Naval Research Laboratory
Washington, DC 20375-5320

NRL/6137/MR—2022/15

Test Facility Data Entry Recommendations

MATTHEW R. FIRST

STEPHANIE H. ROBBINS-WAMSLEY

Marine Biological Engineering Section
Chemistry Division

TIMOTHY P. WIER

VANESSA MOLINA

SCOTT C. RILEY

Excet, Inc.
Springfield, VA

JONATHAN F. GRANT

Battenkill Technologies
Manchester Center, VT

July 26, 2022

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
26-07-2022	NRL Memorandum Report	07/15/2021 – 07/14/2022

4. TITLE AND SUBTITLE

Test Facility Data Entry Recommendations

5a. CONTRACT NUMBER
70Z023-20-KMMS03100

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Matthew R. First, Stephanie H. Robbins-Wamsley, Timothy P. Wier*, Vanessa Molina*, Scott C. Riley*, and Jonathan F. Grant**

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
3616

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/6130/MR--2022/15

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

United States Coast Guard; Environmental Standards Division (CG-OES-3)
2703 Martin Luther King Junior Avenue S.E.
Washington, DC 20593-7509

10. SPONSOR / MONITOR'S ACRONYM(S)
USCG-OES3

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

*Excet, Inc. 6225 Brandon Ave., #360, Springfield, VA 22150; **Battenkill Technologies, Inc., 3429 Richville Road Suite C, Manchester Center, VT 05255; See Appendix A for recommended data tables.

14. ABSTRACT

This report provides guidance on submitting electronic data associated with a Ballast Water Management System (BWMS) Type Approval (TA) Application. The goal of this effort is to simplify the TA Application submission, review process, and data archival as well as to facilitate the meta-analysis of multiple datasets. For a single application, all information will reference a unique Application Identification (ID), which is provided by the USCG upon receipt to the Letter of Intent. Basic descriptive data—including information about the Independent Laboratory (IL) and the BWMS—are stored in the Main Table, which links to separate tables with detailed information on the BWMS (e.g., models used in testing, specifications, and manufacturer's claims) and the IL's Assessment. Next in the hierarchy is a table of Test Cycles. For this set of recommendations, the Test Cycles described are all associated with land-based, Biological Efficacy (BE) testing, although this basic structure will be suitable for other TA Application data, such as shipboard BE testing. A single Test Cycle will have multiple Samples (the next level in the data hierarchy), and each Sample will have multiple Analyses. Analysis tables include discrete laboratory analyses (such as measurements of concentrations of organisms ≥10 and <50 μm, ≥50 μm, E. coli) but also continuous data, such as electronic measurements collected by a flow or pressure sensor installed within piping. Each analysis type has a unique table to store analysis-specific information and results, and the each of these tables will house records from all Samples and all Test Cycles. Finally, we provide a short list of "best practices" and an example set of table templates. Adoption of this approach for transmitting data will likely reduce the effort required by the Test Facilities, the ILs, and the USCG.

15. SUBJECT TERMS

Ballast water management System (BWMS) Type approval (TA) testing Data management
Reporting Biological efficacy (BE) testing

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Matthew First
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	U	32	**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U			(202) 815-1371

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.

EXECUTIVE SUMMARY

This report provides guidance on submitting electronic data associated with a Ballast Water Management System (BWMS) Type Approval (TA) Application. The goal of this effort is to simplify the TA Application submission, review process, and data archival as well as to facilitate the meta-analysis of multiple datasets. For a single application, all information will reference a unique Application Identification (ID), which is provided by the USCG upon receipt to the Letter of Intent. Basic descriptive data—including information about the Independent Laboratory (IL) and the BWMS—are stored in the Main Table, which links to separate tables with detailed information on the BWMS (e.g., models used in testing, specifications, and manufacturer's claims) and the IL's Assessment. Next in the hierarchy is a table of Test Cycles. For this set of recommendations, the Test Cycles described are all associated with land-based, Biological Efficacy (*BE*) testing, although this basic structure will be suitable for other TA Application data, such as shipboard *BE* testing. A single Test Cycle will have multiple Samples (the next level in the data hierarchy), and each Sample will have multiple Analyses. Analysis tables include discrete laboratory analyses (such as measurements of concentrations of organisms ≥10 and <50 µm, ≥50 µm, *E. coli*) but also continuous data, such as electronic measurements collected by a flow or pressure sensor installed within piping. Each analysis type has a unique table to store analysis-specific information and results, and the each of these tables will house records from all Samples and all Test Cycles. Finally, we provide a short list of "best practices" and an example set of table templates. Adoption of this approach for transmitting data will likely reduce the effort required by the Test Facilities, the ILs, and the USCG.

This page intentionally left blank.

Contents

This page intentionally left blank.

TERMS

BE	Biological Efficacy
BWMS	Ballast Water Management System
CFR	Code of Federal Regulation
DOM/C	Dissolved Organic Matter/Carbon
ETV	Environmental Technology Verification
IL	Independent Laboratory
MSC	Marine Safety Center

MM	Mineral Matter
POM/C	Particulate Organic Matter/Carbon
TA	Type Approval
TF	Test Facility
TSS	Total Suspended Solids
TRC	Treatment-Rated Capacity
USCG	United States Coast Guard

This page intentionally left blank.

1. INTRODUCTION

At the time of writing, the U.S. Coast Guard (USCG) has issued Type Approval (TA) certificates to 48 Ballast Water Management Systems (BWMS)[1]. As part of a TA application, an Independent Laboratory (IL) submits the results of testing to the USCG's Marine Safety Center (MSC). Test reports must follow the requirements listed in the Code of Federal Regulations (CFR), as summarized below (Box 1).

Box 1. Relevant requirements codified in the US Code of Federal Regulations (CFR).

The Test Report prepared and submitted by an independent laboratory must be formatted as set out below. The Test Report must include, in addition to the information required by 46 CFR 159.005-11, information as follows:

46 CFR §162.060-34 Test Report requirements

(a) Summary statement with the following information:

(1) Name of the independent laboratory (IL) and all test facilities, subcontractors, and test organizations involved in testing the ballast water management system (BWMS).

(2) Name of manufacturer.

(3) BWMS model name.

(4) The IL's assessment that the BWMS –

(i) Has demonstrated, under the procedures and conditions specified in this subpart for both land-based and shipboard testing, that it meets the ballast water discharge standard requirements of 33 CFR part 151, subparts C and D

…

(e) For each test conducted, summary descriptions of -

(1) Test conditions

…

(4) Results and discussion.

(f) Appendices, including -

…

(3) Data generated during testing and evaluations;

(4) Quality Assurance and Quality Control records;

…

(6) Relevant records and tests results maintained or created during testing;

[1]USCG Marine Safety Center; *Approved BWMS and Status of Applications*; Accessed 07-Jun-2022 https://www.dco.uscg.mil/Portals/9/MSC/BWMS/BWMS_Approval_Status_3JUN22.pdf

Manuscript approved July 24, 2022.

Box 1. Continued

46 CFR §159.005-11 Approval inspection or test report:

(a) Each approval inspection or test report must contain the following:

(1) The name of the manufacturer.

(2) If the inspections or tests are performed or supervised by an independent laboratory, the name and address of the laboratory.

(3) The trade name, product designation (such as model numbers), and a brief description of the equipment or material inspected or tested.

(4) The time, date, and place of each approval inspection and test.

(5) The name and title of each person performing, supervising, and witnessing the approval inspections or tests.

(6) The performance data for each test required in the applicable subpart, including a description of each failure.

46 CFR §162.060-5 Incorporation by reference.

(d) U.S. Environmental Protection Agency (EPA), Environmental Technology Verification Program, National Risk Management Research Laboratory Office of Research and Development, U.S. Environmental Protection Agency, 2890 Woodbridge Avenue (MS-104), Edison, New Jersey 08837.

(1) EPA/600/R-10/146, Generic Protocol for the Verification of Ballast Water Treatment Technologies, version 5.1, (dated September 2010), IBR approved for §§ 162.060-26 and 162.060-28 (ETV Protocol).

Importantly, the CFR requires:

- summary descriptions of the test conditions [46 CFR §162.060-34(e)(1)],
- results [-34(e)(4)], and
- data generated during testing and evaluations [-34(f)(3)].

The large data set generated from these test reports is valuable beyond its primary purpose of TA certification of a single system: together, the data combined from multiple TA applications allow comparisons across BWMS type, trends in performance, typical environmental characteristics, and extremes of test conditions, among other important queries. The meta-analysis of TA data will inform policy makers, but it may also highlight differences among test facilities (TFs), such that the USCG can craft guidance to harmonize differences in test protocols or methods. This report results from a (nearly) year-long effort to digitize the existing library of TA reports from land-based tests. Here, we describe the structure of the database developed to organize and store test data, such that users may pose specific queries, including queries in support of program analysts and approval authorities.

Our intent is for ILs to adopt this standard form to report test data, such as would be included as an alternative to data appendices of the test reports. In this system, data would be transmitted in a common structure, using comma-separated values that are compatible with most spreadsheet and database software (e.g., Microsoft Excel [.xls]). Common languages and file formats such as *.html* and *.xml* are also easily interpreted and allow the end users options for converting the data to preferred formats. ***Records***[2] (i.e., "rows" of the data tables) would contain all relevant ***Fields*** and indices, delimited or structured such that a user could intuitively relate Records across data tables.

Final reports transmitted in a rich-text format (such as .docx) or in a portable document format (.pdf) may contain structured data tables, but the variations in table layout and test formatting complicate the process of importing data. Tables embedded in text also break across pages, have custom formatting, have notations or references[3], or include only summary data. All of these characteristics challenge approaches for extracting data. Currently, there is no requirement to provide **Structured Data**[4] upon submission of the final report for TA, but standardized submission of Structured Data would simplify the reporting requirements for the IL and will simplify the USCG's review, potentially shortening the time and effort required to evaluate TA applications. For example, report reviewers could perform a series of standard queries and generate reports with custom tables and graphs immediately upon receipt of the Structured Data, such that the custom data displays are available for reference during the review of the text.

The purpose of this report is to describe the basic data structure requirements, such that *electronic* data transmitted to the USCG are easily assimilated and interpreted. The terms and concepts described above—Structured Data, Records, Fields, and Indices—are defined specifically herein, with examples relevant to applications for TA of BWMS. This report considers the general template for land-based, biological efficacy (*BE*) testing, such as those outlined directly in the CFR or incorporated into the Environmental Technology Verification (ETV) Program Protocol (See §2). We list the basic data type requirements and the general table structure (§3) and recommended best practices (§4). In the survey of the first ~40 TA packages, we found a variety of terms, but much of the crucial data for land-based testing could be mapped into a limited number of consistently defined Fields. **Appendix A** shows these fields for the "standard" tables, and **Appendix B** provides an example of a **Table Description**, which

[2] Technical terms defined within this document appear in bold typeface at their first use. Common terms (e.g., "Fields", "Records") are capitalized in subsequent appearances, as to indicate they are used in a technical, not general sense.
[3] Such as footnotes.
[4] Here, "Structured Data" is information that conforms to an organization, such as a table with column headings (*Fields*) and indexed rows (*Records*).

provides information not easily included in field headers (such as data types and units for numerical data). Appendix C is a template for data entry in Microsoft Excel format.

Note	This guidance primarily focuses on land-based *BE* testing, though the general format will be adaptable for shipboard tests and operational testing, among others.

2. General Test Design

The ETV[5] describes the procedures for land-based, *BE* testing. This protocol is incorporated by reference into the CFR (46 CFR §162.060-5). A land-based *BE* test—at its core—evaluates whether treatment by a BWMS results in organism concentrations that meet the USCG discharge standard. Test ("challenge") water is characterized by the TF, treated by the BWMS, and discharged; the concentration of living or viable organisms indicates of the whether the BWMS is effective at its purpose, but the ETV requires additional indicators to verify the test integrity, especially the analysis of uptake water and control tank discharge. Figure 1 illustrates the general test design and shows the potential sample types and locations.

[5]US Environmental Protection Agency, *Generic Protocol for the Verification of Ballast Water Treatment Technology*, Accessed 16-Apr-2021; https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P10097A4.txt

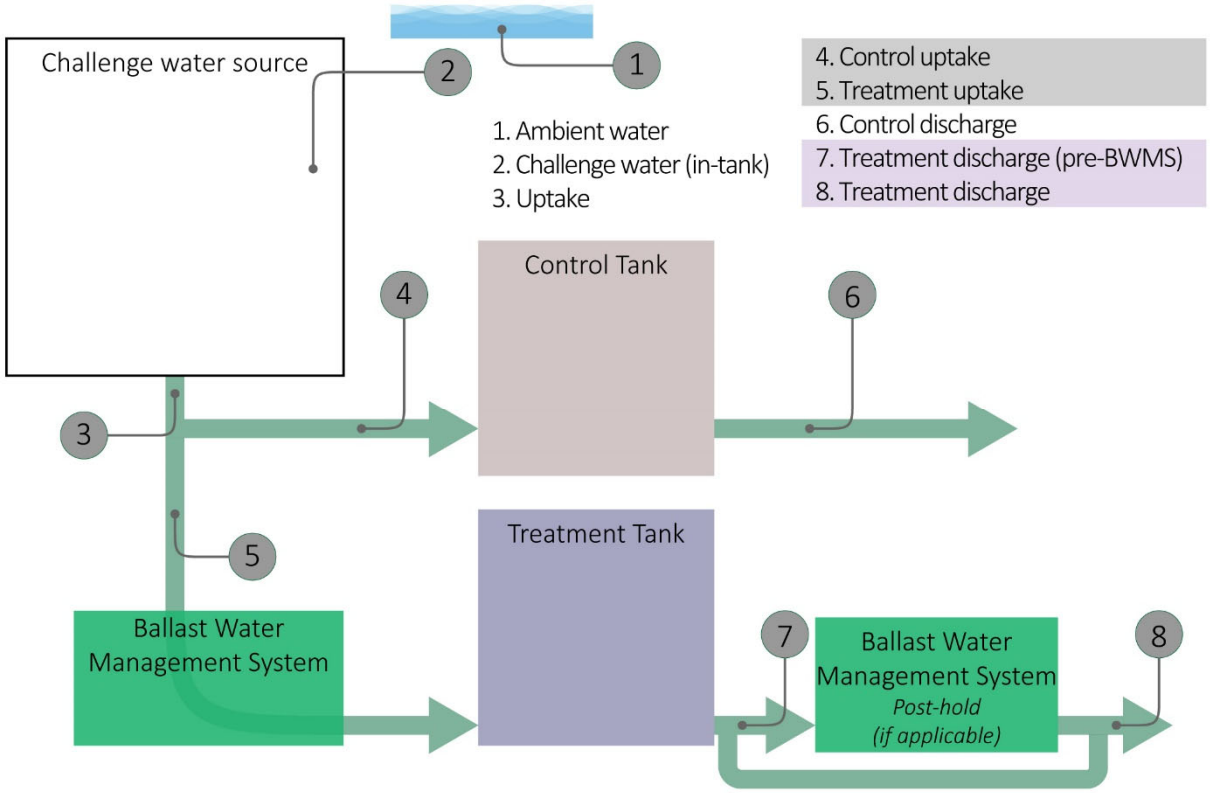

Figure 1. General land-based test schematic. See text for details.

This schematic considers *eight* potential sample sources for a single trial, although not all trials will include each of these. For example, depending on the treatment technologies employed, a BWMS may perform treatment during final discharge, and a sample before the BWMS on treated discharge (i.e., #7) may be collected and analyzed. If treatment occurs exclusively upon uptake or in the tank, only a single treatment discharge sample is expected. This is shown on Figure 1 as the bypass around the BWMS, which still leads to the final sample point (#8, *Treatment Discharge*). The specific design for land-based, *BE* testing will vary based upon the treatment approach. For example, the ETV considered some of the possible treatment options, which are illustrated below (Fig. 2).

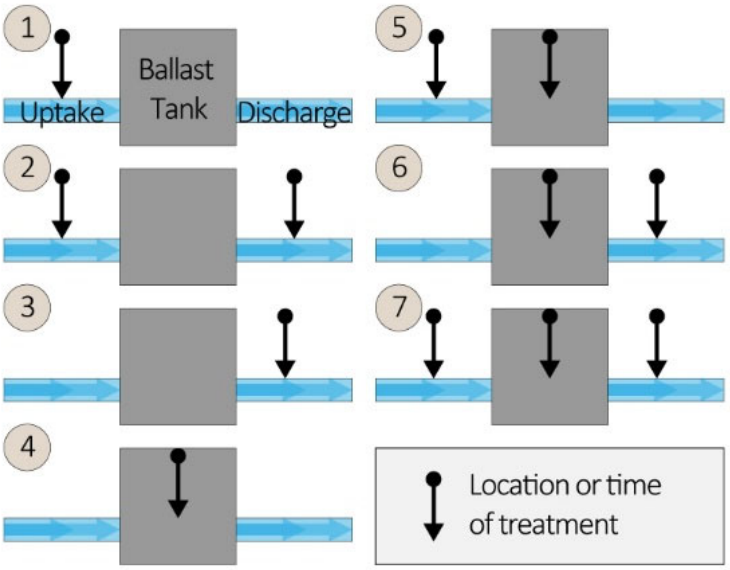

Figure 2. Potential treatment locations or times. Treatments (shown here as arrows) may occur during uptake, while water is in the tank, during discharge, or in any combination. Derived from the ETV §5.4, Table 6.

Test designs may also vary based upon the TF's system configuration. For example, a TF may collect separate samples for both the control and treatment tanks (e.g., #4 and #5 on Fig. 1), or it may collect a single sample, prior to splitting the uptake stream into the control and discharge tanks (#3; Fig. 1). Additional samples—beyond those named in Figure 1—may be defined, but these should be clearly described (see §3.3).

3. Basic Data Structure

The general structure for the data table is shown in Figure 3: It includes a "top tier" of basic information tables, a second tier describing the Test Cycles, a third tier with Samples, and finally, making up the fourth tier are numerous "analyses" tables, each dedicated to a particular analysis. These tables are described in the sections below.

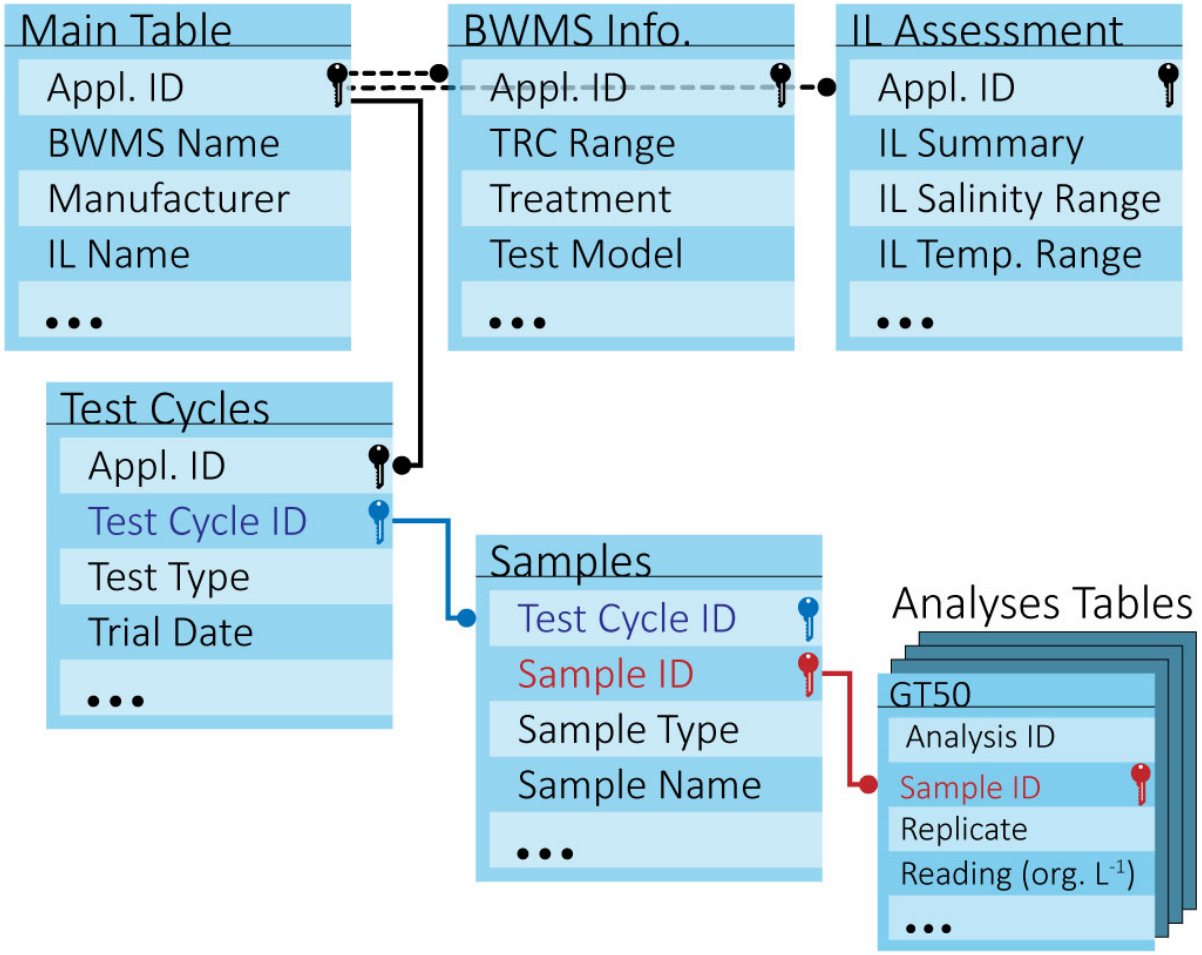

Figure 3. General structure of the database, as arranged by separate data tables. In this figure, unique indices are shown by color-coded "keys" that show the linkages across tables. The figure truncates the list of Fields (column headers on a data table) for simplicity, but a full list of standard Fields are in **Appendix A**. See text for additional details.

3.1. Application ID

All submitted data are organized under a single Application ID. The USCG Marine Safety Center (MSC) will provide ILs with an Application ID upon request or upon receiving a Letter of Intent, as required in 46 CFR 162.060-10(a). This Application ID may correspond to a Project Number used by MSC for internal records. Additional submissions—such as revisions to data or supplemental data received afterwards—will also refer to the same Application ID. The USCG will track multiple submissions, the dates received, and other submission information in an internal database (separate from the database intended for test results). Upon approval, the dataset will be added to the larger database.

3.2. TA Application Main Table

An application submitted to the USCG for consideration will deliver some basic information on the BWMS and its manufacturer, the IL overseeing the testing (as well as sub-laboratories performing specific tests), the manufacturer's claims, and the IL's assessment (as stated in 46 CFR §162.060-34(a)4). In summary test reports, this information is typically conveyed within the first few pages, such as in the Executive Summary or in the front matter of the document. The data in these "top-tier" tables are generally used for quick reference of basic information, and the USCG may use these data for internal reports or to populate information on TA Certificates.

While a particular TA Application will have only one Application ID (essentially, a single record), the USCG's Application Table will have 48+ records, representing TA Applications accepted, rejected, or under review. Data are organized primarily by the TA application package, and if approved, these records will later be identified with the short TA certificate number (e.g., TA-**30**, for 162.060.**30**).

The **Main Table** houses the essential information describing the BWMS, the IL, and the test facility, among other information. At the same level in the hierarchy (but partitioned in separate tables) are the **BWMS Information** and the **IL Assessment**. The BWMS Information describes the treatment system, and it includes basic reference information on the treatment technology, the treatment-rated capacity (TRC), the models used in testing, and the manufacturer's claims (e.g., temperature range) among other items. The IL Assessment is a summary of the findings based upon test results and independent analyses. This includes the general outcome of the tests as well as the operational temperature range, the salinity range, etc.

For the USCG's main TA database, all records are traced back to the **Application ID**. An IL will only submit one application at a time (with a single Application ID), and all the records in their submission will reference that single Application ID. Nested within that single Application ID are the data relevant to the topics required by the CFR:

- Land-based test requirements (46 CFR §162.060-26)
- Shipboard test requirements (§162.060-28)
- Component testing requirement (§162.060-30)
- Active substances, preparations, and relevant chemical testing (§162.060-32)

Here, we describe the data formatting for land-based test data; a subsequent report describing shipboard testing is forthcoming upon the completion of data set of shipboard data for BWMS with existing TAs.

Within a single Application ID, data are organized first by Test Cycle[6], then by Sample, and finally by Analysis (Fig. 3). These are described in the sections below.

3.3. Test Cycle

A **Test Cycle** is a single, independent run of the test performed with specified criteria, such as the salinity of the challenge water. Fifteen (15) or more Test Cycles are expected for each land-based TA application package: 5 independent, replicate Test Cycles at each of 3 salinity ranges. Test Cycles will have a unique ID, and all Test Cycles are described in a data table that describes the test characteristics, dates of testing, and includes information on the challenge water characteristics for organism concentrations, salinity, volume, hold time, and flow rate, among others.

All relevant data on Samples and Analyses reference a Test Cycle, and all the data generated from a given Test Cycle will be traceable to a single **Test Cycle ID.** The Test Cycle data table will describe all aspects of the each Test Cycle, including basic information on the start date, the challenge water preparation or collection date and time, the target salinity ("freshwater, brackish, marine"), etc.

3.4. Samples

As mentioned, a test design may not include all samples listed in Figure 1. The test design may also include samples *not* shown in Figure 1, such as those collected to address specific characteristics of the BWMS or requirements of the BWMS manufacturer, the IL, or the TF. For example, samples may be collected prior to or following a neutralization unit. ***These samples should still be reported, especially when relevant to the performance of the BWMS or used to qualify internal controls or data quality objectives.***

The sample name should—as best as possible—use the naming convention in Figure 1, which indicates:

- **Source:** *Control* or *Treatment* tank (also: ambient water, challenge water tank)
- **Direction:** *Uptake* or *Discharge*
- **Relation to Treatment (BWMS):** *Before* or *After BWMS*

Examples of the main sample types are included in a template for data entry (Appendix C). In-line sensors are treated as unique Samples, with unique characteristics (start and end times, flow rates, etc.). For these instances, the sample name should indicate the sample was collected "in-

[6] As discussed, this description considers only Test Cycles performed during land-based testing. Potentially Test Cycles for shipboard and operational testing can be listed within the Test Cycles table.

line"; See Table 1 as an example). The sampling information is described in a table of Samples; each Sample will have a unique index and descriptions of the collection time, volume, replicate number, and flow rate, among other characteristics.

An important characteristic of the Sample is the collection method, as water samples may be time-integrated (collected over long periods) or discrete grabs (collected almost instantaneously). The field Collection Method will indicate a "time-integrated" or "discrete grab" sampling approach. Sample Start and End Times and Sample Volume can indicate actual sample flow rate. If multiple, unique samples are collected (i.e., grab samples, each with different containers and sample materials), the sample should be identified as a Field Replicate (see Box 2). Finally of note, each sample record will contain a field for text notes, which allow the test personnel to make observations and explanations, as appropriate.

Box 2. Field Replicates, Analytical Replicates, and Readings

The commonly used terms are define below:

Field Replicates (FR): Multiple samples originating from a single sample location during a single Test Cycle. For example, three consecutive samples may be collected from a single port (e.g., control discharge). These samples may represent portions of a larger sample, such as the beginning, middle, and end of the control tank discharge. They may also represent a narrow portion of an operation, e.g., three samples, collected sequentially during the middle of the tank operation. In this database, FR are defined and indexed in the Samples Table.

Example:

Sample ID	Name	FR
SAMP-001	Control Uptake	1
SAMP-002	Control Uptake	2
SAMP-003	Control Uptake	3

Analytical Replicates (AR): Also called "subsamples", AR are taken from well-mixed samples (whether a single sample or field replicates). AR are not independent, but they may be used to measure variation of the values due to random error. In this database, AR are defined and indexed in the specific **Analyses** table.

Readings: In the case where an analytical instrument may be able to re-analyze a discrete aliquot of water, these are referred to as multiple readings. For example, if an instrument measures the fluorescence or water color, a single aliquot (such as in a cuvette) may be read multiple times without compromising the sample integrity. In the case where multiple readings are used, a field labeled "Readings" should be included. In this database, AR are defined and indexed in the specific **Analyses** table.

Example

Analysis ID	Sample ID	AR	Reading
pH-001	SAMP-001	1	1
pH-002	SAMP-001	1	2
pH-003	SAMP-001	1	3
pH-004	SAMP-001	2	1
pH-004	SAMP-001	2	2
pH-005	SAMP-001	2	3

Table 1. Example record for a sample collected to analyze total residual oxidant (TRO). This sample type is not among the core sample types (as described in the ETV), so this example provides guidance on defining a Sample not listed in Figure 1.

Sample ID	SAMP-001
Sample name	Treatment Discharge after Neutralization
Purpose of Sample	Treatment Residual Oxidant analysis
Collection Method	Grab sample
Notes	In-line, isokinetic sample port located on the discharge line approximately 10 m following the neutralization injection port (refer to diagrams and pictures, as appropriate)
Sample Processing	Whole water (unfiltered)
Target Sample Volume (mL)	500
*Flow Rate (min.; mL h^{-1})**	3000
*Flow Rate (max.; mL h^{-1})**	12000
Analyses Planned	Temperature (via digital thermometer); °C DPD free chlorine; mg L^{-1}
Sample Collection Time	Approximately the mid-point of discharge (20-40 minutes following discharge)
Collection Date	3/1/2021
Collection Start Time	12:00:00
Collection End Time	12:01:45
*Final Sample Volume (mL)**	150
Field Replicate	Number of field replicates
Source name	Treatment Discharge

*Units should be included in the Field Name but should also be defined in a separate Table Description (see §4.1), in this case, Flow Rates (both minimum and maximum) are mL hour $(h)^{-1}$; Sample Volume is in mL.

Note	All numeric values in a field should be reported in the sample unit (e.g., mL or L, but not both). Where both units are used, it is preferable to convert all values to a single unit. *If necessary*, a field with mixed units must have a separate Field indicating the unit.

3.5. Analyses

A single sample will generally have many analyses, and those samples collected directly for *BE* tests will supply aliquots for the analysis of organisms ≥50 µm, ≥10 and <50 µm, indicator

microbes, and other water characteristics. Analysis data are stored in separate tables, each dedicated to a single parameter (e.g., organisms ≥50 µm, *E. coli*, dissolved organic carbon, etc.).

Each of the **Analyses** tables will have specific Fields (i.e. column headers) for critical information, including sample volumes, aliquot volumes, quantity measured (e.g., living organisms detected) and calculated result (**Final Concentration**). The Fields should provide sufficient information to trace the Final Concentration to the factors used in its calculation. **Analysis Start** and **End Times, Notes**, and ***Data Flags*** should also be included, but additional information collected by the laboratory (reagent lot numbers, analyst identifications) *are not* needed in the Analyses table. Typical Fields used in Analyses tables are shown below: Fields for raw and summary data are in Figures 4 and 5, respectively.

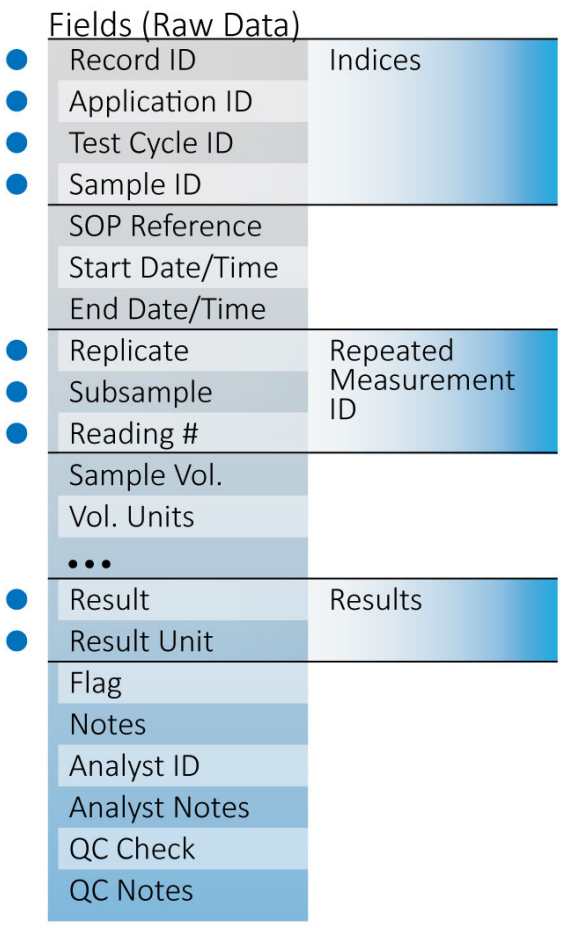

Figure 4. Typical Field names used for Analyses tables with **Raw Data**. Groups of Fields include Indices, Repeated Measurement Identifiers (ID) and Results (see descriptions below). Fields marked with a blue symbol are essential, whereas other fields may be included and give context for the result. Ellipsis indicate space for additional fields (e.g., dilution, aliquot volume, etc.).

Indices: The Record ID is a unique index for each row of the Analyses table. Values are typically alpha-numeric with consecutive integers (e.g., R-0001, R-0002, using leading zeros). Other indices are used in queries to find, sort, or group particular data sets or Records.

Repeated Measurement ID: A sample may have been partitioned into several vessels (e.g., Field Replicate 1, 2, and 3), but the analysis may report multiple subsamples or readings (see **Box 2** for further details).

Results: The final, numerical outcome of the analysis and, in a separate field, the units of the results, including any non-numerical information (e.g., <, >, etc.).

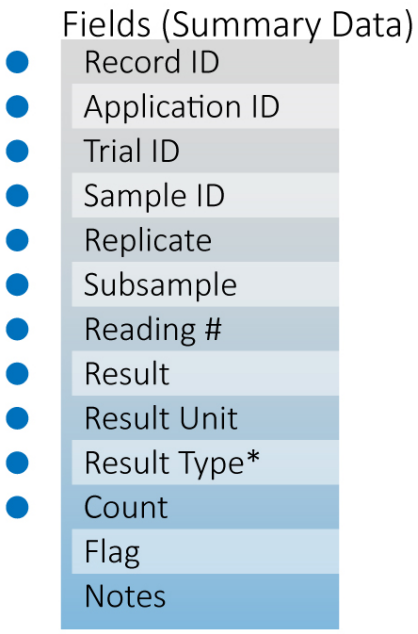

Figure 4. Typical Field names used for Analyses Tables with **Summary Data**. The general groups shown in Figure 4 apply to these Fields, although they are not shown here.

*The main difference between the two types of Analyses results is that the "Result Type", which here should indicate whether the Result is the arithmetic mean, standard deviation, minimum, maximum of the range of values, for example. "Count" is the numerical size of the population considered in the analysis (e.g., the number of values used to calculate the Result.

If the summary data includes multiple values, such as the mean, the standard deviation (SD), the standard error, and the coefficient of variation, e.g., each value is reported as a separate Record (row). All summary values would be found in a query for that Sample ID:

Query:	[Sample ID] = "SAMP-001"; [Analysis Table] = pH				
Results:	SAMP-001	Mean of Range:	7.8	Count:	6
		SD of Range:	0.12	Count:	6
		CV of Range:	1.5	Count:	6
		Min. of Range:	7.6	Count:	6
		Max. of Range:	7.9	Count:	6

4. Best Practices and Pitfalls

4.1. Include table descriptions

A Table Description will list the field name and the data type, typically: "text", "numerical", "date", "time", "binary", or "index". For numerical data, a separate field (typically of type "text") should be provided for units (e.g., "living org. mL^{-1}", "°C", etc.). An example table description for Total Suspended Solids (TSS) is in **Appendix B**.

4.2. Handling continuous, in-line measurements

Test parameters such as salinity, flow rate, pressure, chlorophyll *a* (*in situ* fluorescence), and pH may be reported (periodically or continuously) by in-line sensors. For the purposes of this database, each measurement location—whether supplying a single sensor or a manifold supplying multiple sensors—is considered a unique Sample (e.g., "Treated Discharge In-line").

At a certain location, each unique sensor or reading type is considered an Analysis, even if there are multiple analyses collected by a single instrument. ***For continuous readings, the full data set is not required.*** However, a full (or subsampled or averaged) data set may be transmitted if the readings are:

- Indexed by Sample ID (in this case, e.g., a manifold or sensor array on the uptake treatment line), where each Sample ID is unique for the entire set of trials
- Timestamped, such that each Record (*row*) is unique
- Clearly defined, where the Fields of each record are unambiguous (and they include units either in the field name or in a separate Table Description*;* see Appendix B)

4.3. Delineating raw vs. summary data

Data compilers should report raw data such that each row represents a single, discrete value. A table with summary values—such that would be shown in the body of a report—is not necessary. Summary data may be challenging to incorporate into the main database, but if it is prepared in addition to the raw data tables, it should be clearly marked as summary data, and reported data should be described as "mean", "standard deviation", "minimum", "maximum", of the data range considered (see Box 3). Summary tables should also conform to the guidance above and avoid mixing text and numerical data.

The IL may also choose to report summary data, but should include at least the following information, as it relates to the range of readings:

- Size (number of readings/records included in the data range)
- Arithmetic Mean or Average
- Standard Deviation
- Minimum (value of the range)
- Maximum (value of the range)

Box 3. Raw Data and Summary Data

Raw Data represent a single measurement point, whereas **Summary Data** represent two or more measurement points, displaying the mean, median, minimum or other statistical description of the set of values. Examples of Raw and Summary Data are shown below (AR = Analytical Replicate).

Raw Data:

Analysis ID	Sample ID	AR	Reading	Value
pH-001	SAMP-001	1	1	7.6
pH-002	SAMP-001	1	2	7.8
pH-003	SAMP-001	1	3	7.7
pH-004	SAMP-001	2	1	7.9
pH-004	SAMP-001	2	2	7.9
pH-005	SAMP-001	2	3	7.8

Summary Data:

Sample ID	AR	Value	Description	Count
SAMP-001	1	7.7	Mean of Range	3
SAMP-001	2	7.9	Mean of Range	3

Summary Data should indicate the description of the statistical operation resulting in the values (in this case "Mean of Range") and the count of data included in the calculation. It may be possible to summarize the raw data in at different levels: In the case above, the mean summarized the Readings. It would be also possible to summarize the Analytical Replicates, as shown below:

Sample ID	Value	Description	Count
SAMP-001	7.8	Mean of Range	6

4.4. Avoid text in numerical data fields

Numerical fields must contain numbers only. Any text data, including units and inequality signs (“>”,”<”, “±”, etc.), will prevent numerical operations, such as calculating the mean values. Use a period (“.”) to mark decimals. Other symbols for grouping digits (e.g., commas “,”) are not needed, and they may be interpreted as text by some software packages. Units are expected to be the same throughout the field, but potentially, a separate column may show the units if they vary among records. Missing data may be intentional, but an empty cell may be interpreted as a zero (0) value. To prevent this, and to avoid adding text (such as “Null” or “NaN” for “Not a Number”) into a numerical field, data generators should add a column to identify (*flag*) issues. A standard set of “*flags*” may appropriate for all Analyses Tables, but whether standard or specific to a single analysis, the *flags* should be defined in the Table Descriptions.

Sample ID	Analytical Rep.	Reading	Flag Code
SMPL-001	1	2.4	0
SMPL-001	2		1
SMPL-001	3	3.1	0
SMPL-002	1	9820	0
SMPL-002	2	7895	0
SMPL-002	3		2

Figure 5. Excerpt of a hypothetical data table, shown here to indicate data flags for describing missing data. In this example, two values are missing (cells shaded in green). In some instances, users may interpret missing values as zero (0), rather than “null”. The “Flag Code” indicates the reason for missing data. Flag codes are defined by the data generator (e.g., the analyst or the TF). In this case, “0” indicates “no flag”; “1” would indicate value is below the detection limit; and “2” above the detection limit. See Appendix B for recommended Flag Codes.

4.5. Use an appropriate precision for values

Numerical values should be reported with the appropriate level of precision. This is most noticeable when calculation result (e.g., the mean number of organisms in a small aliquot) is multiplied by a correction factor, such that the resulting value contains more significant digits than the initial measurement. The level of precision should match the TF’s pre-defined levels of precision as described in their Standard Operating Procedures (SOPs), Test Quality Assurance Plan, or Quality Assurance Project Plans.

4.6. Link each Analysis Table to an analytical protocol

Analyses Tables should reference an SOP, so that the data Fields are related to processes and measurements in the SOP. This improves traceability of the data, allowing end-users to put the information in context. Revisions to the SOP would not likely require a separate Analysis Table,

but tracking the SOP version will also help, e.g., if the precision of the results changes within a table.

5. Benefits of Structured Data

The USCG, by statute, requires data are submitted with the TA application. The specific requirement that the data are submitted as structured, electronic data is not currently codified, but ILs and TFs should consider adopting this approach, as it will likely hasten the review process, reducing the time spent finding and interpreting data, especially given the variety of test reports submitted by multiple ILs. Reviewers will be able to rapidly arrange and summarize data, perform custom queries, and quality checks. Submitting structured, electronic reduces the IL's effort in generating text-based versions of raw or summary data, minimizing potential transcription errors or formatting issues. For these reasons, adoption of the approaches described herein benefit the ILs as well as the USCG.

6. Acknowledgements

This work was funded by the U.S. Coast Guard (USCG) Office of Operating and Environmental Standards (OES-3), Agreement 70Z023-20-KMMS03100, Task 2.8, although it does not represent the official policy of the USCG. We are grateful for guidance and programmatic support from Richard Everett, Regina Bergner, LT Carl Hendrickson (OES) and LTCDR Kirstin Haas (USCG Marine Safety Center). We appreciate those involved with the command and security reviews of this document.

APPENDIX A: RECOMMENDED TABLES

Note	For simplicity in this report, the following tables transpose the Field Names as Records (rows), but the Field Names are actually column headers. For example, the Table 1 has 8 (eight) columns. In some cases, tables may have 100s of rows. See Appendix C for actual data tables.

Tier 1: Main Table, BWMS Information, IL Assessment

Table 1: Main

Field #*	Field Name*	Data Type
1	Application ID	Index (Primary Key)
2	BWMS Name	Text
3	BWMS Manufacturer	Text
4	IL Name	Text
5	IL Address	Text
6	Application Date	Date/Time
7	Manufacturer Address	Text
8	Model	Text

* Same column headers used for all tables; headers omitted in subsequent tables

Table 2: BWMS Information

1	Application ID (key)	Index (Primary Key)
2	BWMS Treatment Type	Text
3	BWMS TRC	Text
4	Land-based test model	Text
5	Shipboard test model	Text
7	Salinity Range**	Text
8	Temperature Range**	Text
9	Hold time Range**	Text
10	Other limitations**	Text

** Manufacture claims, which may differ from IL's assessment

Table 3: IL Assessment

1	Application ID (key)	Index (Primary Key)
2	IL Summary Statement	Text
3	IL Salinity Range	Text
4	IL Temperature Range	Text
5	IL Hold Time Range	Text
6	IL Water characteristics limitations	Text
7	IL Other limitations	Text

Tier 2: Test Cycles

Table 4: Test Cycles

1	Application ID	Index: Link to Tier 1 Tables
2	Test Cycle ID	Index (Primary Key)
3	Test Type (e.g., Land-based, shipboard)	Text
4	Salinity Category	Text
5	Salinity Trial Number (x of y)	Text
6	Trial Start Date	Date/Time
7	Trial End Date	Date/Time
8	Treatment Uptake: Start	Date/Time
9	Treatment Uptake: End	Date/Time
10	Control Uptake: Start	Date/Time
11	Control Uptake: End	Date/Time
12	Treatment Discharge: Start	Date/Time
13	Treatment Discharge: End	Date/Time
14	Control Discharge: Start	Date/Time
15	Control Discharge: End	Date/Time
16	Trial Validity	Text
17	Trial Notes	Text

Tier 3: Samples

Table 5: Sample

1	Application ID	Index: Link to Tier 1 Tables
2	Test Cycle ID	Index: Link to Tier 2 Table
3	Sample ID	Index (Primary Key)
4	Sample Name	Text
5	Field Replicate	Number
6	Collection Start Time	Date/Time
7	Collection End Time	Date/Time
8	Sample Total Volume	Number
9	Sample Processing	Text
10	Sample Final Volume	Number
11	Sample Notes	Text

Tier 4: Analyses Tables

List of expected analyses

Table #	Parameter Name	Table Name
Organism Concentrations		
1	Organisms ≥50 µm	"GT50"
2	Organisms ≥10 and <50 µm	"10-50"
3	*Escherichia coli*	"Ecoli"
4	Intestinal enterococcus	"Entero"
5	*Vibrio cholerae* (O1 and O139)	"Vibrio"
Challenge water and test conditions		
6	Chlorophyll a	"Chl-a"
7	Total Suspended Solids	"TSS"
8	Dissolved Organic Matter (Carbon)	"DOM"
9	Particulate Organic Matter (Carbon)	"POM"
10	Mineral Matter	"MM"
11	Water Temperature	"Temp"
12	Water Salinity	"Salinity"
13	Water pH	"pH"
14	Heterotrophic bacteria	"Heterotrophs"
Test engineering parameters		
15	Ballast System Flow	"BSysFlow"
16	Ballast System Pressure	"BSysPressure"

General format of Analyses Tables (also see Fig. 4 and 5 in the main document)

Tables 6--22: Analyses (General Format)		
1	Application ID	Index: Link to Tier 1 Tables
2	Test Cycle ID	Index: Link to Tier 2 Table
3	Sample ID	Index: Link to Tier 3 Table
4	Analysis ID*	Index (Primary Key)
4	SOP Reference	Text
5	Analysis Start Date/Time	Date/Time
6	Analysis End Date/Time	Date/Time
7	Replicate**	Number
8	Subsample**	Number
9	Reading**	Number
10	Sample Volume ϕ	Number
11	Sample Volume Units ϕ	Text
12	Result	Number‡
13	Result Description	Text (e.g., mean, standard deviation, if summary data)
14	Result Units	Text
15	Flag	Text
16	Notes	Text
17	Analyst ID	Text
18	Analyst Notes	Text
19	QC Check	Text
20	QC Notes	

*Note: this may also be termed “parameter ID”; it is a unique ID for each record in the table

**See Box 2 in the main document for a disambiguation among these three terms

ϕ Multiple volumes may be necessary (e.g., aliquot volume, concentrated sample volume, etc.); add fields or sets of fields as needed.

‡Also, this may be a logical (binary) value (e.g., pass/fail)

APPENDIX B: EXAMPLE TABLE DESCRIPTION

Note: A table description is a short explanation of the information in a spreadsheet table. The description is included in a separate text file (.txt, .docx, .rtf), and it includes information on the fields (including data type and units, as necessary). A hypothetical example is below, although a particular table will have a different set of fields and information, as appropriate.

Data Table Name:	TSS
Type:	Analysis, raw data
Protocol Reference:	TSS SOP Rev.02
Field Descriptions:	Listed below

Field Name	Type	Units/Description
TSS ID	Index	Unique Index (Primary key); format: "TSS-000"
Test Cycle ID	Index	Index to Test Cycle ID; format: "TEST-000"
Sample ID	Index	Index to Sample ID; format: "SAMP-000"
Start Time	Date/Time	Time of analysis start; format "YYYY-MM-DD HH:mm:ss"
Volume	Number	Volume in used in analysis; units: L
Filter mass	Number	Initial filter mass; units: mg
Dry mass	Number	Filter + solids mass after drying; units: mg
TSS result	Number	Calculation: ([Dry mass]-[Filter mass])/[Volume]; units: mg L^{-1}
Data Flags	Category	See category codes below
Notes	Text	Indicate observations if necessary

Data Flags

Flag Code	Description
0	No Flag
1	Below limit of detection
2	Above limit of detection
3	Not performed
4	Other (see notes)

APPENDIX C: DATA ENTRY TEMPLATE

The following pages contain example tables with column headings and example data in the first row. Tables can be reproduced as spreadsheet or database tables.

Main Table

Application ID	BWMS Name	BWMS Manufacturer	IL Name	IL Address	Application Date	Manufacturer Address	BWMS Model
APID-001	Example BWMS	Examples LLC	Indy Labs, LLC	100 Independence Ave; Washington, DC 20371	6/1/2021	100 Example Street; Binary City, USA 10101	Ex-BWMS-Pro-250
…							
…							

BWMS Information Table

Application ID (key)	BWMS Treatment Type	BWMS TRC	Land-based test model	Shipboard test model	Salinity Range**	Temperature Range**	Hold time Range**	Other limitations**
APID-001	Filtration/UV	100-5000	Ex-BWMS-Pro-250	Ex-BWMS-Pro-500	0-4 PSU	0-35 °C	24 h	UVT >50%
…								
…								

IL Assessment Table

Application ID (key)	IL Summary Statement	IL Salinity Range	IL Temperature Range	IL Hold Time Range	IL Water characteristics limitations	IL Other limitations
APID-001	…	0-25 PSU	4-35°C	>24 h	UVT >50%	N/A
…						
…						

Sub-Laboratory Table

Application ID (key)	SL Name	Type of Sub-Laboratory	SL Address	SL Point of Contact
APID-001	SL Example Labs, LLC	Land-based TA testing	101 Lab St.; Binary City, 10101	Jane Doe; 000-101-1010
…				
…				

Test Cycle Table

Test Cycle ID	Application ID	Test Type (e.g., Land-based, shipboard)	Salinity Category	Salinity Trial Number (x of y)	Trial Start Date	Trial End Date	Treatment Uptake: Start	Treatment Uptake: End	Control Uptake: Start	Control Uptake: End	Control Discharge: Start	Control Discharge: End	Trial Validity	Trial Notes
TC-001	APID-001	Land-based	Freshwater	1 of 5	1/1/2001	1/2/2001	8:00	8:55	1/1/2001	1/2/2001	8:00	8:55	Valid	….
…														
…														

Sample Table

Sample ID	Test Cycle ID	Sample Type	Sample Name	Collection Method	Field Replicate	Collection Start Time	Collection End Time	Sample Total Volume	Sample Processing	Sample Final Volume	Sample Notes
S-001	TC-001	Viable/Vital Organisms	Control Uptake	Discrete	FR-01	8:00	8:05	20 L	None	20 L	…
…											
…											

Analysis Table (Example)

Application ID	Test Cycle ID	Sample ID	Analysis ID*	SOP Reference	Analysis Start Date/Time	Analysis End Date/Time	Replicate (AR)	Reading**	Sample Vol. ϕ	Sample Vol. Units ϕ	Result	Result Description	Result Units	Flag	Notes	Analyst ID	Analyst Notes	QC Check	QC Notes
APID-001	TC-001	S-001	pH-001	SOP.pH.Rev01	1/1/2001	1/1/2001	1	1	20	L	7.6	N/A Raw	-	0	…	…	…	…	…
…																			

Page intentionally left blank