DRAFT

Standard Operating Procedure

**DWM Water Quality Data Processing and Validation--- Laboratory Data**

CN 56.61

Contents

[I Applicability 2](#_Toc338326315)

[II Overview 2](#_Toc338326316)

[III Pre-Requirements to Initiate Procedures 2](#_Toc338326317)

[IV Assumptions 3](#_Toc338326318)

[V Validation Procedures for Laboratory Data 3](#_Toc338326319)

[VI. Reporting Rules for Preliminary (QC2) and Final (QC4) Laboratory Data Tables 9](#_Toc338326320)

[VII. Data Validation Reports 11](#_Toc338326321)

[APPENDIX A 12](#_Toc338326322)

[*Symbols and Qualifiers Used for DWM Laboratory Data* 12](#_Toc338326323)

[APPENDIX B 14](#_Toc338326324)

[*DWM Validation Criteria for Laboratory Data* 14](#_Toc338326325)

[APPENDIX C 16](#_Toc338326326)

[*Description of Visual Basic Scripts and Modules* 16](#_Toc338326327)

[EDD2LIMS.wsf 16](#_Toc338326328)

[LabDataXREF.wsf 17](#_Toc338326329)

[LabDataQA.wsf 19](#_Toc338326330)

[FinalizeLabData\_QC2.bas 23](#_Toc338326331)

[LabData\_Formulas.bas 27](#_Toc338326332)

[Function GetQual 27](#_Toc338326333)

[Function GetFinalDecision 28](#_Toc338326334)

[APPENDIX D 30](#_Toc338326335)

[*QC3-level review guidelines (LAB DATA)* 30](#_Toc338326336)

[APPENDIX E 31](#_Toc338326337)

[*Summary of Automated and Manual QC Checks on Laboratory Data (QC1-QC3)* 31](#_Toc338326338)

# I Applicability

These procedures apply to the following types of data generated by DWM from the 2006 monitoring year onward:

1. Laboratory data (e.g., microbiological, chemical, etc.)
2. Associated fieldsheet information (aka, fieldsheet “metadata”)

Each year, DWM monitoring staff collect water quality samples for laboratory analysis at selected sites (streams, rivers, lakes and ponds throughout the State). Resulting data are provided to DWM from the WES state laboratory, DWM’s lab and contract labs via the WES Laboratory Information Management System (LIMS) and Electronic Data Deliverables (EDDs). LIMS data are provided via monthly MS Access extracts. EDDs are provided using a standard DWM EDD format. While these data are “final” from a laboratory perspective, they undergo further review by DWM to evaluate data quality.

Note: Lab data collected historically up through the 2004 monitoring year were validated using other procedures (see CN 56.0, 56.1). Lab data collected in the 2005 monitoring year were validated using interim procedures, some of which formed the basis for the procedures described in this SOP. As a result, the format of the 2005 lab data output varies from that of the 2006-2010 output (for example).

# II Overview

These procedures are implemented by DWM’s data management team and are currently applied to individual year data sets. These procedures document DWM’s approach to organizing, processing, validating and finalizing water quality data from laboratories and from the fieldsheet record. While lab data may be qualified or censored during this process, the fieldsheet metadata are essentially provided “as-is” and tangentially to the primary data for use as needed. “Technical” QC checks on QC1-level laboratory data include field duplicate precision, field blank “hits”, frequency of QC sample collection per survey “trip”, violations of analytical holding times, non-adherence to field and/or laboratory methods, tidal influences, non-representativeness, analytical accuracy issues, and other factors affecting data quality. These QC checks are compared to current validation criteria, based on QAPP data quality objectives. Following technical QC review, QC2-level data are provided to project managers for QC3-level review. After any required edits are made to the data files, data become final (QC4-level).

# III Pre-Requirements to Initiate Procedures

The following information is required in order to initiate these procedures:

1. Raw LIMS extract (most recent version containing data being reviewed)
2. Raw EDDs from all labs
3. Combined LIMS+EDD Master file (following #1 and #2)
4. Proofed electronic fieldsheet files
5. Acceptance criteria and decision matrices for lab data ( accept vs. qualify vs. censor)
6. Current DWM reporting rules for data (rounding rules, significant figures)
7. Current DWM station file (geo-referencing)
8. Program Software: MS Access, MS Excel, Visual Basic

Database staff should be well-versed in MS EXCEL, Visual Basic (VB) coding and process documentation.

# IV Assumptions

These procedures have the following underlying assumptions:

1. Proofed fieldsheet files are accurate, until shown to be inaccurate. Any required changes to fieldsheet information is documented and transferred to other databases as needed.
2. LIMS (WES) and EDD (all other labs) data files are accurate until shown to be inaccurate (including that reported lab data have met internal laboratory QC acceptance limits)

# V Validation Procedures for Laboratory Data

1. When all surveys have been completed and the resulting laboratory data received for the data year-set, export final LIMS data from the WES laboratory (e.g., all 2011 data) in EXCEL format to w/dwm/data/raw data/LIMS. . Go to: [Y:\LIMSAnalyticalReports\BRP-DWM-WP\extracts.mdb](file:///Y:\LIMSAnalyticalReports\BRP-DWM-WP\extracts.mdb), open the year-dataset, select the entire file contents, then use “external data”/excel tool to export to w/dwm/data/raw data/*year*/LIMS.
2. Assemble all “other” (non-WES) lab data EDDs for the year-set of data undergoing validation and The EDDs can be from the DWM lab, contract labs, EPA lab, etc. (as applicable). Ensure all EDDs are available at w/dwm/data/rawdata (**see Appendix C: EDD2LIMS.wsf for description of required columns and column order in the master EDD file**).
3. When validation of lab data begins, repeat step 1 above and copy the LIMS file to W:\Data\laboratory\_QA\YYYY\LIMS (this is the working directory where changes to the LIMS data are made); name the file “LIMSYYYY\_MMDDYY.xlsx”, where YYYY is the data year being processed, and MMDDYY is the month-day-year that the LIMS data was extracted from the LIMS database
4. Compile all EDDs from “other” labs into one Master EDD file using the standard EDD template format (EXCEL). This is typically done “manually”. Save to: W:\Data\laboratory\_QA\YYYY\EDDs (this is the working directory where changes to the EDD data are made). Rename files to current date when new data are added. Before proceeding to Step 5, review master EDD file for date errors, dilution rate vs. result errors, dilution rate vs. MDL errors, etc. and make corrections.
5. Combine Master EDD file with the extracted LIMS DATA file to create one ALL-LAB-DATA file (EXCEL) via the following steps:
   1. Run VB script EDD2LIMS-V3.wsf to append master EDD file to LIMS file
   2. Combined file is automatically saved in W:\Data\laboratory\_QA\YYYY\LIMS-EDD Data\combined, where YYYY is the 4-digit year being processed.
6. Review ALL-LAB-DATA file to ensure consistency throughout for the following. Where corrections are needed, MAKE CHANGES TO THE INDIVIDUAL LIMS and/or EDD MASTER FILES (not the combined file):
   1. Erroneous differences due to syntax (e.g., remove spaces as needed so that same data are made equivalent) (this is done in the code)
   2. SampleID
      1. must start with the 7 digit OWMID (XX-XXXX, basinID+dash+4 digits) in order to be properly matched to the field sheet meta data OWMID
      2. any characters preceding the OWMID must be omitted, trailing characters are OK
      3. any character substituted for the dash must be corrected, e.g. some labs may implement zero instead of dash
      4. after the field sheet cross referencing is completed, the reviewer has another chance to catch any unmatched OWMIDs, correct the combined LIMS+EDD file and re-run the code
   3. Non-project data
      1. special projects or regional data with meta data not electronically entered: see 9c below (these data are automatically excluded by the code, OR can be deleted manually from the individual LIMS and/or EDD file)
      2. Analyte-specific data (e.g. fish toxics data): see 8a (these data are automatically excluded by the code when identified in the hold-time file)
   4. Correct analyte names that may not be unique for code matching purposes if different methods are used (e.g. Hardness may become Hardness-EDTA)
   5. Add missing values to ANALYSISTIME column
   6. Correct blank cells
7. Cross-reference the ALL-LAB-DATA file with the definitive and QC’d Fieldsheet metadata file via the following steps:
   1. With respect to data validation for any given year, check that each MetaData file has the same number of columns, that columns are in the same order, and that column headers are identical to 2011 file, (2005-2010 format was the basis for the process design, but changes made to the WQD necessitated changing the MetaData file format).
   2. Pre-review metadata file for missing records based on no water, ice-out, no access, stagnant, etc. where samples may not have been collected. Since these situations have no sample data, they are excluded in the process. NOTE: This was the case for 2005-2009 data validations. For 2010, missing data was accounted for in the final data in order to document no water conditions primarily, but also other istuations in which samples were not collected. For 2011, new field “NOT\_COLLECTED” was added to electronic field sheet to pre-identify missing samples (this step is no longer necessary)
   3. Run VB code LabDataXREF-V6.wsf (ALL-LAB-DATA file from step 5b is used as input along with the field sheet meta data file for the data year being processed)
   4. QC1 file is automatically saved to W:\Data\laboratory\_QA\YYYY\QC1 for data year YYYY.
   5. Perform 100% completeness check and reconcile any unmatched OWMIDs identified on the sheets called Unmatched FS OWMIDs and Unmatched Lab Cust Sample Num and make necessary changes to the field sheet metadata file, the LIMS extract and/or the EDD master file as needed. ~~the ALL-LAB DATA (combined) file~~. ). Also, disgard project data that will not undergo validation (e.g., non-DWM projects, regional BST, fish toxics, audit QC, etc.---these data can be moved to a “DO NOT USE” sheet within the EDD master or LIMS file, otherwise they will be excluded later in the process by default (dropped if no field sheet match).
   6. Perform additional error checks (see Meta Data Checks sheet in QC1 file) and correct in the source file(s) as necessary:
      1. Check that sample type is correct (should not be “in situ…”)
         1. Make corrections in WQD validation database after verifying sample type from paper field sheet
      2. Check that collection date reported by lab matches field sheet start (i.e. sampling) date
      3. Check that collection time reported by lab matches field sheet start (i.e. sampling) time
      4. Check that analysis date/time reported by lab occurs after field sheet sampling date/time
   7. When starting a new set of error checks, make a copy of the latest LIMS and/or EDD file, and make changes to the copy, so as to preserve the “history” of each set of code runs. The latest versions of the LIMS and EDD files will then be used to make the combined LIMS+EDD file and the QC1 file (field sheet meta data cross-referenced with the LIMS+EDD file). These steps should be repeated for as many times as necessary. No changes should be made to the combined EDD+LIMS file or the QC1 file to correct any errors (only save these files if you need to preserve a filter or add a column to track issues that have already been addressed)
   8. Repeat above steps as needed, based on subsequent re-runs. Note: changes made to either the LIMS EXTRACT or EDD master file will require repeating step 5 above to recombine the files (required input for XREF code in step 7b). ~~Therefore, make required changes to the combined LIMS/EDD file.~~
8. Following creation of QC1 file and error reconciliation, create Hold Time file and Units file for year being processed:
   1. Hold time file:
      1. Save a copy of the hold time criteria file with today’s date
      2. Open the hold time template file located in the QC1 folder, copy and paste the “hold time shell” worksheet to the hold time workbook on a new sheet labeled with the data year
      3. Populate the right hand columns with the final parameter name, final method, bottle group, hold times (most data can be copied from the prior year’s sheet or from earlier sheets if necessary; otherwise see latest version of W:\DWM\SOP\CN 365.0 – QAPP\_ DWM Monitoring Program\_2010-2014.doc for hold time values by year)
   2. Units file:
      1. Save a copy of the units file with today’s date
      2. Select UNIT column in the QC1 file and copy and paste to a new blank worksheet in the Units file
      3. Select newly pasted column, click on the “Data” tab on the MS Ribbon, select “Remove Duplicates” and select “ my data has headers” in the resultant pop-up box, click OK
      4. Add column for standardized units spellings called “UNITS\_DWM”
      5. Populate UNITS\_DWM column
      6. Rename worksheet to year being processed, e.g. “2007”
   3. Update the above worksheets/files as needed when new data are added to ALL-LAB-DATA file during the OWMID reconciliation process
9. Auto-validate ALL-LAB-DATA file for the “basic-4” QC issues (field blank contamination (b), field duplicate precision (d), frequency of QC samples (f) and analytical sample holding time violations(h)) using “ACCEPT”, “QUALIFY” and “CENSOR” options. The steps for this are as follows:
   1. Run VB code LabDataQA-V12.wsf (QC1 file from step 7c with most current date is used as input)
   2. QC2 file is automatically saved to W:\Data\laboratory\_QA\YYYY\QC2 for data year YYYY
   3. Review Excluded Data sheet to make sure all samples are included (should have already been addressed in step 7 above)
      1. Add any missing sample data and field sheet data
      2. Add newParam values (these are the “final” DWM reported analyte names) based on Hold Time Criteria file
      3. Add decisions for b, d, f, and h qualifiers
   4. Review automated decisions for b, d, f, and h qualifiers and make changes as needed to the “b Result”, “d Result”, “f Result”, and “h Result” (using Accept, Qualify, or Censor keywords)
10. “Manual” Review: Check/edit the draft ALL-LAB-DATA file (QC2 working file). Generally follow the sequence outlined below to edit automated QC decisions (Accept, Qualify, Censor). Note: Auto-QC decisions are conservative by design; best professional judgement during manual review is used to re-evaluate individual decisions that were based on set criteria. Justify all manual changes by adding comments to the QC2 comments column in the working file. Notebook notes should also be taken to document the review process.
    1. **Generate year-specific Reporting Rules file**, based on a copy of the auto-generated Hold Time file. Ensure consistency between the files, revising the Reporting Rules file as need (e.g., combining DWM and SMART for lar same analyte). “PARAM” (Hold Time file)=”LAB NAME (Reporting Rules file). “FINAL PARAM” (Hold Time file)=”DWM NAME (Reporting Rules file).
    2. If not already done via code, identify each test as deriving from either a **lake or river sample**: add “Lake” or “River” to the Survey Type column on the QC2 working sheet. Use project codes, lake-specific metadata and other available information (E.G., Secchi data) to make this determination. Classify impounded rivers sampled as lakes using DWM lakes fieldsheet as “Lakes” so the associated Secchi depth data will be reported in QC2/QC4 deliverables.
    3. **Secchi depth metadata** indicating Secchi>max depth (e) or Secchi time outside the approx. “solar noon” time (+/- 3 hours) of 10am-4pm (m). Insert qualifiers as needed in Secchi Qual column and populate Secchi Decision column. Also, Secchi depth metadata indicating that the Secchi disk hit bottom means that the reported Secchi result (if no “>” used) is an underestimate of the true condition, so the data must be censored or qualified (using BPJ), and annotated in Special Notes with qualifier “a” and explaining that the reported Secchi result is inaccurate (i.e., >reported value). Also, because Secchi Depth is repeated for each OWMID, apply Secchi-depth-“QUALIFY” to all analyte rows associated with a particular OWMID, BUT DO NOT APPLY ACTUAL QUALIFIERS MEANT FOR SECCHI TO OTHER PARAMETERS BY USING THE INDIVIDUAL QUALIFIER COLUMNS (i.e., use Secchi QUAL column only). Where found, “-9”s (i.e., 0) will be automatically changed to 0.0.
    4. **Populate a/e/j/m/p/r/t decision columns with “accept” for all data.** These will be adjusted as needed based on further review.
    5. **Transfer lab qualifiers (and review other laboratory internal QC data, lab comments, etc.)** that relate to data quality, such as lab QC outside acceptance limits for accuracy (a). In most cases, lab qualifiers are transferred directly to DWM data using the same or similar DWM qualifier symbol (as appropriate). Note: In some cases, lab’s decisions to report data may differ from DWM’s viewpoint on how these same data should be reported. For example, WES’ “M” qualifier is not carried forward by DWM. Verify no auto-code and that lab qual carrying forward is 100% manual (RC)
    6. **Apply “e” quals/censors.** Examples include: “Total” vs. “Partial” results for split samples (e.g., fecal coliform < E. coli (a fecal coliform organism), dissolved fraction > total, Secchi depth > max. depth, etc. Use BPJ for qualify vs. censor.
    7. **Apply “t” quals/censors.** Review location information for coastal samples to determine potential for tidally-influenced data. Also, associated attended probe conductivity data can be reviewed, if available, to estimate salinity of grab samples. Use BPJ for qualify vs. censor.
    8. **Review automated decisions/edit as needed** for each of the “BASIC-4”: hold time (h), field duplicate precision (d), frequency of QC samples (f), and field blank contamination (b). Make changes to appropriate “qualifier” result column by overwriting decision(s) for these fields to Accept, Qualify, or Censor (as appropriate). Best professional judgement is used to re-evaluate individual (conservative) auto-decisions. For example, a decision may be changed from censor to qualify for a result that just exceeds the censor threshold or that is based on a low-number-effect (high RPDs for low value numbers) in order to retain the data as qualified (usable). Also, for example, an automatic trip effect qualifying analyte-trip samples due to censoring of the associated QC sample could be manually changed to censoring ALL samples (depending on the nature of the problem).
    9. **Review field audits** **/apply decisions as needed** . Field audits may reveal issues related to data quality, such as adherence to sample collection procedures. Use BPJ for qualify vs. censor.
    10. **Review external laboratory audit results/apply decisions as needed** for potential accuracy (a), precision (d) or other problems. These are typically QC samples prepared by DWM and sent to the lab(s) or purchased Proficiency Testing (PT) studies sent to the lab by a 3rd party. Use BPJ for qualify vs. censor.
    11. **Remove any QC sample data** remaining that was not previously removed during QC1 data processing. Place removed data rows into the “excluded data” folder.
    12. **Review fieldsheet comments** **/apply decisions as needed** . Fieldsheet comments often indicate real or potential effects to data quality, such as tidal conditions (t), lack of adherence to method (m), sample preservation issues (p) and lack of representativeness (r). In addition to fieldsheet comments, brackish or saline conditions can be verified using attended probe conductivity data, if available. Use BPJ for qualify vs. censor.
    13. **Misc. communications (e.g., e-mail)** related to data quality.
    14. **Use of buckets for sample collection** is not standard protocol and results in method not followed (m).
    15. Review each qualifier column for **3 or more qualifiers**. This may indicate the need for censoring if the datum was not already (automatically) censored.
    16. **Correct newResult column as needed**, based on any field sheet comments (e.g. switched samples), obvious outliers or erroneous results, etc. If the result is “Missing”, it may be able to be recovered. Make same changes to the Result column
    17. **“New Units” field to be auto-populated later (leave blank)**
    18. **Review auto-trip-effect decisions** (censoring of QC sample(s) results in minimum qualification of associated trip samples). Random checks. Note: If manual changes are made to decisions for QC samples (e.g., from censor to qual), subsequent additional manual changes may be necessary to the associated trip samples (change to accept or leave as qual.)
    19. **Check “final decision” column for completeness** (all rows are auto-filled), including confirming that “No Result” decisions are based on missing data
    20. **Review all CENSORED data** for final acceptability.
11. Automatically finalize lab data file by applying significant figures and other reporting rules and create individual project files
    1. Make edits to the “read me” sheet in the file called “QC2 lab template.xlsx”
    2. Run VBA macro FinalizeLabData:
       1. Enable macros (if needed)
       2. Check that all “new” cells are not blank (i.e., all columns to the right of new “trip”). Also, remove any row shading, if any previously used to check row-mismatching.
       3. Go to Developer tab🡪Macros button🡪select FinalizeLabData and hit Run
       4. As macro runs, address error messages as they appear by aborting run and editing the working file as needed.
    3. Macro will:
       1. Apply reporting rules based on analyte/method specific criteria (see **W:\DWM\SOP\CN 056.5 - SOP\_DWM data reporting rules\_10-20-11.xls**) and no data (--), missing (\*\*), no water (^^), or censored (##) symbols to data using hold time and units files?
       2. Apply formatting
       3. Create “YYYY\_laboratory\_QC2 final all data” sheet based on “YYYY\_laboratory\_QC2 working” sheet, where YYYY is the data year; the final all data sheet is a formatted copy of the working sheet and includes all duplicate, field blank, river and lake tests
       4. Create “YYYY\_laboratory\_QC2 final lakes” sheet based on “YYYY\_laboratory\_QC2 final all data” sheet; this sheet includes all lake tests and duplicate tests (lower ID or uncensored test result or if both are censored, then lowest ID is used)
       5. Create “YYYY\_laboratory\_QC2 final rivers” sheet based on “YYYY\_laboratory\_QC2 final all data” sheet; this sheet includes all river tests and duplicate tests (lower ID or uncensored test result or if both are censored, then lowest ID is used)
       6. Create “YYYY\_laboratory\_QC2 final duplicates” sheet based on “YYYY\_laboratory\_QC2 final all data” sheet; this sheet includes all duplicate test pairs by analyte
       7. Create “YYYY\_laboratory\_QC2 final field blanks” sheet based on “YYYY\_laboratory\_QC2 final all data” sheet; this sheet includes all field blank tests
       8. Create “YYYY QC2 lab data summary” sheet with analyte test counts, number of surveys, and number of stations, summarized by project
       9. Create individual project files based on filtering of “Project Name” column on “YYYY\_laboratory\_QC2 final all data” sheet
          1. Adds “read me” tab imported from file = **QC2 lab template.xlsx**
          2. Adds individual data sheets: “Lab data- Lakes”, “Lab data - Rivers”, “Lab data – Duplicates”, “Lab data – Field Blanks”
       10. Create an additional “all projects” file
           1. Adds “read me” tab imported from file = **QC2 lab template.xlsx**
           2. Adds individual data sheets: “Lab data- Lakes”, “Lab data - Rivers”, “Lab data – Duplicates”, “Lab data – Field Blanks”
           3. Adds “YYYY QC2 lab data summary” sheet
    4. Review “YYYY\_laboratory\_QC2 final all data” sheet and “YYYY QC2 lab data summary” sheet, making changes to the working sheet or the macro if necessary (DO NOT make changes to the final sheets, these are meant to be based solely on the data as presented on the working sheet and these sheets get overwritten when the macro is re-run)
       1. Check for blank cells: the only columns that can have blanks are Mile Point, Secchi Depth qualifiers, Result Qualifiers, and QC OWMID
       2. Check for “rounding not applied” results
       3. Check for “<” or “>” that require manual rounding (in the working sheet) (e.g., <0.02000)
       4. Check that reporting rules have been applied to all data
       5. Check final data file for center alignment/left justified cells (all)
       6. Random QC checks (optional)
    5. **QC3 Review**: Notify PIs/QC3 reviewers via e-mail that QC2 data are ready for review. Provide guidance as appropriate. For QC3-level review guidelines, see Appendix D.
    6. **Changes based on QC3 review**: Compile staff QC3 (project-level) comments in a separate sub-folder under “QC3”. For required changes, open “YYYY\_laboratory\_QC3\_date” (this QC3 version of the working file is generated via a code module). Enable macros (if needed), then make manual changes to the working file as needed. Document changes in the “QC3 Comments” column. Once all changes have been made, make sure there are no blank cells in content-required fields, then run the macro “Finalize lab data” to generate the QC4 files. Resolve any errors and re-run macro if needed (overwriting previous files). Once the macro is finished, review files to make sure all looks copacetic. Copy QC4 final data tables to the definitive repository location for final data (e.g., w/dwm/data/QC4 data) and remove the QC2 files from the QC2 sub-folder (as the data have changed status).

# VI. Reporting Rules for Preliminary (QC2) and Final (QC4) Laboratory Data Tables

Each lab that DWM employs may have different reporting conventions and there may errors and/or inconsistencies in data presented in EDDs. As a result, DWM has established standard reporting rules for data deliverables, which ensure consistent data presentation with regard to issues such as significant figures, measurement units, formatting conventions and rounding rules. The majority of these rules are automated via VB code and/or Excel macros, as explained above.

1. Rounding to remove insignificant digits is performed as needed based on Standard Methods , 21st Edition guidance. Rounding procedures are also applied to data to account for appropriately decreasing resolution as concentration ranges increase.
2. Where dilutions have occurred at the lab, the corresponding increases in the method detection limit and upper quantitation limit are typically accounted for in the lab result, but not always (e.g., if a lab reports “ND” rather than “<10” for a 10X dilution with a MDL of 1; in this case the translation of “ND” equates to a “<10” result, not “<1”).
3. In cases where different lab methods are used for the same analyte resulting in different resolution capabilities , the data are reviewed to ensure that the application of standard reporting rules is appropriate.
4. Non-detect results shown as “ND” are transformed to less than the numeric method detection limit for the analyte in question (e.g., <.005).
5. By convention for field duplicate results, the lower OWMID# is taken as the primary result for reporting and the higher OWMID# is made the duplicate result, which is only presented in the context of quality control (precision).
6. For bacteria results of “TNTC” (too numerous to count), the result is transformed to greater than the upper quantitation limit (UQL) (e.g., >1600 CFU/100 mls.) defined in the original EDD.
7. Non-censored but qualified results are shown with the qualifier symbol immediately adjacent (to the right in a separate column) to the numeric result.
8. The data elements for final data tables include the following metadata fields:

| RIVERS |  | LAKES *(****lake-specific in bold****)* |
| --- | --- | --- |
| Project |  | Project |
| Watershed |  | Watershed |
| SARIS\_PALIS\_CAMIS |  | SARIS\_PALIS\_CAMIS |
| Water Body |  | Water Body |
| Unique ID |  | Unique ID |
| Station ID |  | Station ID |
| Station Description |  | Station Description |
| Mile Point |  | Mile Point |
| Latitude (dec. degrees) |  | Latitude (dec. degrees) |
| Longitude (dec. degrees) |  | Longitude (dec. degrees) |
| Field Sheet Log |  | Field Sheet Log |
| Survey Type |  | Survey Type |
| Sample OWMID |  | Sample OWMID |
| QC OWMID |  | QC OWMID |
| QC Type |  | QC Type |
| Sample Date |  | FS Start Date |
| Sample Time |  | FS Start Time |
| Flow Condition |  | **Lake level** |
|  |  | **Max Depth** |
|  |  | **Sample Depth** |
|  |  | **Relative Sample Depth** |
|  |  | **Secchi depth** |
|  |  | **Secchi depth Qualifiers** |
| Analyte |  | Analyte |
| UNITS |  | UNITS |
| Result |  | Result |
| Result Qualifiers |  | Result Qualifier |
| Analysis Method |  | Analysis Method |
| Special Notes (e.g., flow composites) |  | Special Notes (e.g., flow composites) |

1. All analytes are shown in one data element field (column): DWM Characteristic Code. This format differs from probe data tables where each individual analyte is provided in separate columns.
2. Sample depths for river samples are not provided since most of these samples are collected just beneath the water surface (approx. 0.2 meters). Fieldsheet comments indicating collection at depths significantly different from the assumed depth should be noted in the Special Notes field.
3. Blank cells in the final data tables are generally avoided by using “--“ to indicate no data. As stated above, blank cells should only show for the following fields: Mile Point, Secchi Depth qualifiers, Result Qualifiers, and QC OWMID.
4. Errata to QC4 data files. While every attempt is made to avoid changes to QC4 data, changes can be made to QC4 files if needed by issuing formal errata describing the change(s) and generating revised QC4 file(s). Errata documentation should include the following:
5. Errata: description of change, name of changed file, file locations where update has taken place, …..
6. New file(s)
7. Reference Files:

# VII. Data Validation Reports

For each year-set of data that has been validated, a brief summary report is produced by DWM QA Analyst to document what took place. These reports are given Control Numbers (CN) and placed on DWM’s network drive as a reference: [w\dwm\SOP](file:///\\dep.govt.state.ma.us\enterprise\Worcester-Workgroup\DWM\SOP).

# APPENDIX A

# *Symbols and Qualifiers Used for DWM Laboratory Data*

The following data qualifiers or symbols are used in the MassDEP/DWM data validation process for qualified and censored LABORATORY DATA. Decisions regarding censoring vs. qualification for specific, problematic data are made based on a thorough review of all pertinent information related to the data.

**DWM General Symbols:**

“ ## ” = Censored data (i.e., data that has been discarded for some reason; check qualifier symbol for cause(s)).

“ \*\* ” = Missing data (i.e., data that should have been reported, but were not for any reason other than no water).

“ -- ” = No data (i.e., data not collected nor intended)

“ ^^ “ = No water (i.e., a special case of missing data due to dry/no water conditions)

[ ] = A result reported inside brackets has been “censored”, but is shown for informational purposes (e.g., high blank results).

**DWM Laboratory Data Qualifiers:**

“ a ” = accuracy as estimated at WES Lab via matrix spikes, PT sample recoveries, internal check standards and lab-fortified blanks did not meet project data quality objectives identified for program or in QAPP.

“ b ” = blank Contamination in lab reagant blanks and/or field blank samples (indicating possible bias high and false positives).

“ d ” = precision of field duplicates (as RPD) did not meet project data quality objectives identified for program or in QAPP. Batched samples may also be affected.

“ e ” = not theoretically possible. Specifically, used for bacteria data where colonies per unit volume for e-coli bacteria > fecal coliform bacteria, for lake Secchi and station depth data where a specific Secchi depth is greater than the reported station depth, and for other incongruous or conflicting results.

“ f ” = frequency of quality control duplicates did not meet data quality objectives identified for program or in QAPP.

“ h ” = holding time violation (usually indicating possible bias low)

“ j ” = ‘estimated’ value; can be used for lab-related issues where certain lab QC criteria are not met and re-testing is not possible (as identified by the WES lab only). Also used to report sample data where the sample concentration is less than the ‘reporting’ limit or RDL and greater than the method detection limit or MDL (mdl< x <rdl). Also used to note where values have been reported at levels less than the mdl. Also used for estimated ranges based on known metadata.

“ m ” = method SOP not followed, only partially implemented or not implemented at all, due to complications with sample matrix (eg. sediment in sample, floc formation), lab error (eg. cross-contamination between samples), additional steps taken by the lab to deal with matrix complications, lost/unanalyzed samples, use of expired reagents and missing data.

“ p ” = samples not preserved per SOP or analytical method requirements.

“ r ” = data may not be representative due to circumstances and/or conditions at the time of sampling, including the possibility of “outlier” data.

“ s ” = field sheet recorded data were used to accept data (i.e., not data electronically recorded in a data logger or in cases where data logging is not possible (e.g., single-probes)).

“ t ” = tidal influence likely (not indicative of freshwater flow)

# APPENDIX B

# *DWM Validation Criteria for Laboratory Data*

|  |  |  |  |
| --- | --- | --- | --- |
| **FIELD BLANKS** | | | |
| <MDL | =MDL | >MDL | >>MDL |
| accept | accept | b (qualify) | b (censor, BPJ and/or >2X RDL) |
| NOTES: | | | |
| 1) If "ND", then OK (show as "<*MDL value*" (if available), then "<*RL value*"; if no MDL or RL, flag for error check | | | |
| 2) If <X, OK | | | |
| 3) "Trip Effect": If field blank censored, qualify all other same-analyte samples for that survey trip | | | |

|  |  |  |
| --- | --- | --- |
| **QC FREQUENCY** | | |
| FB and FD (both collected) | FB or FD (only one type collected) | No field QC collected on crew trip |
| accept | accept | f (qualify) |
| NOTES: | | |
| 1) In general due to most programmatic QC samples meeting acceptance limits, censoring for lack of QC samples for any given trip is avoided; use BPJ in cases of chronic disregard of QC sampling to decide if censoring for such an extreme case is warranted. | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **HOLDING TIME** | | | |
| <HT | =HT (based on one decimal place) | >HT | >>HT |
| accept | accept | h (qualify) | h (censor, BPJ and/or >HT+10%) |
| NOTES: | | | |
| 1) Use hold time lookup table for specific year being validated | | | |
| 2) If no analysis date AND time, at least qualify | | | |
| 3) If no analysis time, interpolate to add times using 5 min increments (if start and/or end known), OR add 5 min sequentially onto last time provided for each sample with unknown time within a batch (to master LIMS/EDD file). Adding 1/2 to 1 hour for lunch not needed (not that critical) | | | |
| 4) All times using 24 hour time format | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **FIELD DUPS (bacteria)** | | | | | | | | | | | | | | | |
| </= 50 CFU/MPN | | | | 51-500 CFU/MPN | | | | 501-5000 CFU/MPN | | | | >5000 CFU/MPN | | | |
| <30% RPD | =30% RPD | >30% RPD | >>30% RPD | <20% RPD | =20% RPD | >20% RPD | >>20% RPD | <10% RPD | =10% RPD | >10% RPD | >>10% RPD | <5% RPD | =5% RPD | >5% RPD | >>5% RPD |
| accept | accept | d (qualify) | d (censor, BPJ and/or 2X DQO (60%)) | accept | accept | d (qualify) | d (censor, BPJ and/or 2X DQO (40%)) | accept | accept | d (qualify) | d (censor, BPJ and/or 2X DQO (20%)) | accept | accept | d (qualify) | d (censor, BPJ and/or 2X DQO (10%)) |

|  |  |  |  |
| --- | --- | --- | --- |
| **FIELD DUPS** | | | |
| <20% RPD | =20% RPD | >20% RPD | >>20% RPD |
| accept | accept | d (qualify) | d (censor, BPJ and/or 2X DQO (40%)) |
| NOTES: | | | |
| 1) For "<X" results, use "X" in RPD calculation (if "0" result, use MDL (or minimal if no MDL provided) value for calculation and result | | | |
| 2) "Trip Effect": If field duplicates censored, qualify all other same-analyte samples for that survey trip. BPJ used to determine if trip samples should also be censored. | | | |
| 3) Manually review all qualify and censor auto-decisions for low number effect (LNE); edit decisions using BPJ as necessary | | | |
| 4) For bacteria data, a separate table is used and RPDs are calculated using log-10-transformed data. | | | |
| **FIELD DUPS (Secchi)** | | | |
| <10% RPD | =10% RPD | >10% RPD | >>10% RPD |
| accept | accept | d (qualify, BPJ) | d (censor, BPJ) |
| NOTES: | | | |
| 1. Manually review field sheets for duplicate Secchi readings, manually calculate RPDs, and apply qualifiers as needed. 2. Applicable when duplicate Secchi readings are required by SOP and available for evaluation | | | |
|  | | | |

# APPENDIX C

# *Description of Visual Basic Scripts and Modules*

## EDD2LIMS.wsf

* General purpose:
  + To compile all laboratory data into one file, joining a “master” laboratory EDD data file to a LIMS Excel file for the data year being processed
* Required input files:
  + Excel file exported from latest update of LIMS database “extracts.mdb” (located at : [Y:\LIMSAnalyticalReports\BRP-DWM-WP\extracts.mdb](file:///Y:\LIMSAnalyticalReports\BRP-DWM-WP\extracts.mdb)) for data year being processed (placed in directory: W:\DWM\Data\laboratory\_QA\YYYY\LIMS-EDD Data\LIMS)
  + Compiled master EDD file for data year being processed (placed in directory: W:\DWM\Data\laboratory\_QA\YYYY\LIMS-EDD Data\EDDs)

**LIMS file column order**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** | **K** | **L** | **M** | **N** | **O** | **P** | **Q** | **R** | **S** |
| **SAMPLE NUMBER** | **CUSTOMER**  **SAMPLE NUMBER** | **PARAM** | **RESULT** | **QUALIFIER** | **UNITS** | **DETECT LIMIT** | **Report DL** | **METHOD** | **ANALYSIS TIME** | **CUSTOMER ID** | **PROJECT ID** | **COLLECT DATE** | **COLLECT TIME** | **SITE** | **SITE Locator** | **Comments** | **APPROVED BY** | **QC1 COMMENTS\*** |

\*Column must be manually added to LIMS file prior to running EDD2LIMS code

**EDD file column order**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** | **K** | **L** | **M** | **N** | **O** | **P** | **Q** | **R** | **S** | **T** |
| **LAB ID** | **LabS Num** | **Field Samp Num** | **Analyte / Characteristic** | **Sample Fraction\*** | **Result** | **Lab Qual** | **Res Comm** | **Units** | **MDL** | **RDL** | **UQL** | **Analytical Method** | **Anal Date** | **Anal Time** | **Site Locator** | **Collect Date** | **Collect Time** | **Project/other** | **QC1 Comments** |

\*Column not included

* Output files:
  + **YYYY\_LIMS\_EDDs\_mm-dd-yyyy.xlsx**, where YYYY is the data year being processed, mm, dd, and yyyy are the month, day, and year the combine file was created (combined file is placed in directory: W:\DWM\Data\YYYY\LIMS-EDD Data\combined)

**Combined file column order**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LIMS Column** |  | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** |  | **I** | **J** | **J** | **K** | **L** | **M** | **N** | **O** | **P** | **Q** | **R** |  | **S** |
| **EDD Column** | **A** | **B** | **C** | **D** | **F** | **G** | **I** | **J** | **K** | **L** | **M** | **N** | **O** |  | **S** | **Q** | **R** |  | **P** | **H** |  |  | **T** |
| **Combined**  **Column** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** | **K** | **L** | **M** | **N** | **O** | **P** | **Q** | **R** | **S** | **T** | **U** | **V** | **W** |
|  | **LAB ID** | **SAMPLE**  **NUMBER** | **CUSTOMER**  **SAMPLE NUMBER** | **PARAM** | **RESULT** | **LAB QUALIFIER** | **UNITS** | **MDL** | **RDL** | **UQL** | **METHOD** | **ANALYSIS DATE** | **ANALYSIS TIME** | **CUSTOMER ID** | **PROJECT ID** | **COLLECT DATE** | **COLLECT TIME** | **SITE** | **SITE Locator** | **Comments** | **APPROVED BY** | **Dummy Column** | **QC1 COMMENTS** |

* Approximate run time:
  + 3-5 minutes

Procedure:

1. Enter the 4-digit data year to process when prompted by the code
2. Navigate to the master EDD file when prompted by the code, click Open
3. Navigate to the LIMS file when prompted by the code, click Open
4. check for directory to hold combined LIMS and EDD data file
5. define column map for EDD to LIMS (headers in combined file are identical to LIMS file)
6. open LIMS file
   1. rename combined data worksheet to “YYYY LIMS+EDDs”
   2. add worksheet called “File Process Info”for recording macro name, input file names and paths, date script was run (starting and ending times)
   3. add column on data sheet for AnalysisDate before AnalysisTime column (date and time are split from LIMS AnalysisTime column; EDD file already has analysis date and time in separate columns)
   4. add column on data sheet for UQL (to be added from EDD file)
7. open EDD file
   1. copy EDD data to appropriate columns on combined data worksheet (as defined in step 5 above)
8. standardize date and time formats for columns: AnalysisDate, AnalysisTime, CollectDate, CollectTime

## LabDataXREF.wsf

* General purpose:
  + To “cross-reference” or match each laboratory analyte sample ID with the associated field sheet meta data, using the OWMID
* Required input files:
  + Combined LIMS and EDDs laboratory data file: **YYYY\_LIMS\_EDDs\_mm-dd-yyyy.xlsx**
  + Field sheet meta data file for year being processed: **MetaDataYYYY.xlsx** (located in directory: W:\DWM\Data\RawData\YYYY, where YYYY is the data year being processed)
* Output files:
  + **YYYY\_laboratory\_QC1\_mm-dd-yyyy.xlsx**, where YYYY is the data year being processed, and mm, dd, and yyyy are the month, day, and year the QC1 file was created (QC1 file is placed in directory: W:\DWM\Data\laboratory\_QA\YYYY\QC1)
* Approximate run time:
  + 2 hours
  + Speed depends on network traffic, personal computer used, input file sizes

Procedure:

1. Enter the 4-digit data year to process when prompted by the code
2. Navigate to the field sheet meta data file when prompted by the code, click Open
3. Navigate to the combined LIMS + EDD file when prompted by the code, click Open
4. Navigate to the folder for storing the QC1 file
5. Create QC1 file from copy of LIMS + EDD file
6. Add new sheets to QC1 file
   1. File Process Info: information to be stored includes when vbscript was run, start and end times for processing, location of input files
   2. Unmatched FS OWMIDs: stores all OWMIDs from the field sheet meta data file that have no matching record in lab data column “Customer Sample Number”
   3. Unmatched Lab Sample Numbers: stores all Customer Sample Numbers (regardless of whether the number appears to start with an OWMID) without a match to the field sheet file
   4. Unmatched Lab OWMIDs: stores all Customer Sample Numbers by analyte that appear to start with an OWMID without a match to the field sheet file
   5. Unique Unmatched Lab OWMIDs: stores all unique occurrences of Customer Sample Numbers without a match to the field sheet file
7. Make a copy of field sheet meta file and open the copy
8. Sort OWMID column in the meta data file so that blanks are at the end of the column; copy header from field sheet file
9. Insert blank columns to left of lab data in the QC1 file, for later insertion of meta data
10. write header from meta data file to LIMS + EDD file
11. write header from QC1 sheet to Unmatched Lab Sample Numbers and Unmatched Lab OWMIDs sheets
12. Find the Customer Sample Number column on the QC1 sheet and sort in ascending order
13. get the first OWMID from the meta data file, look for all matches in the LIMS+EDD file in the CustomerSampleNumber column (with and without dash)
    1. if match is found, insert the field sheet meta data into the matched rows in the LIMS+EDD file
    2. if match is not found, write field sheet meta data to Unmatched FS OWMID sheet
14. locate date and time columns inserted from field sheet meta data file (StartDate, StartTime, Date, FSTime) and correct the formats
15. add new columns to QC1 sheet at end of used range
    1. Trip: to identify trips (first 7 characters of FSLOG)
    2. ID like OWMID: to identify CustomerSampleNumbers with patterns like OWMID
    3. CustSamp#Prefix: to identify prefixes in CustomerSampleNumber column (first 2 characters)
16. loop through CustomerSampleNumber to identify IDs with OWMID-like pattern at the beginning (with and without the dash)
    1. if match is found, place Y in “ID like OWMID” column
       1. if OWMID is blank indicating no field sheet meta data match, copy row to Unmatched Lab OWMIDs sheet
    2. if match not found, place N in “ID like OWMID” column
    3. copy all unmatched lab numbers (identified by OWMID = blank) to Unmatched Lab Sample Numbers sheet (regardless of whether the pattern is like an OWMID)
17. find unique unmatched lab OWMIDS
    1. copy results on Unmatched Lab OWMIDs sheet to Unique Unmatched Lab OWMIDs sheet
    2. find unique unmatched chemistry OWMIDs by removing duplicates
    3. extract first two characters from OWMID and add to Prefix column (to enable easy filter of CustomerSampleNumber by “basin”)
18. add CustSamp#prefix and Trip to QC1 sheet
19. color code column headings (field sheet columns = tan, lab sample data = blue, extra columns (“ID like OWMID” and “CustSamp#Prefix”) = gray)

## LabDataQA.wsf

* General purpose:
  + To perform automated QA decisions for hold-time violations (h), field blank contamination (b), field duplicate precision (d), and frequency of QC samples (f)
* Required input files:
  + Field sheet cross-referenced laboratory data file: **YYYY\_laboratory\_QC1\_mm-dd-yyyy.xlsx**
  + Hold time file: **Hold Time Criteria.xlsx** (located in W:\DWM\Data\laboratory\_QA\reference\_files)
  + VBA modules
    - FinalizeLabData\_QC2.bas
    - LabData\_GetFinalDecision.bas
* Output files:
  + **YYYY\_laboratory\_QC2\_mm-dd-yyyy.xlsx**, where YYYY is the data year being processed, and mm, dd, and yyyy are the month, day, and year the QC2 file was created (QC2 file is placed in directory: W:\DWM\Data\laboratory\_QA\YYYY\QC2)
* Approximate run time: 1 hour

Procedure:

1. User enters data year to process at prompt
2. User navigates to QC1 (field sheet cross-referenced) file when prompted, clicks Open
3. Check for presence of hold time file, station ID/waterbody file, and VBA macro files
4. User navigates to the location for storing the QC2 file, clicks OK
5. Copy QC1 file to QC2 location and rename to final output file name
6. Rename QC1 data sheet to “YYYY\_laboratory\_QC2 init”, delete unneeded sheets that were used to track unmatched sample numbers in QC1 file, create “Excluded data” sheet to store IDs associated with analytes not included in QA process
7. Create file process info sheet to store start and end time of program run, file names
8. Create sheet to store list for QC result pull-downs
9. Define named range for QCResult list
10. Import VBA modules
    1. Add LabData\_GetFinalDecision module (formulas)
    2. Add FinalizeLabData module (code to create final sheets)
11. Check for data to exclude from QC2 sheet and cut and paste to Excluded Data sheet:
    1. blank OWMIDs: those Customer Sample Numbers without a match to the field sheet meta data
    2. blank Customer Sample Numbers (also have no match to the field sheet meta data)
    3. Analytes not included (analytes identified from Exclude column in hold time file)
12. Insert new columns on QC2 init sheet for QA checks, QC results, new fields, etc.
13. Start QA calcs
    1. Copy lab result from “Result” column to “newResult” column
       1. If Result is blank, newResult = Missing
       2. Check for “<” at start of Result and get value to the right
          1. If right side is numeric, and
             1. doesn’t equal MDL or RDL value, then newResult = “Check Result”
             2. equals MDL or RDL value, then newResult = Result as reported (i.e. <0.150 remains the same)
          2. If right side is not numeric, and
             1. Equals “MDL” text, then newResult = <MDL value (e.g., Result = “<MDL”, newresult = “<0.15”)
             2. Equals “RDL” text, then newResult = <RDL value (e.g., Result = “<RDL”, newresult = “<1”)
             3. If MDL or RDL value is blank and result = <MDL or <RDL, then newResult = Missing
       3. Check for “>” at start of Result and get value to the right
          1. If right side is numeric, then use newResult = Result as reported
          2. If right side is not numeric, and
             1. Equals “UQL” text, then newResult = >UQL value (e.g., Result = “>UQL”, newResult = “>1600”)
             2. Doesn’t equal “UQL” text, then newResult = Missing
       4. Check for text in Result (not starting with “<” or “>”)
          1. If Result = “\*\*”, then newResult = “Missing”
          2. If Result = “ND”, and
             1. MDL is not blank, then newResult is < MDL value
             2. If MDL is blank and RDL is not blank, then newResult is < RDL value
          3. If Result = “TNTC”, then newresult = “>”UQL value (unless UQL is blank, then newResult = “Missing”)
          4. all other values get Missing for newResult
       5. if Result = “0” exactly, then newResult = “Check Result”
    2. Hold-time calculations
       1. Populate newFStime column
          1. Combine field sheet date with either StartTime or CollectTime when available
          2. If times or dates are missing (indicated by -1), then newFSTime equals “No Time”
       2. calculate hold time as newFSTime minus (AnalysisDate + AnalysisTime)
          1. if only dates are available, use dates
       3. lookup max hold time in “Hold Time Criteria.xlsx” file based on “Param” column in QC2 file; populate “Max HoldTime (days)” column in QC2 file
          1. if Param value in QC2 file is blank or if hold time value is blank, then “Max Hold Time” value = “No Value Found”
       4. Calculate hold time exceedance as (hold time calculation/max hold time)-1, where hold time calculation is from step 13.b.ii. and where max hold time is from step 13.b.iii. (inserted into column labeled “HoldTime Exceedance”)
          1. If no hold time calculation or no max hold time, then hold time exceedance = “Can Not Calculate”
          2. If Dates only were used to calculate hold time and max hold time is < 1 day, then Exceedance value = "Dates Used/Max HT < 1 Day"
          3. Based on QC criteria table for hold time exceedances, apply Accept, Qualify, or Censor to “h Result” column
    3. Add newParam, Major Analyte Category, and Bottle Group column values based on looking up “Param” column values in QC2 file and matching to hold time file
    4. Field Blank calculations
       1. For each row where QCTypeName is “Field Blank”, get newResult value and MDL value
       2. Calculate the DLRatio as newResult/MDL and populate field called “QA-Blanks (Result/MDL)”
          1. If MDL is blank then use RDL instead (if both MDL and RDL are blank, then ratio = “Can Not Calculate”
          2. If newResult has a “<” symbol, then
             1. Ratio is automatically flagged as “Non Detect”
       3. Based on QC criteria table for field blanks, apply “Accept”, “Qualify”, or “Censor” to “b Result” column
          1. Note: if ratio is “Non Detect”, the field blank result is Accept
          2. Non “Field Blank” rows get “N/A” in the “b Result” column for not applicable
    5. Duplicate Calculations
       1. For each row where QCTypeName = “Duplicate”, get sample OWMID and QC OWMID, analyte and Major Analyte Category
       2. For each analyte, get sample OWMID and QC OWMID result values
       3. Calculate relative percent difference (RPD) between each duplicate pair for each analyte and populate the “QA-DUPES (RPD)” column for each duplicate pair
          1. For all analytes except those in the Bacteria Major Analyte Category: RPD = (2\*(ABS((Sample Result – QC Result)/(Sample Result + QC Result))))\*100
          2. For analytes in the Bacteria Major Analyte Category, the Sample and QC Results are first log-transformed before calculating RPD
          3. If Sample Result or QC Result is “Missing” then RPD equals “Can Not Calculate”
       4. Count the number of duplicates found for each analyte and OWMID and populate the “DUPES” count column (to aid in determining whether duplicates are missing)
       5. Based on QC criteria table for duplicates, apply “Accept”, “Qualify”, or “Censor” to “d Result” column
          1. Non “Duplicate” rows get “N/A” in the “d Result” column for not applicable
          2. a separate QC criteria table is used for Bacteria analytes
    6. QAQC frequency check
       1. For each trip # and Bottle Group in the QC2 file, count the number of duplicate rows, field blank rows, and “other” (i.e. Routine Sample) rows
       2. Populate the “QA-QCfreq” column
          1. If number of duplicates = 0, then “NO BLANKS”
          2. If number of field blanks = 0, then “NO DUPES”
          3. If number of field blanks = 0 and number of duplicates = 0, then “NO QC”
       3. Based on the QC criteria table for QC frequency, apply “Accept” or “Qualify” to the “f Result” column
    7. Check if any field blank and duplicates are censored for each trip/analyte combination
       1. For each trip and each analyte combination
          1. check for any field blank rows that were censored (“b Result”)
          2. check for any duplicate rows (pairs will have the same result) that were censored (“d Result”)
       2. if any field blank result = “Censor” then,
          1. apply “Qualify” to all other trip/analyte rows in the “b Result” column
          2. if multiple field blank rows exist, leave those results as is
          3. apply "Trip/analyte blank censored" to the "QA-Blanks (Result/MDL)" column
       3. if any duplicate pair result = “Censor”, then
          1. apply “Qualify” to all other trip/analyte rows in the “d Result” column
          2. if multiple duplicate rows exist, leave those results as is
          3. apply "Trip/analyte duplicate censored" to the "QA-DUPES (RPD)" column
14. apply pull-down list to each QC result column (to allow manual additions/changes to QC Results during QC2 review): h, b, d, f, a, e, m, p, r, t, Secchi Result columns
15. sort columns by OWMID and parameter and apply formatting
16. autofilter the cells
17. copy the QC2 init sheet to the QC2 working sheet (this sheet is where QC2 review edits and comments are made)
18. insert formulas for determining final QC decision (“Final Decision” column) and combined qualifiers (“Combined Qualifiers” column) as changes/additions are made to the QC Result columns
19. insert worksheet change code onto QC2 working sheet to automatically highlight cells in yellow when values are changed
20. record ending time for code on “File Info Process” sheet

## FinalizeLabData\_QC2.bas

* General purpose:
  + To finalize laboratory data, including application of reporting rules, formatting, creation of project files
* Required input:
  + Files:
    - QC2 file: **YYYY\_laboratory\_QC2\_mm-dd-yyyy.xlsx,** where YYYY is the data year being processed, and mm, dd, and yyyy are the month, day, and year the QC2 file was created (QC2 file is placed in directory: W:\DWM\Data\laboratory\_QA\YYYY\QC2)
    - Units file: **Units.xlsx** (file location: W:\DWM\Data\laboratory\_QA\reference\_files)
    - Methods file: **Methods.xlsx** (file location: W:\DWM\Data\laboratory\_QA\reference\_files)
    - Template file for inserting read me tab in project files: **QC2 lab template.xlsx** (file location: W:\DWM\Data\laboratory\_QA\templates\)
    - Hold time file: **Hold Time Criteria.xlsx** (file location: W:\DWM\Data\laboratory\_QA\reference\_files)
    - Station ID/Water Body information file: **rptStaidUnique\_ID\_02-24-11\_LatLongBasin.xlsx** (file location: W:\DWM\Data\laboratory\_QA\reference\_files)
  + QC2 File Worksheets:
    - YYYY\_laboratory\_QC2 working
* Output:
  + Files:
    - Project folder (called “YYYY\_lab\_QC2\_Project\_Files\_mm-dd-yyyy”, where mm is month, dd is day, and yyyy is year that QC2 file was generated using **LabDataQA.wsf**; YYYY is the data year) containing individual Excel project files (named “*Project*\_YYYY\_ labdata \_QC2.xlsx”, where Project is the project name and YYYY is the data year)
  + QC2 File Worksheets:
    - YYYY\_lab\_QC2 final all data
    - YYYY\_lab\_QC2 final lakes
    - YYYY\_lab\_QC2 final rivers
    - YYYY\_lab\_QC2 final field blanks
    - YYYY\_lab\_QC2 final duplicates
    - YYYY lab data summary
* Approximate run time: 2-3 minutes

Procedure:

1. check for presence of directory and files
2. announce to the user what macro will do
3. define columns on final sheet and working sheet
4. check for invalid or blank entries in specific 'working' sheet columns: Survey Type, Final Decision, Secchi Decision, newResult
5. check for presence of project files before continuing, give user opportunity to delete
6. create final worksheets (all data, lakes, rivers, field blanks, field duplicates, lab data summary); if old versions are found then delete these first)
7. check for blank values in newUnits, newMethod, and Major Analyte Category
   1. if blanks are found, automatically overwrite all values in these columns
   2. if no blanks are found, ask user whether the fields should be updated
8. copy field sheet and lab data columns from “working” to “final all data” sheet
9. insert headers on “final all data” sheet
10. add station and water body information based on matching Unique\_ID
    1. if no match was found to Unique\_ID
       1. add “NO MATCH FOUND” to station and water body columns on “final all data” sheet
       2. announce to user that the missing Unique\_IDs need to be added to the station/Water Body file, and the macro needs to be re-run
11. apply reporting rules to lab data based on **DWM reporting rules\_10-20-11.xls** (located at: W:\DWM\Data\laboratory\_QA\reference\_files)
    1. apply “##” to data cells for censored data (Result or Secchi Depth)
    2. apply “- -“ to depth data cells
       1. for Survey Type = River: if Sample Depth, Secchi Depth, Relative Depth, Maximum Depth cells are blank
       2. for chlorophyll a: if Relative Depth is blank
    3. apply “\*\*” to data cells to indicate missing data
       1. where Result = “Missing”
       2. where dates/times = -1
       3. where Flow Condition = blank or “Not Recorded”
       4. for Survey Type = “Lake”, where Secchi Depth, Sample Depth, Relative Depth, Maximum Depth is blank
    4. apply “^^” to data cells to indicate “NO WATER”
       1. where Flow Condition = “NO WATER” (all Survey Types): apply to Result, Analyte, Units, Analysis Method
       2. where Flow Condition = “NO WATER” (Survey Type = “Lake”): apply to Sample Depth, Relative Depth, Maximum Depth, Secchi Depth
       3. reset Result Qualifier and Secchi Qualifier columns to blank
    5. format cells for significant figures based on analyte
       1. if result is < MDL or RDL value, or > UQL value, leave value as reported from lab (do not apply significant figures)
       2. format depths to 1 decimal place regardless of data value range
12. make corrections to data formats
    1. standardize date and time formats
    2. combine sample depths 1 and 2 to create final Sample Depth (Lakes only)
       1. if sample depth 1 = -9, then sample depth 1 = 0
       2. if no sample depth2 (blank), then
          1. if depth1 = -8, then Sample Depth is missing
       3. if 2 depths reported, then
          1. if depth 2 not reported (blank), then Sample Depth is missing
          2. if depth1 = -8, then sample depth 1 = 0
       4. if analyte = “chlorophyll a”, then
          1. if Relative Depth = “- -“ (no data) and depth 2 is blank, then Sample Depth is missing
    3. convert Relative Depth codes to phrases (s = Surface; nb = Near bottom; m = Mid-depth)
    4. for lakes with mile point = 999, reset values to blank
    5. correct cases where Connecticut mispelled (source of mispelling is electronic field sheet)
    6. delete extra columns used for applying formatting rules above
13. apply formatting to cells
    1. apply colors and borders and fonts to cells; left justify and center vertically
    2. set column widths
14. add filename and path to File Process Info sheet
15. create summary sheet of sample counts (YYYY QC2 lab data summary)
16. split “Final all data” sheet into 4 worksheets: lakes, rivers, field blanks, duplicates
    1. field blanks sheet
       1. sort “Final all data” sheet by QC Type Name
       2. filter “Final all data” sheet for QC Type Name = “Field Blank”
       3. copy rows to “Field blanks” sheet
          1. if no blanks found, then write “No Field Blanks for YYYY”
       4. delete unnecessary columns (all depth columns; QC OWMID; Flow Condition)
    2. lakes sheet
       1. sort “Final all data” sheet by Survey Type, then by QC Type Name
       2. filter “Final all data” sheet for Survey Type = “Lake” and QC Type Name = “Routine Sample”
       3. copy rows to “Lakes” sheet
          1. if no Lake rows found, then write “No Lake Projects for YYYY”
       4. rename “Flow Condition/Lake Level” column to “Lake Level”
    3. rivers sheet
       1. sort “Final all data” sheet by Survey Type, then by QC Type Name
       2. filter “Final all data” sheet for Survey Type = “River” and QC Type Name = “Routine Sample”
       3. copy rows to “Rivers” sheet
          1. if no River rows found, then write “No River Projects for YYYY”
       4. rename “Flow Condition/Lake Level” column to “Flow Condition”
    4. duplicates sheet
       1. sort “Final all data” sheet by QC Type Name
       2. filter “Final all data” sheet for QC Type Name = “Duplicate”
       3. copy rows to “Duplicates” sheet
          1. if no duplicates found, then write “No Duplicates for YYYY”
       4. sort duplicates sheet by Analyte, sample OWMID, QC OWMID (puts pair #s together by analyte)
       5. for each duplicate pair by analyte, copy lower OWMID to lakes and river sheets, respectively
          1. if lower OWMID was censored/missing/no data/no water, then use next higher ID (in case of quadruplicates), if both have been censored, use lower ID
          2. if only one ID found, then copy that row
       6. delete unneeded columns from duplicates sheet (Depth columns; Flow Condition)
    5. delete Depth columns from Rivers sheet (after duplicates have been copied)
    6. sort individual sheets to preserve sort order for project file export (each sheet is sorted by Project, then Watershed, then Sample Date, then Sample Time
17. export project files (each file will contain a worksheet for lakes, rivers, blanks, and duplicates)
    1. get unique project names from final all data sheet
    2. create individual project files
       1. name file as "*Project*\_*YYYY*\_labdata\_QC2.xlsx"
       2. copy “read me” sheet from QC2 lab template.xlsx
          1. add hyperlink to field sheet meta data file
       3. create Rivers, Lakes, Duplicates, Field Blanks sheets
          1. filter Final all data sheet and copy rows to the appropriate sheet in project workbook
          2. if no rows found, then print “No River Projects for YYYY”, etc. to appropriate sheet
          3. split and freeze 1st row on each sheet
          4. protect each sheet
       4. delete blank/unused sheets
       5. set file attribute as read-only
       6. add summary counts for QC2 Final, by project to “YYYY QC2 lab data summary” sheet
    3. sort Final all data sheet by Project, Watershed, Sample Date, Sample Time
    4. create All Projects workbook (include Lakes, Rivers, Field Blanks, Duplicates sheets; Data summary sheet)
18. final clean up of QC2 file
    1. autofilter each final sheet, freeze top row
    2. write date macro was run to File Process Info tab

## LabData\_Formulas.bas

(This VBA module contains 2 VBA functions, GetQual and GetFinalDecision, described separately below)

### Function GetQual

* General purpose:
  + To combine all DWM qualifiers (a, b, d, e, f, h, m, p, r, t) into one cell (comma-separated)
* Required input:
  + Files:
    - QC2 file: **YYYY\_laboratory\_QC2\_mm-dd-yyyy.xlsx,** where YYYY is the data year being processed, and mm, dd, and yyyy are the month, day, and year the QC2 file was created (QC2 file is placed in directory: W:\DWM\Data\laboratory\_QA\YYYY\QC2)
  + QC2 File Worksheets:
    - YYYY\_laboratory\_QC2 working
  + QC2 working sheet columns:
    - Qualifier columns for a, b, d, e, f, h, m, p, r, t
    - Each cell in a given qualifier column is filled out with N/A, Accept, Qualify, Censor
* Output:
  + Individual cell results on QC2 working sheet, Combined Qualifiers column
* Approximate run time: seconds

Procedure:

1. For each unique qualifier cell, check values
   1. if Decision is Qualify or Censor, then add the qualifier letter to a combined qualifier string
   2. if Decision is Accept, then no letter for that qualifier is added to the string
   3. assign cell value of combined string as GetQual function result
2. NOTE:
   1. Code DOES NOT distinguish whether a qualifier cell has been left blank (i.e. blank cells are treated as Accept—no qualifier letter is added to the string)

### Function GetFinalDecision

* General purpose:
  + To apply final QC decision (Accept, Qualify, Censor) based on all qualifier decision columns for a particular OWMID/analyte
* Required input:
  + Files:
    - QC2 file: **YYYY\_laboratory\_QC2\_mm-dd-yyyy.xlsx,** where YYYY is the data year being processed, and mm, dd, and yyyy are the month, day, and year the QC2 file was created (QC2 file is placed in directory: W:\DWM\Data\laboratory\_QA\YYYY\QC2)
  + QC2 File Worksheets:
    - YYYY\_laboratory\_QC2 working
  + QC2 working sheet columns:
    - Qualifier columns for a, b, d, e, f, h, m, p, r, t; Combined Qualifiers column
    - Each cell in a given qualifier column is filled out with N/A, Accept, Qualify, Censor
    - Combined Qualifiers column contains the formula to calculate the combined qualifiers (or user has manually overwritten the cell contents with non-represented qualifiers)
* Output:
  + Individual cell results on QC2 working sheet, Final Decision column

Approximate run time: seconds

Procedure:

1. For each unique qualifier cell, check values
   1. if Decision is Censor, then set censorflag to -1 (only one qualifier is required to make the Final Decision equal Censor)
   2. if Decision is blank, then set blankflag to -1 (any required qualifier cell that is blank results in a blank Final Decision)
   3. determine Final Decision
      1. if censorflag = -1, then Final Decision = “Censor”
      2. if censorflag = 0 and combined qualifier cell is not blank, then Final Decision = “Qualify”
      3. if censorflag = 0 and combined qualifier cell is blank, then Final Decision = “Accept”
      4. if blankflag = -1, then Final Decision is blank (this overrides 1.c.i. through 1.c.iii.)
2. NOTES:
   1. Final Decision is left blank if any qualifier cell is blank; this allows user to filter the Final Decision column for blanks and add the missing qualifier Decisions
   2. User can manually override the formula’s calculated Final Decision when best professional judgment is required
   3. To retract any BPJ decision the formula must be manually copied and pasted back into the cell

# APPENDIX D

# *QC3-level review guidelines (LAB DATA)*

During the QC2 stage, data not meeting DWM validation criteria or otherwise suspect have been either qualified or censored.   For QC3-level review, the general goal is to try to avoid having to make changes to the data after they are “published”.  The current QC3 process is basically an external check on the data processing and QC2 review.  Questions to ask include, but are not limited to:

1. Is the dataset **complete**?   Are there missing data that are not represented in the QC3 export?

2)    Are there **errors in the QC2 technical assessment**, or was something missed (e.g., omissions)?

3)    Do there appear to be any **inexplicable outliers or unreasonable data** that are probably not real?

4)    Are there any indications that any of the data are not **representative** of the station and site conditions?   Is there reason to believe, for example, that one or more samples may have been collected improperly for a specific reason or in error (e.g., in a mixing zone of an outfall, in a backwater pool)?  We did look for these potential situations, but you may have information to add to the evaluation.

5)   Is there **anything “broken” in the data files** (e.g., a feature that does not/did not work)?

For QC# Reviewers: If you do have comments and/or suggested edits , please DO NOT MAKE CHANGES TO THE DATA FILES, but provide written comments to me so we can document any changes that may be made as a result of your reviews.

Upon completion of the review, any required edits are made by QC2 staff.  At that point, the data are considered final (QC4-level) and moved to the “QC4 data” sub-folder, where they will remain.   Once published, the data will be at QC5-level (but the final data table will still reside in the QC4 data location).

# APPENDIX E

# DRAFT

# *Summary of Automated and Manual QC Checks on Laboratory Data (QC1-QC3)*

NOTES:

1. This list shows data quality checks only, not all tasks performed.
2. Checks are listed in the general order in which they occur
3. In cases where both code and BPJ are used, the manual component is typically a review/edit mode of the automated process
4. Some checks listed under Code are actually performed “manually” by the staff person directly applying the code

| **QC Check** | **Code/Macro** | **Manual/BPJ** |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Fieldsheet data entry QC | no | yes |
| EDD (non-WES) error identification | no | yes |
| Discrepancies re: dates between fieldsheet and LIMS/EDD | Yes? | no |
| Syntax errors (ALL LAB DATA file) | yes | no |
| Format errors (ALL LAB DATA file) | yes | no |
| Format errors (Fieldsheet metadata) | yes | no |
| Identify unmatched FS OWMIDs | yes | no |
| Identify unmatched Lab sample #s | yes | no |
| Identify unmatched Lab OWMIDs | yes | no |
| Unmatched-ID file review/edit | no | yes |
| Units/Methods confirmation | yes | no |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Exclusion of non-project data | yes | no |
| Field blank contamination | yes | yes |
| Field Duplicates > DQO | yes | yes |
| Holding time violations | yes | yes |
| Poor QC sample frequency | yes | yes |
| Trip effect (b, d) decisions | yes | no |
| Lake/River sample designation (2005-2010 only) | no | yes |
| Secchi depth “issues” | no | yes |
| Confounded results (e) | no | yes |
| Lab Qualifier carryover | no | yes |
| Lab comments “issues” (misc.) | no | yes |
| Fieldsheet comment “issues” (misc) | no | yes |
| Lab audit (external) evaluations | no | yes |
| Field audit evaluations | no | yes |
| Misc. communication “issues” | no | yes |
| Qualifier “baggage” assessment | no | yes |
| Automated decision review/edit (b, d, f, h) | no | yes |
| Final result review/edit | no | yes |
| Tidal influence evaluation | no | yes |
| Trip effect random checks | no | yes |
| Completeness checks | yes | yes |
| Obvious outliers | no | yes |
| Final checks on censorings | no | yes |
| Review/approval of excluded data | no | yes |
| Application of reporting rules (inc. NDs, rounding, TNTCs, etc.) | yes | no |
| Review/edit data output | no | yes |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| QC3 project-level review (missing data, outliers, etc.) | no | yes |
| Review/approve/edit output based on QC3 review | no | yes |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |