and detailed, and can be accessed at https://www.postgresql.org/docs/.

Postgres automatically folds all schema, table, and column names to lowercase unless they are double-quoted. All PHIDB schema, table, and column names are lowercase and do not require double-quoting in queries. If uppercase or mixed-case identifier names are used in queries, and are not quoted, they will match the lowercase names. If uppercase or mixed-case identifier names are used, and they are double-quoted, they will not match identifier names in the PHIDB, and an error will result.

Character data stored in Postgres data tables is case-sensitive. When using lookup values, identifiers, or text values as selection criteria while querying the database, the query must exactly match the case of data in the database. For example, chemistry data reported on a dry-weight basis have a *meas_basis* code of "DryWt", and queries that use alternate case such as "drywt" or "Drywt" will fail to match the data. Users may refer to the design document valid values tables to match the case of the database.

Standard SQL is a declarative language, and Postgres also includes a procedural language, PL/pgSQL, that provides features such as conditional tests and looping. Documentation for PL/pgSQL is available at https://www.postgresql.org/docs/12/plpgsql-overview.html.

The *PostGIS* extension has been installed in the PHIDB to support spatial data storage and analysis. PostGIS data types are used in the PHIDB to store sampling locations and field event boundaries. Documentation for PostGIS is available at https://postgis.net.

The *tablefunc* extension has been installed in the PHIDB to support cross-tabbing of query results. Documentation for the tablefunc extension is available at https://www.postgresql.org/docs/12/tablefunc.html.

4.2 PHIDB Schemas

In Postgres, database tables and functions are grouped together into named collections called *schemas*. All PHIDB tables and functions are in a schema named "idb". The database is configured to look for tables and functions in the "idb" schema, so the schema name need not be specified when referring to base tables in the PHIDB.

4.3 Table Joins for Data Selection

When retrieving or summarizing data, selection criteria may be based on attributes of locations, samples, analytical results, and other elements of the data model that are implemented in separate database tables. SQL "join" statements should be used to link the different tables to which criteria will be applied. Because the PHIDB data model uses separate tables to distinguish between collections, main (interpretive) samples, and sample splits, the sequence of join statements may include several tables other than those to which selection criteria will be applied. Such sequences of join statements are commonly used, particularly for chemistry data. The following subsections provide examples of such sequences of join statements; these are both illustrations of the way in which tables are to be joined and also potentially useful as code snippets that can be pasted into SQL scripts.

4.3.1 Chemistry Data

The following conceptual query example includes tables from *d_location* (allowing selection by sampling coordinates or other location attributes) down to *d_lab_result* (allowing selection by analyte, qualifier, or other attributes of a chemistry measurement).

```
SELECT
     <column list>
FROM
     d_location as loc
     inner join d_field_event_location as evloc
          on evloc.location = loc.location
     inner join d_field_collection as coll
          on coll.field_event_id = evloc.field_event_id
          and coll.event_location = evloc.event_location
     inner join d_field_sample as fs
          on fs.field_event_id = coll.field_event_id
          and fs.collection_id = coll.collection_id
     inner join d_analytical_sample as ss
          on ss.field_event_id = fs.field_event_id
          and ss.main_sample_id = fs.main_sample_id
     inner join d_lab_result as lr
          on lr.field_event_id = ss.field_event_id
          and lr.analytical_sample_id = ss.analytical_sample_id
WHERE
     <selection criteria>;
```

These join statements follow the structural relationships between database tables that are described in the *Portland Harbor Interim Database (PHIDB) Design Summary*.

These joins could be created in the opposite order. PHIDB tables have indexes on both primary and foreign key columns to maximize the performance of common sequences of joins such as this one.

The following conceptual query example shows joins from *d_location* down to *d_calculated_chemistry*.

```
SELECT
     <column list>
FROM
     d_location as loc
     inner join d_field_event_location as evloc
          on evloc.location = loc.location
     inner join d_field_collection as coll
          on coll.field_event_id = evloc.field_event_id
          and coll.event_location = evloc.event_location
     inner join d_field_sample as fs
          on fs.field_event_id = coll.field_event_id
          and fs.collection_id = coll.collection_id
     inner join d_calculated_chemistry as cc
          on cc.field_event_id = fs.field_event_id
          and cc.main_sample_id = fs.main_sample_id
WHERE
     <selection criteria>;
```

4.3.2 Non-Chemistry Data

In addition to chemistry data, the PHIDB can store data on toxicity test results, species abundance measurements, and qualitative and quantitative observations made on collections and samples. The following conceptual query examples illustrate SQL join statements that link these types of information to sample and location information. However, as of October 2021, no data are returned in the PHIDB for the following queries. It is anticipated these data will be added in the future.

4.3.3 Sample Measurements

```
SELECT
      <column list>
FROM
      d_location as loc
      inner join d_field_event_location as evloc
            on evloc.location = loc.location
      inner join d_field_collection as coll
            on coll.field_event_id = evloc.field_event_id
            and coll.event_location = evloc.event_location
      inner join d_field_sample as fs
            on fs.field_event_id = coll.field_event_id
            and fs.collection_id = coll.collection_id
      inner join d_sample_measurement as sm
            on sm.field_event_id = fs.field_event_id
            and sm.main_sample_id = fs.main_sample_id
WHERE
      <selection criteria>;
```

4.3.4 Sample Observations

```
SELECT
      <column list>
FROM
      d_location as loc
      inner join d_field_event_location as evloc
            on evloc.location = loc.location
      inner join d_field_collection as coll
            on coll.field_event_id = evloc.field_event_id
            and coll.event_location = evloc.event_location
      inner join d_field_sample as fs
            on fs.field_event_id = coll.field_event_id
            and fs.collection_id = coll.collection_id
      inner join d_sample_observation as sobs
            on sobs.field_event_id = fs.field_event_id
            and sobs.main_sample_id = fs.main_sample_id
WHERE
      <selection criteria>;
```

4.3.5 Toxicity Test Results

```
SELECT
      <column list>
FROM
      d_location as loc
      inner join d_field_event_location as evloc
            on evloc.location = loc.location
      inner join d_field_collection as coll
            on coll.field_event_id = evloc.field_event_id
            and coll.event_location = evloc.event_location
      inner join d_field_sample as fs
            on fs.field_event_id = coll.field_event_id
            and fs.collection_id = coll.collection_id
      inner join d_tox_test_result as tt,
            on tt.field_event_id = fs.field_event_id
            and tt.main_sample_id = fs.main_sample_id
      inner join d_tox_test_batch as tb
            on tb.lab_name = tt.lab_name
            and tb.tox_test_batch = tt.tox_test_batch
WHERE
      <selection criteria>;
```

4.3.6 Species Abundance Measurements

```
SELECT
      <column list>
FROM
      d_location as loc
      inner join d_field_event_location as evloc
            on evloc.location = loc.location
      inner join d_field_collection as coll
            on coll.field_event_id = evloc.field_event_id
            and coll.event_location = evloc.event_location
      inner join d_field_sample as fs
            on fs.field_event_id = coll.field_event_id
            and fs.collection_id = coll.collection_id
      inner join d_species_abundance as abund,
            on abund.field_event_id = fs.field_event_id
            and abund.main_sample_id = fs.main_sample_id
WHERE
      <selection criteria>;
```

4.4 Data Summarization

The following sections illustrate several methods to simplify summarization of PHIDB data.

4.4.1 Summarizing Chemistry Data

The PHIDB can store detailed analytical chemistry data, including results for individual laboratory replicates, but for the purpose of data analysis, these results commonly should be summarized up to the level of the main (interpretive) sample. The detailed representation of analytical results allows correspondingly detailed data selection criteria to be applied, but presents a challenge to data summarization. See Appendix A for fully functional SQL queries that produce output when run against the PHIDB

The PHIDB includes two functions that can be used to simplify the summarization of analytical results. The functions take a specification of the rows of the *d_lab_result* table that are to be used, and return a table of chemistry data summarized up to the level of the main sample. These functions carry out the following operations:

- Standardizes concentration units so that all results for the same material analyzed, fraction analyzed, analyte, measurement basis, and dimension (e.g., mass/mass or mass/volume) have the same units.

- Hierarchically averages concentrations: laboratory replicates are averaged first, then multiple analyses for the same analytical sample are averaged, then results for multiple analytical samples for the main sample (field split) are averaged.

- Qualifiers are propagated through the averaging steps; if any result is estimated (*J*-qualified), the result will be estimated; if any result is detected (not *U*-qualified), the result will be detected; and if any result is rejected (*R*-qualified), the result will be rejected.

- Significant digits are propagated through the averaging steps. The significant digits value for the result will represent a precision that is no greater than the least precise value that is averaged, where precision is represented by the place of the least significant digit (as in one's place, ten's place, etc.). No rounding is performed during averaging.

The two functions are named *chemavg_halfdl* and *chemavg_detpref*. They differ in their treatment of non-detects when a detected value and an undetected value are to be averaged, as follows:

- The *chemavg_detpref* function discards the undetected value and uses only the detected value, or the average if there are multiple detected values. DMP Section 6.2 calls for selecting the minimum non-detect result value as the RD decision data which is consistent with this function.

- The *chemavg_halfdl* function includes non-detects. The function computes an average using the detected value and half the undetected value. This additional function is included for completeness but is not consistent with the DMP.

The primary key of the table that is produced consists of the columns *field_event_id, main_sample_id, material_analyzed, fraction_analyzed, analyte, basis,* and *units.* Attribute columns of the table are *result_value* (e.g., the concentration), *sig_figs, undetected, estimated,* and *rejected.* The last three of these are Boolean values, just as in the *d_lab_result* table.

These functions average across sample delivery groups, preparation methods, and analytical methods. If summarization by analytical method (for example) is desired, the summarization function should be carried out for separate data subsets, where each subset contains data for a different analytical method.

The functions take one argument, which is the name, as a string, of a temporary table or view that contains all of the desired rows of the *d_lab_result* table. The following code snippet illustrates the use of the *chemavg_halfdl* function.

```
CREATE TEMPORARY VIEW selected_chem AS
SELECT
      lr.*
FROM
      d_lab_result as lr
      <join statements as needed>
WHERE
      <selection criteria as needed>;

SELECT
      *
FROM
      chemavg_halfdl('selected_chem');
```

4.4.2 Creating Chemical Qualifier Strings

Data qualifiers are stored in the PHIDB as Boolean values to simplify data selection and for better performance during data summarization. However, single-letter codes are ordinarily used for data qualifiers when producing data tables. The following code snippet illustrates the conversion of the multiple Boolean qualifiers to a single string of qualifier codes.

This will produce an empty string when there are no qualifiers. If SQL null value is desired instead of an empty string, this expression can be wrapped in another case statement that produces a null if all of the Boolean qualifiers are false. This will produce an empty string when

```
case when undetected then 'U' else '' end
     || case when estimated then 'J' else '' end
     || case when rejected then 'R' else '' end as qualifiers
```

there are no qualifiers. If an SQL null value is desired instead of an empty string, this expression can be wrapped in another case statement that produces a null if all of the Boolean qualifiers are false.

```
SELECT
     loc.*
FROM
     d_location as loc
     inner join (
          SELECT
               ST_Difference(a.spatial_extent, b.spatial_extent)
               as sel_area
          FROM
               (SELECT spatial_extent FROM d_field_event
               WHERE field_event_id = 'Event_A') as a
               cross join
               (SELECT spatial_extent FROM d_field_event
               WHERE field_event_id = 'Event_B') as b
          ) as anotb
          on ST_Within(loc.loc_geom, anotb.sel_area);
```

4.5 Unit Conversions

The lookup table for units (*l_unit*) includes two columns, *dimension* and *factor*, that can be used to convert (standardize) units in a data summarization query. The *dimension* column contains abbreviated codes indicating whether the associated units have dimensions of length, mass, mass/mass, mass/volume, or others. The *l_dimension* lookup table defines the dimension codes. The *factor* column in the *l_unit* lookup table contains a numeric value that can be used to convert values that have the same dimension but expressed in different units.

To convert a numeric value (e.g., a depth) from one unit to another, the steps to follow are:

- Join the data table to one copy of the *l_unit* table on the unit code of the value to be converted.

- Join to another copy of the *l_unit* table from which only the row with the target units are selected.

- In the SELECT clause of the query, multiply the numeric value to be converted by the *factor* value from the first *l_unit* table and divide it by the *factor* value from the second *l_unit* table.

This process is illustrated in the following code snippet, which converts sample depths to centimeters. The output includes both the original and converted values.

```
SELECT
      field_event_id,
      main_sample_id,
      sample_depth_upper,
      sample_depth_lower,
      sample_depth_units,
      sample_depth_upper * u1.factor / u2.factor
            as sample_depth_upper_cm,
      sample_depth_lower * u1.factor / u2.factor
            as sample_depth_lower_cm,
      u2.unit as sample_depth_units_cm
FROM
      idb.d_field_sample as fs
      inner join idb.l_unit as u1
            on u1.unit = fs.sample_depth_units
      inner join idb.l_unit as u2 on u2.dimension = u1.dimension
WHERE
      u2.unit = 'cm';
```

An alternative method of performing the join is as follows:

```
SELECT
     field_event_id,
     main_sample_id,
     sample_depth_upper,
     sample_depth_lower,
     sample_depth_units,
     sample_depth_upper * u1.factor / u2.factor
          as sample_depth_upper_cm,
     sample_depth_lower * u1.factor / u2.factor
          as sample_depth_lower_cm,
     u2.unit as sample_depth_units_cm
FROM
     idb.d_field_sample as fs
     inner join idb.l_unit as u1
          on u1.unit = fs.sample_depth_units
     cross join (select * from idb.l_unit where unit='cm') as u2;
```

Some table columns that contain unit codes will only contain unit codes having a single dimension. The column *d_field_sample_depth_units* is an example. Other columns containing unit codes may contain codes of multiple dimensions—*d_lab_result.result_units* is an example. In the latter case the target units should be specified in a temporary table that contains columns for the dimension and the target units, and the dimension value used to join that table to the copy of the *l_unit* table that is joined to the data table of interest (e.g., *d_lab_result*). The temporary table that is used to specify target units could also have additional columns, such as *analyte* and *material_analyzed*, to provide more detailed specifications for unit conversions.

5 ADDITIONAL GUIDANCE

Representatives of performing parties or DEQ needing additional guidance on accessing data from the PHIDB should send inquiries to: data-submittal@phssidb.com or contact your EPA Remedial Project Manager.

APPENDIX A

SQL Queries for Data Selection

APPENDIX A

This appendix presents SQL code that illustrates the data selection and summarization methods described in Section 4.4. These examples are fully functional and will produce output when run against PHIDB. They may serve as a foundation for creating different customized data queries.

Selection of Analytical Chemistry Data

This example illustrates a simple query of analytical chemistry data that produces an output table containing a subset of sample attributes and analytical chemistry measurements. No data summarization is performed by this query, although chemistry data qualifiers are translated from Boolean flags to letter codes. Sampling coordinates are converted to latitude and longitude values. The query returns laboratory results and does not include calculated chemistry.

```
SELECT
     evloc.field_event_id,
     evloc.event_location,
     ST_Y(loc.loc_geom) as latitude,
     ST_X(loc.loc_geom) as longitude,
     ST_SRID(loc.loc_geom) as spatial_reference_system,
     coll.sample_date,
     fs.sample_material,
     lr.material_analyzed,
     lr.analyte,
     lr.result,
     lr.result_units,
     lr.basis,
     case when lr.undetected then 'U' else '' end
          || case when lr.estimated then 'J' else '' end
           as qualifiers,
     lr.sig_figs
FROM
     d_location as loc
     inner join d_field_event_location as evloc
          on evloc.location = loc.location
     inner join d_field_collection as coll
          on coll.field_event_id = evloc.field_event_id
          and coll.event_location = evloc.event_location
     inner join d_field_sample as fs
          on fs.field_event_id = coll.field_event_id
          and fs.collection_id = coll.collection_id
     inner join d_analytical_sample as ss
          on ss.field_event_id = fs.field_event_id
          and ss.main_sample_id = fs.main_sample_id
     inner join d_lab_result as lr
          on lr.field_event_id = ss.field_event_id
          and lr.analytical_sample_id = ss.analytical_sample_id
WHERE
     coll.field_event_id = 'PDI_2018'
     and coll.collection_material = 'Sediment'
     and fs.sample_depth_upper = 0
     and not lr.rejected
     and lr.reportable;
```

The following query illustrates an export format example that returns unsummarized laboratory results and summarized calculated chemistry results using the UNION statement. This query selects all sediment depth data that is reportable and not rejected for the 'PDI_2018' field event. Chemistry data qualifiers are translated from Boolean flags to letter codes. Sampling coordinates are converted to latitude and longitude values. The main and analytical sample IDs are included in addition to analytical method and method detection limit. This statement can be modified in both of the where clauses to only select surface sediment by removing the comment dashes "—" in front of the 'fs.sample_depth_upper = 0' statement.

```
SELECT
      evloc.field_event_id,
      evloc.event_location,
      ss.main_sample_id,
      ss.analytical_sample_id,
      fs.sample_depth_upper,
      fs.sample_depth_lower,
      fs.sample_depth_units,
      ST_Y(loc.loc_geom) as latitude,
      ST_X(loc.loc_geom) as longitude,
      ST_SRID(loc.loc_geom) as spatial_reference_system,
      coll.sample_date,
      lr.date_analyzed,
      lr.lab_name,
      fs.sample_material,
      lr.material_analyzed,
      lr.analyte,
      lr.analytical_method,
      lr.result,
      lr.result_units,
      lr.basis,
      lr.mdl,
      lr.mdl_units,
      case when lr.undetected then 'U' else '' end
            || case when lr.estimated then 'J' else '' end as
qualifiers,
      lr.sig_figs
FROM
      d_location as loc
      inner join d_field_event_location as evloc
            on evloc.location = loc.location
      inner join d_field_collection as coll
            on coll.field_event_id = evloc.field_event_id
            and coll.event_location = evloc.event_location
      inner join d_field_sample as fs
            on fs.field_event_id = coll.field_event_id
            and fs.collection_id = coll.collection_id
      inner join d_analytical_sample as ss
            on ss.field_event_id = fs.field_event_id
            and ss.main_sample_id = fs.main_sample_id
      inner join d_lab_result as lr
            on lr.field_event_id = ss.field_event_id
            and lr.analytical_sample_id = ss.analytical_sample_id
WHERE
      coll.field_event_id = 'PDI_2018'
      and coll.collection_material = 'Sediment'
      --and fs.sample_depth_upper = 0
      and not lr.rejected
      and lr.reportable
UNION
Select
    evloc.field_event_id,
      evloc.event_location,
      ss.main_sample_id,
      ss.analytical_sample_id,
      fs.sample_depth_upper,
```

```
      fs.sample_depth_lower,
      fs.sample_depth_units,
      ST_Y(loc.loc_geom) as latitude,
      ST_X(loc.loc_geom) as longitude,
      ST_SRID(loc.loc_geom) as spatial_reference_system,
      coll.sample_date,
      Null as date_analyzed,
      Null as lab_name,
      fs.sample_material,
      cc.material_analyzed,
      cc.analyte,
      'Calculated Chemistry' as analytical_method,
      cc.result,
      cc.result_units,
      cc.basis,
      Null as mdl,
      Null as mdl_units,
      --Null as idl,
      --Null as idl_units,
      case when cc.undetected then 'U' else '' end
            || case when cc.estimated then 'J' else '' end as
qualifiers,
      Null as sig_figs
FROM
      d_location as loc
      inner join d_field_event_location as evloc
            on evloc.location = loc.location
      inner join d_field_collection as coll
            on coll.field_event_id = evloc.field_event_id
            and coll.event_location = evloc.event_location
      inner join d_field_sample as fs
            on fs.field_event_id = coll.field_event_id
            and fs.collection_id = coll.collection_id
      inner join d_analytical_sample as ss
            on ss.field_event_id = fs.field_event_id
            and ss.main_sample_id = fs.main_sample_id
      inner join d_lab_result as lr
            on lr.field_event_id = ss.field_event_id
            and lr.analytical_sample_id = ss.analytical_sample_id
      inner join d_calculated_chemistry as cc
            on cc.field_event_id = ss.field_event_id
            and cc.main_sample_id = ss.main_sample_id
WHERE
      coll.field_event_id = 'PDI_2018'
      and coll.collection_material = 'Sediment'
      --and fs.sample_depth_upper = 0
      and not cc.rejected;
```

Summarization of Chemistry Data

This example illustrates the summarization of surface sediment analytical chemistry data for the 'PDI_2018' field event up to the level of the interpretive sample, and the merging of those analytical results with calculated values that have been stored in the database. The summarization step makes use of the *chemavg_halfdl* function. Qualifiers are converted from Boolean flags to character codes for both the summarized analytical chemistry data and the calculated chemistry data. Sampling location coordinates are converted to latitude and longitude values. The following messages will be displayed the first time that this code is run using an interactive query tool.

```
NOTICE: table "tmpt_selchem" does not exist, skipping
NOTICE: table "tmpt_dyndefunit" does not exist, skipping
NOTICE: table "tmpt_chmstdunits" does not exist, skipping
NOTICE: table "tmpt_avg_lrep" does not exist, skipping
NOTICE: table "tmpt_avg_analsamp" does not exist, skipping
NOTICE: table "tmpt_avg_intsamp" does not exist, skipping
```

```
CREATE TEMPORARY VIEW selected_samps AS
SELECT
      fs.field_event_id,
      fs.main_sample_id
FROM
      d_field_collection as coll
      inner join d_field_sample as fs
            on fs.field_event_id = coll.field_event_id
            and fs.collection_id = coll.collection_id
WHERE
      coll.field_event_id = 'PDI_2018'
      and coll.collection_material = 'Sediment'
      and fs.sample_depth_upper = 0;

CREATE TEMPORARY VIEW selected_chem AS
SELECT
      lr.*
FROM
      selected_samps as fs
      inner join d_analytical_sample as ss
            on ss.field_event_id = fs.field_event_id
            and ss.main_sample_id = fs.main_sample_id
      inner join d_lab_result as lr
            on lr.field_event_id = ss.field_event_id
            and lr.analytical_sample_id = ss.analytical_sample_id
WHERE
      not lr.rejected
      and lr.reportable;

SELECT
      evloc.field_event_id,
      evloc.event_location,
      ST_Y(loc.loc_geom) as latitude,
      ST_X(loc.loc_geom) as longitude,
      ST_SRID(loc.loc_geom) as spatial_reference_system,
      coll.sample_date,
      fs.sample_material,
      lr.material_analyzed,
      lr.analyte,
      lr.result_value,
      lr.units,
      lr.basis,
```

```
      case when lr.undetected then 'U' else '' end
            || case when lr.estimated then 'J' else '' end
            as qualifiers
FROM
      d_location as loc
      inner join d_field_event_location as evloc
            on evloc.location = loc.location
      inner join d_field_collection as coll
            on coll.field_event_id = evloc.field_event_id
            and coll.event_location = evloc.event_location
      inner join d_field_sample as fs
            on fs.field_event_id = coll.field_event_id
            and fs.collection_id = coll.collection_id
      inner join chemavg_halfdl('selected_chem') as lr
            on lr.field_event_id = fs.field_event_id
            and lr.main_sample_id = fs.main_sample_id
UNION
SELECT
      evloc.field_event_id,
      evloc.event_location,
      ST_Y(loc.loc_geom) as latitude,
      ST_X(loc.loc_geom) as longitude,
      ST_SRID(loc.loc_geom) as spatial_reference_system,
      coll.sample_date,
      fs.sample_material,
      calc.material_analyzed,
      calc.analyte,
      calc.result,
      calc.result_units,
      calc.basis,
      case when calc.undetected then 'U' else '' end
            || case when calc.estimated then 'J' else '' end
            as qualifiers
FROM
      selected_samps as ss
      inner join d_field_sample as fs
            on fs.field_event_id = ss.field_event_id
            and fs.main_sample_id = ss.main_sample_id
      inner join d_field_collection as coll
            on coll.field_event_id = fs.field_event_id
            and coll.collection_id = fs.collection_id
      inner join d_field_event_location as evloc
            on evloc.field_event_id = coll.field_event_id
            and evloc.event_location = coll.event_location
      inner join d_location as loc
            on loc.location = evloc.location
      inner join d_calculated_chemistry as calc
            on calc.field_event_id = fs.field_event_id
            and calc.main_sample_id = fs.main_sample_id
WHERE
      not calc.rejected;
```