



MassDEP

**Massachusetts Department of Environmental Protection
Bureau of Water Resources
Division of Watershed Management
Watershed Planning Program**

STANDARD OPERATING PROCEDURE

External Data Management and Reviews

CN 616.0

June 2025 – June 2027

**Prepared and
Approved by:**

Robert Smith

Robert F. Smith, External Monitoring and Data
Coordinator

Date: June 13, 2025

Approved by:

Jasper Sha

Jasper, Sha, QA Analyst

Date: 6/17/25

Approved by:

R. Chase

Richard F. Chase, Section Chief, Data Management
and Water Quality Assessment Section

Date: 8/12/25



List of Revisions

Revision Date	Revision	Pages #s	CN/ (Old CN if applicable)	Initials
June 2025	Original	19	NA	RFS



Table of Contents

Section	Page #
Definitions	4
Scope and Application	4
Summary of Method	4
Safety Considerations	5
Personnel Qualifications/Responsibilities	5
Equipment and Supplies	5
Detailed Procedures	5
External data acquisition	6
External data database	7
Continuous data management	8
Geospatial data management	8
Biological data management	8
Geospatial data review	8
Water quality and biological data review	10
Feedback to data providers	13
Data harmonization and export for analyses	13
Analyses and assessment	14
Quality Control	15
Data & Records Management	15
Data Software	15
References	15
Appendices	16



Definitions: The following definitions and acronyms are used in this SOP.

Agency: Any governmental organization (municipal, state, or federal), tribal nation, or other government sponsored organization.

Assessment Unit: AU: a waterbody segment partitioned to represent homogeneity in physical, biological, or chemical conditions at a spatial scale appropriate to characterize the SWQS attainment status of each AU. AUs should represent a waterbody or a portion of a waterbody larger than the reach associated with a single monitoring station. See CN 618.0 (MassDEP 2025a).

External data: Data collected by an agency, non-profit, or other entity not associated with MassDEP.

External Data Database: EDB; this is the database used to house discrete water quality data.

External data provider: The term referencing the suite of potential entities that WPP may acquire external data from.

Harmonization: The process of making corrections, format changes, structure changes, etc. to tabular data files (e.g., R data frames, Excel Sheets, database tables, etc.) to ensure interoperability among systems resulting in accurate, consistent, and efficient data management, summary, and analysis products.

Internal data: Data collected by WPP.

Organization Code: OrgCode; This is a unique code assigned to external data providers by WPP for record keeping purposes.

Report number: The number, corresponding to a directory in the SharePoint, that is assigned to an individual external data submittal. Report numbers are assigned sequentially when submissions are made.

Station: The location where a water quality and/or biological sample was collected.

Station definition: A sample location's unique identifier, waterbody name, general description, coordinates (i.e., latitude and longitude), and other metadata (e.g., AU association, basin code, etc.).

Station ID: A unique code used to identify the station. All Station IDs have the official organization code added as a prefix to the data provider's submitted Unique ID.

Scope and Application: External data management and analyses are required for MassDEP to assess surface water quality in accordance with requirements set forth in § 305(b) and § 303(d) of the federal Clean Water Act (CWA). This SOP describes the general procedures for acquiring, reviewing, and managing data from non-MassDEP (i.e., "external") sources for all uses, but the primary purpose of this data is to inform WPP's actions to satisfy CWA regulatory requirements. The procedures described herein describe general processes that may require ad-hoc alterations to achieve project specific goals. Detailed descriptions of these fine-resolution procedures are beyond the scope of this SOP.

Summary of Method: This SOP covers WPP's general procedures for acquiring, reviewing, and managing water quality and biological data collected by non-MassDEP entities (i.e., "external data"). WPP acquires external data from diverse sources, which require diverse procedures for all phases of the data review and management process. Data acquisition, screening, storage, harmonization, review, export, and analyses will differ depending on the data source, structure, composition, completeness, and quality. Several data harmonization steps are required to prepare data for: (a) specific procedures in the data review and management process and (b) use in WPP's analytical processes (including assessment). The



ultimate goal of data review and management is to assign a usability score to the data that classifies its utility for use in assessing water quality. Data may be characterized as usable for (1) educational purposes; (2) screening purposes; or (3) water quality assessments. Thus, external data have multiple uses for informing WPP activities, and the procedures in this SOP support data review and management activities for all WPP activities.

Safety Considerations: No safety procedures are required. This SOP covers data management only.

Personnel Qualification/Responsibilities: WPP utilizes a team of staff scientists with expertise in data management. No specific certifications or training are required, but certain staff must have knowledge of and the ability to work with Microsoft Access databases. All activities are centralized in the Water Quality Assessment and Data Management Section. The External Monitoring and Data Coordinator is assigned by WPP to coordinate the data management and review processes described in this SOP. The External Monitoring and Data Coordinator controls access to all raw data and databases required for external data management and review. Access control includes coordinating data dissemination to WPP staff as needed using consistent procedures that maintain data integrity.

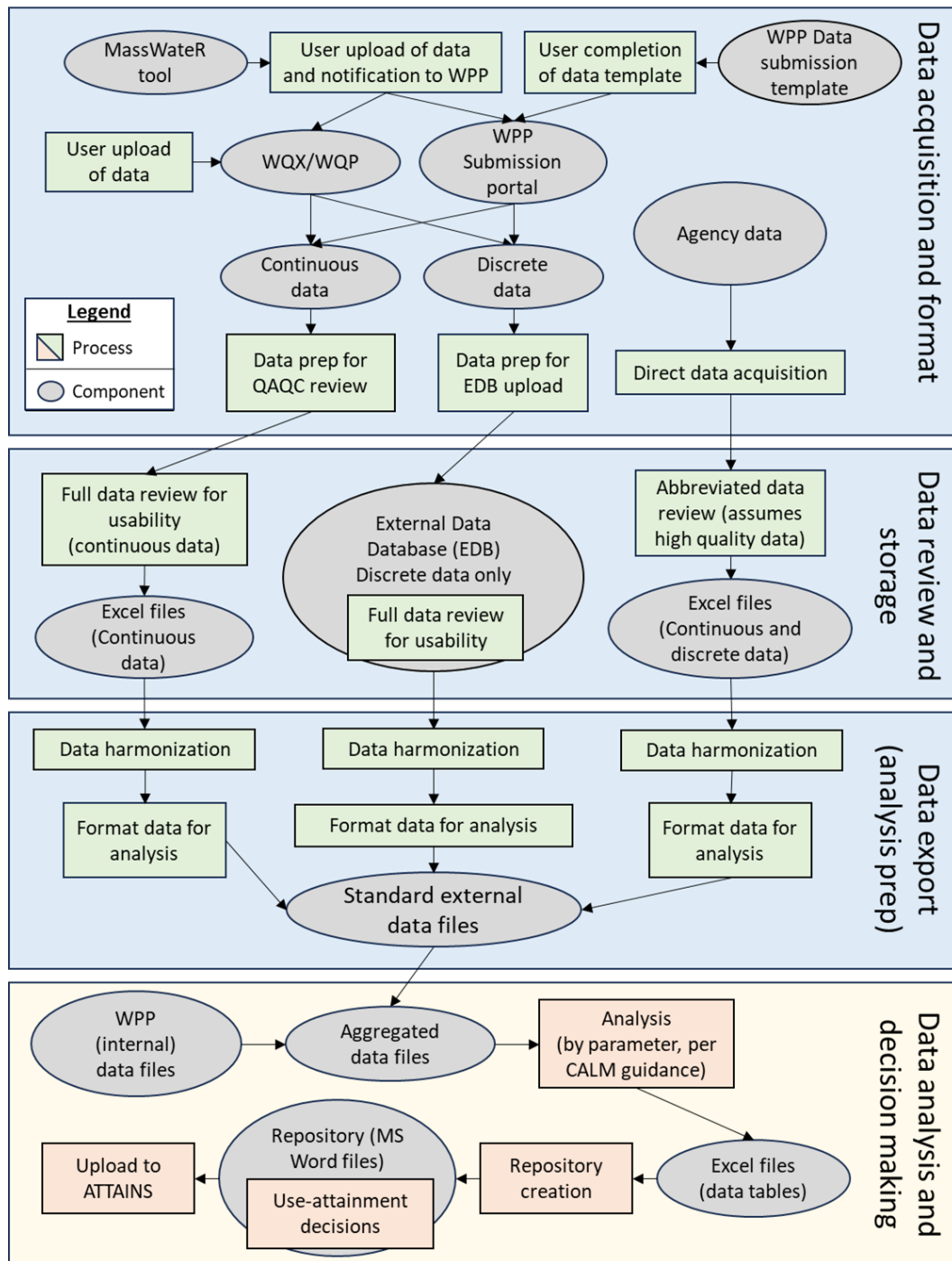
Equipment and Supplies: No specific equipment or supplies are required other than standard MassDEP computing equipment (including Microsoft Office software).

Detailed Procedures: WPP regularly seeks relevant surface water quality data from non-MassDEP sources (i.e., “external” data) for assessing surface water quality in accordance with requirements set forth in § 305(b) and § 303(d) of the federal Clean Water Act (CWA). In assessing surface water quality in the Commonwealth and in preparing lists of impaired waters to meet CWA requirements, states must consider all existing and readily available data and information. In addition to data collected by WPP (i.e., “internal” data), WPP solicits external data and information from any and all potential non-MassDEP data providers. External data are managed and reviewed using standard procedures as outlined in the following SOP. External data providers include, but are not limited to:

1. Volunteer monitoring organizations
2. Watershed conservation organizations
3. Lake associations
4. Conservation organizations
5. Tribal nations
6. Cities, towns, and municipalities
7. Federal and state agencies

WPP maintains a website (<https://www.mass.gov/guides/external-data-submittals-to-the-watershed-planning-program>) with instructions for potential external data providers that describes: (a) requirements for data submission; (b) how to submit data; (c) the general review process and data usability requirements; and (d) information about the current assessment cycle relevant to external data submittals (e.g., data submission deadlines). WPP also maintains a guidance document with fine-resolution detail on certain processes (Smith 2025).

The processes for reviewing and managing external data consist of data acquisition, review, storage, export/preparation, and analysis. The analytical processes are described in the current Consolidated Assessment and Listing Methodology guidance document (e.g., MassDEPb 2025) and are not covered in this SOP. The following diagram provides a general overview of the entire process:



External data acquisition: External data are obtained through: (a) submittals to MassDEP by external data providers; (b) data harvesting by WPP staff from readily available data sources (including EPA's Water Quality Portal; WQP); and (c) direct data requests to data providers (only applicable to agencies). WPP contracted projects are generally considered internal data, but these data are occasionally managed within the external data system.

Submittals to WPP from external data providers are required to use (a) the WPP Data Submittal Portal (i.e., the "portal") or (b) EPA's Water Quality eXchange (WQX) with a notification to WPP through the



portal with any additional documentation not uploaded to WQX. The portal is an email account (WQData.Submit@mass.gov) that external data providers use to send data and associated submittal files to WPP (or to send a notification of upload to WQX).

Data submittals directly to WPP through the portal must use the data file templates provided on WPP's website. The template facilitates data upload to the database and data review and management (see below). The template is an Excel spreadsheet with sheets for: (1) discrete water sample data; (2) analyte field QC data; (3) probe field QC data; (4) lab QC data; (5) probe calibration data; (6) continuous data; (7) continuous data QC data; and (8) biological data. Non-tabular data (e.g., image files, aesthetics data, GIS files, etc.) can be submitted separately from the template through the portal. Data submitted through the portal using the template undergo preliminary review to correct obvious errors prior to upload, including, but not limited to: (a) misspelled analyte names; (b) date/time formatting errors; (c) empty cells, merged cells, and other spreadsheet formatting issues; and (d) incomplete data entries (especially for geospatial data). All data submittals are required to include: (1) data file(s) using WPP's template and (2) the statement of data integrity. Additional files (e.g., reports and project QAPP(s)) are not required but may aid in data review and are submitted at the discretion of the data provider or may be requested by WPP staff during review.

Data submitted to WQX are harvested from the Water Quality Portal (WQP) using the R statistics software and R-package "dataRetrieval". Data providers must email the portal any required or optional files not uploaded to WQX. Data obtained from WQP and directly from other sources are harmonized in preparation for upload to the database using custom R scripts. The harmonization process transforms the data into the structure used for the template to utilize a singular process for data upload and review.

All data submittals are logged in a spreadsheet and assigned a report number for tracking. Report numbers are assigned sequentially as data are acquired by WPP. The data submittal log includes additional metadata required for tracking the review process including (but not limited to): data origin/provider, submission date, data years, data contact information, notes on the types of data included, notes on review status, and notes on the review status and outcomes.

Data files, databases, and all required files for reviewing and managing external data are stored in a MassDEP managed Microsoft SharePoint storage system. The External Monitoring and Data Coordinator is the only WPP staff with full control of data files (other than IT support and administrative staff). Data submissions are stored in directories labeled by the assigned report number. Access to raw data and review files is provided to specific staff who assist with data reviews. Harmonized and reviewed (i.e., "finalized") data files are provided to WPP staff as tabular files exported from the relevant database or as copies of reviewed continuous data files. All data files originally submitted by the data provider are maintained, and calculations, corrections, notes, and other edits are made to copies of original files. A specific directory is maintained for geospatial data created as part of the geospatial review and stationing procedure (see below).

External Data Database: WPP utilizes two Microsoft Access databases for uploading, reviewing, scoring, storing, and managing discrete water quality data. This database system is called the "External Data Database" (i.e., EDB). One database is the "front end", which includes queries and forms designed to facilitate data upload and review (see below). Once upload and review are complete, queries transfer the data to the other "back end" database. The "back end" database is the data warehouse that facilitates data storage and management. Queries historically were used to create data exports, but WPP currently utilizes R scripts to format, filter, and harmonize external data for analytical processes (see below). The EDB is provided to WPP staff for data entry and/or QAQC activities through a "check-out" procedure where a copy of the database is made, the original archived, and the copy provided directly to the staff by the External Monitoring and Data Coordinator. The EDB is checked in when work is complete. All transactions are recorded in a log with an explanation of the work performed.

The EDB was recently evaluated by a contractor through an EPA sponsored project that included developing a strategic plan for improving external data management. The strategic plan suggested that



WPP replace the EDB with an SQL based database, which is planned by using funds from a SFY2024 EPA Exchange Network grant awarded to WPP.

Continuous data management: WPP currently does not have a system for continuous data management. Continuous data is reviewed and managed through manual processes that use copies of raw data files and customized processes (including R scripts) for each data submission. A system will be developed using funds from an EPA Exchange Network Grant.

Geospatial data management: Geospatial data is maintained in the EDB and an ESRI geodatabase. Data upload to the EDB includes the data provider's station definition information (e.g., station name, description, and coordinates). During the data review process (see below), station definitions provided by the external data provider are harmonized to determine if they match historical submittals or are new stations. The EDB contains only the original geospatial data from the submission, and the geodatabase includes point feature classes containing the original data and the harmonized (i.e., "final") data. Data is considered "final" when the geospatial review is complete and WPP staff verified the complete station definition information. The point feature class of stations also includes the following station definition information:

1. The Assessment Unit (AU) the station is associated with.
2. The type of waterbody the station is located on (e.g., stream, lake, estuary, etc.; see Appendix 1)
3. The basin the station is located within.
4. Notes on the stationing procedure.

The geodatabase point feature class of stations also includes fields for tracking reviews including a notes field (see description of the stations audit below). The point feature class of stations includes stations for all submitted data regardless of usability score. Stations that are unclear are assigned coordinates for the location of the Statue of Liberty (40.69, -74.045) and are maintained for record keeping and data management purposes.

The geodatabase was created in 2024 to support a reliable and standardized management process for external data stations. The creation of the external data station geodatabase included a full audit of all external data stations, station definitions, AU associations, and other station metadata. The audit reconciled disparate geospatial data sources and created a single unified system that permanently supports all external data processes. Metadata in the attribute table of the point feature class of points maintains notes collected and generated during the audit.

Biological data management: Submission of benthic macroinvertebrate data is rare. Fish population and other biological data have never been submitted by external data providers. Benthic macroinvertebrate data are not stored in the EDB but were historically uploaded to the MAbenthos system. Any future submissions of biological data will be held temporarily as raw data files and utilized for assessment if determined to meet data quality requirements. These data are usually not assessment quality due to the rigorous taxonomic requirements.

Geospatial data review: Geospatial data submitted with water quality and/or biological data are reviewed primarily to ensure accuracy of the sample location. Samples taken at stations with inaccurate or incomplete station definitions are censored (i.e., marked for removal from analyses) or WPP staff work with data providers to correct station definition information. Data submittals with overall poor geospatial documentation and/or high rates of inaccuracies, however, can result in a decision that the entire dataset is not usable for assessment purposes.

The geospatial review occurs after all raw data is upload to the EDB. Final station definitions are linked to data in the EDB through the EDB's record number, which is assigned to each observation when uploaded. The EDB's 'EDDMain' table is exported to a tabular format, and R scripts are used to generate a tabular file that filters for the new upload(s) and lists each unique combination of (1) organization name; (2) waterbody, (3) Station ID; (4) latitude; (5) longitude; and (6) station description. Each unique combination is an entry in the geospatial dataset requiring review (and is assigned a reference number). The tabular geospatial data of new stations is joined (using R) to a tabular export validated point feature



class's attribute table based on Station ID. This allows for existing final station information to be linked to the new station if the station exists in the feature class of validated station definitions, which reduces manual data entry steps. R scripts harmonize the data by adding fields and changing field names to match those in the validated point feature class of station definitions. This step includes adding a station reference field (ExtData_station_REF) and assigning the appropriate value to each entry. The station reference is a unique identifier for the station entry (and required to link EDB records to validated station definitions). The tabular data file is loaded into a GIS as a point feature class for review. The following is performed for the geospatial review with work done directly to the feature class's attribute table:

1. Check for blank entries in Station ID, latitude, longitude, and station description fields (may cause an upload error). "Missing" is entered for all empty fields. Entries with missing latitude and/or longitude will not load, and a value of 0 is entered for the missing field.
2. Point features matched to existing stations linked from the join to the validated data are examined to confirm the station definition and all other information.
3. Point features of new stations are examined manually, and all validated geospatial data are entered into the fields of the attribute table designated as "Final". Original geospatial data remain in the attribute table for reference.
4. AU associations, waterbody type, basin information, and other metadata are entered as needed during the review of geospatial data submitted by the data provider.
5. Station definition data are verified by another WPP staff member.
6. When complete, the edited point feature class of new stations is appended to the existing point feature class of validated station definitions. The combined point feature class is renamed with a sequential version number. Multiple reviews (and appends) can be done before finalizing a new version of the validated station definition file.
7. A tabular data file that crosswalks EDB records numbers to the station reference number (for each unique entry in the geodatabase) is updated using R scripts.

Geospatial review of new data may reveal that previous station definitions were: (a) incorrect as provided by the data provider and/or (b) incorrectly interpreted by WPP during the geospatial data (e.g., associated with the wrong AU). Errors uncovered during review of new stations should be addressed by reconciling entries in the point feature class of currently validated station definitions.

The following are general rules for review but are not inclusive of all situations WPP staff must address during geospatial review:

- Station entry is listed as Do Not Use (i.e., "DNU") when no waterbody or description is provided (i.e., only coordinates are submitted).
- Station entries with a waterbody but no description can be used if the coordinates indicate the station is clearly associated with the listed waterbody. Stations are marked as DNU if the coordinates do not clearly align with the listed waterbody.
- Station entries with coordinates that are geographically far from the listed waterbody and/or from the location referenced in the station description are temporarily listed as DNU. WPP reaches out to the organization for clarification. Without further clarification, the station entries remain DNU.
- Stations submitted by the same data provider with names that are similar but have different coordinates are temporarily marked as DNU. WPP reaches out to the organization for clarification. Without further clarification, the station entries remain DNU.
- Stations submitted by the same data provider with different names but have the same (or similar) coordinates, waterbody name, and station description are assumed by WPP to be the same station. All stations take on the same final station definition.

BPJ is commonly used to make determinations about geospatial data, but a DNU is generally assigned to all data associated with stations with missing information and/or careless record keeping leading to unclear station definition information.

Certain agency data may be stored separately from the EDB, which requires the station data to be stored separately. In this case, a dedicated point feature class of validated station definitions is maintained.



Water quality and biological data review: All data acquired by WPP from external sources are reviewed for potential usability, but the review procedures may differ for certain data sources. WPP generally does not review data from federal agencies and assumes that this work was completed under an approved QAPP and meets rigorous QAQC requirements. In all cases, original data are copied, and all reviews (which may require calculations and other direct edits to data files) are done on data file copies. Reviews for all data types include:

1. Determining if the organization has an approved QAPP for the project/program that generated the data.
2. Determining if the submittal is complete (e.g., includes QA data) and in a format that can be reviewed by WPP. This includes confirming that the statement of data integrity is included.
3. Determining if a certified or otherwise qualified laboratory performed the work to generate the data (if applicable).

Recent guidance from EPA has indicated WPP can (and should) examine how data are submitted as part of assessing usability for assessment (e.g., data submitted to us in a format that requires substantial data wrangling may have its usability score lowered). Any data submitted in a narrative form only would not be reviewed and automatically determined unusable for assessment purposes.

Review of discrete water quality data is the most commonly performed review by WPP. This review utilizes the EDB for entering notes and assigning usability scores to data grouped by the EDRID. An EDRID is a unique code assigned to data grouped by: (1) report number; (2) analyte; (3) data year; and (4) organization code. This is the historical grouping for data usability score assignments (but see below details on per-datum flags that also relate to data usability). Review scores and review comments are entered directly in the front end database (and are transferred to data tables in the back end database).

The review process starts with data clean up and formatting steps (i.e., “data wrangling”). Excessive needs for data wrangling are considered in the scoring. The rationale for using data organization as a review criteria is that data providers who are unorganized or ignore data reporting guidelines are likely to make mistakes during data collection and data management activities. A manual review of the data evaluates numerous criteria related to data quality including, but not limited to:

- All analytes measured have matching QAQC entries.
- Data Quality Objectives (DQO) are checked and reported properly.
- Units are correctly assigned to analytes and presented (e.g., spelled) correctly.
- Minimum and maximum values indicate possible error(s) in measurements.
- Laboratory used for any chemical analyses are certified or otherwise qualified.
- Field and laboratory methods used are appropriate.
- Reporting of audits and/or training demonstrate competency.

The scoring system employed for characterizing data usability for assessment (below) provides the framework through which all review criteria are applied to determine usability. Data wrangling steps include a specific review for “reasonableness” that examines certain criteria such as the minimum and maximum of values and the units reported. Summary statistics are calculated as needed, but no formal set of statistical analyses are currently employed (inclusion of statistical analyses will be implemented with the new database; see above).

WPP employs a scoring system that structures data reviews of usability for assessment. The system is a grading system that bins grades into three tiers:

- **Level 1 – Educational/Stewardship:** These data generally do not meet the rigor (i.e., accuracy, precision, frequency, comparability, overall confidence, etc.) required for use in waterbody assessments. This data is not used for 305(b) or 303(d) decision-making or to guide future sampling efforts.
- **Level 2 – Screening:** These data may meet the data quality objectives in the submitter’s QAPP, but the data generally fail to meet one or more WPP criteria required for direct use in assessments (i.e., fail to meet criteria used in WPP’s monitoring program QAPP approved by the



EPA). Screening-level are typically only used to direct future sampling efforts or as supporting evidence only.

- **Level 3 – Regulatory/Assessment:** These data are typically the result of extensive planning, attention to detail, relatively stringent data quality objectives, training, standard field and lab procedures, metadata collection, project organization and data verification. These traits create data that are scientifically sound and legally defensible. These data are usable for 305(b) and 303(d) decision making.

An important difference between Level 1 and Level 2 is that Level 1 data are generally: (a) submitted to WPP generated from work without an approved QAPP and/or (b) generated through work by a laboratory that isn't certified or otherwise qualified to perform the work. Similarly, data submitted with exceedingly poor or missing geospatial data are assigned a score in Level 1. The data review process immediately stops when the data are determined to be Level 1 and a score is assigned (see below). Thus, Level 1 data may appear high quality, but WPP will not use this data for any program planning and/or project implementation.

Review scores assigned to EDRID data groups indicate if the data should be: (a) censored; (b) used with caution; or (c) used without caveats. Notes are added that identify the specific observations (i.e., datum) that result in the EDRID level score (e.g., samples from Station A should be censored but all other data is usable without caveats). These notes about datum level evaluations are vital for recently developed procedures to add per-datum flags. Only the EDRID level usability scores are stored in the EDB. The usability scores are:

Level	Usability score	Description of usability score
Level 3	A+ _ASSESS	All data should be considered <u>usable</u> by DWM for assessment purposes without caveat
Level 3	A- _ASSESS	All data appear to be <u>usable</u> for assessment purposes, but some data should be used with caveat (as noted) due to special circumstances.
Level 3	B _ASSESS	Some of the data appear to be usable (with caution), as explained in the review comments and summary
Level 2	C _SCREENING	All data should be considered <u>usable</u> by DWM for screening purposes without caveat
Level 2	D _SCREENING	All data appear to be <u>usable</u> for screening purposes, but some data should be used with caveat (as noted) due to special circumstances.
Level 2	E _SCREENING	Some of the data appear to be usable (with caution), but other data should not be used, as explained in the review comments and summary
Level 1	F	None of the data should be used by WPP for any purpose
Level 1	I	Data usability inconclusive, due to lack of information
Level 1	S	Data is considered for educational purposes only. No specific usability score is defined.

Data providers generally submit data with the intention for it to be used for assessment (i.e., scored as Level 3). Thus, in practice, WPP's review generally "moves" the score down during the data review process by using Level 3 scores as the starting point (i.e., benchmark). Final usability scores are applied subjectively to data based on the WPP's staff interpretation of review criteria. No formal rubric of scoring exists. WPP is, however, in the process of developing a defined rubric for scoring data as part of the database development project (see above). The following are examples of common criteria used in the review:

- Agency-approved QAPP
- State-certified (or otherwise qualified) laboratory performing analyses (parameter-specific)
- Documented QAQC activities and data quality assessments
- Number of valid results (vs. required)
- Representative, documented, and accurately described sampling locations
- Training



- Data Verification
- Internal field and/or lab audit(s)
- Project organization
- DQOs relatively generally stringent and comparable to WPP's
- Lab/calibrated instrument vs. kit use
- Level of documented QAQC (e.g., instrument calibration)
- External field and/or lab audit(s) by agency/other
- Calibration of instruments prior to use
- Inspection/maintenance activities (as needed)
- Sufficient metadata documentation
- Voucher sample (biological)

Reporting QAQC data is vital for data reviews. The following specific QAQC information is evaluated using the following steps:

1. Check that QAQC data are provided for all parameters.
2. Check that QAQC data are provided for all instrument measurements.
3. Check that QAQC samples are collected at the frequency required in the project QAPP.
4. Check that QAQC statistics are correctly calculated.
5. Check that QAQC samples meet DQOs in the project QAPP (and are comparable to WPP's DQOs).
6. Check that raw data failing QAQC are clearly flagged or censored or not included in the data submission.

In addition to direct review of tabular data files, WPP staff make general assessments of the entire data submittal. This includes examining the following:

- Review any accompanying reports for context and additional information (a report, however, is not required).
- Examine the age of the data.
- Examine notes, reports, etc. for indications of "non-representative conditions" (e.g., sample locations are tidally influenced, wet-weather events, temporary construction runoff, etc.).
- Examine sample frequency.

Any criteria above, or any other criteria determined by the reviewer to be important for the data's quality, that the reviewer determines fail to meet WPP's threshold for usability results in a scoring decrease. Scores are applied to and notes written to each EDRID data group. Scores of A- and B are the only scores that indicate a portion of the data isn't uniformly usable or unusable for assessment purposes. Thus, the per-datum notes in the review narrative are vital for creating per-datum flags during the data harmonization and export process (see below).

Data reviews commonly require WPP staff to email data providers for additional information and/or to clarify details of sampling methodology, QAQC data, and other aspects of the data to create the most accurate usability score possible and to abide by EPA's directive to use any and all readily available data. Required follow up with data providers is commonly identified during preliminary data reviews.

Reviews for continuous data generally follow the same procedures as described above and utilize the same or similar criteria. The following are unique to continuous data reviews:

- Additional review of measurement units to ensure congruency among all data in the continuous set.
- Checks are made to ensure that side-by-side field readings used a calibrated instrument.
- Data is compared to the nearest National Weather Service (NWS) weather data, especially with daily air temperature and precipitation.
- A graphical representation is created to aid in the review.
- Trim continuous data to remove anomalous measurements during deployment, maintenance, and retrieval activities (if applicable).



Continuous data also have additional criteria for assigning a usability score that include, but are not limited to:

- Quality and frequency of probe maintenance during the deployment periods
- Instrument drift
- Potential water level fluctuations and potential out-of-water periods of time
- Probes buried in sediment or sand during deployment
- Instrumental anomalies in the data record that were not censored
- Data records not trimmed to the period of deployment, maintenance, or retrieval

The EN grant funded project to develop a new database will also develop new storage options and efficiencies for reviewing continuous data.

External biological data are rarely submitted, but the review process is similar to the that use for water quality data. The following are additional criteria used to assign a usability score for biological data:

- Taxa level resolution is not fine enough for assessment purposes.
- Lack of voucher sample collection and expert verification.
- Non-standard and/or incomparable field methodology.
- Poor metadata.

Feedback to data providers: WPP provides all external data providers a standard data submission feedback form (see appendix 2). The form details the reasons for the usability scores assigned to the data and general suggestions for potential improvements. The form notes that information on specific improvements should be obtained by the data provider through a virtual or in-person meeting with the External Monitoring and Data Coordinator and other relevant WPP staff.

Data harmonization and export for analyses:

The EDB lacks many common data standardization controls. Thus, data exported from the EDB requires extensive post-processing as part of a post-export harmonization procedure that aligns external data structures with internal data. All harmonization processes utilize R scripts. The following files are required:

1. EDB data table ('EDDMain' table in the EDB)
2. EDB review table ('ExtDataReview' table in the EDB)
3. Organization name and code table
4. Table of detection quantification limits
5. Per-datum flag logic table (see below)
6. Per-datum flag codes list (see below)
7. Tabular data from attribute table of point feature class of validated station definitions
8. Crosswalk table of EDB record numbers to station entry references

The R scripts perform the following actions for the data harmonization process:

1. Join fields across data tables as needed for data exports (e.g., join final Station ID to observations).
2. Standardize analyte names.
3. Remove and/or reconcile non-numeric entries based on the details below.
4. Standardize date and time entries.
5. Add fields needed for data exports that may be populated or left empty for calculations in later steps (e.g., sample month, date-time, and flag code)
6. Convert the field of results to a numeric field.
7. Apply per-datum flags using the per-datum flag logic table (see below).
8. Add standardized organization names and codes.
9. Add a field for depths and perform specific data conversions as needed (in progress).
10. Remove duplicate entries.

Removing and/or reconciling non-numeric result entries is needed to: (a) address censored and flagged per-datum observations; and (b) reconcile result entries indicating the observation is below detection limits or above quantification limits. Appendix 3 includes a list of non-numeric values that should be censored or otherwise flagged due or reconciled to account for detection and quantification limits. WPP's



procedure is to replace minimum detection and upper quantification limit entries with the procedure's limit value. When a "<" or ">" symbol is used to indicate detection and quantification limit entries respectively, the operator is removed and the remaining numeric value is used as the result. Text indicating "below detection" or "above quantification limit" are entered for the result are replaced with values reported for the specific procedure if available from the data provider in the submittal. The per-datum flag logic table references the EDRID number and limit value, and R scripts use logic statements to replace character entries in the data with: (a) the appropriate numeric value or (b) "NA" if the detection or quantification limit is not available (which results in the removal of the entry from the dataset during a downstream harmonization process). Per datum flags are applied through a complex process that references a logic table created for all entries in the EDB. WPP staff manually enter values in the logic table to indicate the dates or months and/or stations that require a flag code applied to each datum. Each EDRID data group can include observations from multiple stations and sample dates/months. An A- or B score indicates that some data can be used with caution or censored. The logic table is created by referencing the notes entered during the review to indicate what specific samples (by date/month or station) should be flagged. The following flags are applied:

Flag code	Explanation
CN	Data censored, remove from data
UC-g	Use with caution because B score (i.e., generic B score)
UC-d	Use data with caution as precision of field duplicates (as RPD) did not meet project data quality objectives identified for program or in QAPP
UC-i	Use with caution because post-survey checks outside typical acceptance ranges at post-field calibration checks.
UC-a	Use with caution due to concerns over accuracy
UC-b	Use with caution due to blank contamination in lab reagent blanks and/or field blank samples (indicating possible bias high and false positives).
UC-f	Use with caution because frequency of QC duplicates did not meet data quality objectives for program or QAPP
UC-p	Use with caution because samples not preserved per SOP or analytical method requirements
UC-u	Use with caution because of unstable probe readings, due to lack of sufficient equilibration time prior to final readings, non-representative location, high-variable water quality conditions.

Once harmonization is complete, the data must be filtered for date, location, usability score, and/or any parameter as needed for exports. R scripts format the data for: (a) raw data output; (b) harmonized data output; and (c) WQP output. Fields are renamed, removed, and/or added as needed for each export type. Data are exported as .csv files with a date stamp affixed to the file name.

Analyses and assessment:

External data are used for multiple programmatic processes but primarily for assessment decision making. The 2024 Consolidated Assessment and Listing Methodology (CALM) guidance document (MassDEPb 2025) describes the current analytical framework for utilizing water quality and biological data for assessment decision making.

Quality Control: No quality control procedures or materials are required for these data management procedures. QAQC is performed on data during the data entry process, and standard data management procedures are followed (e.g., creating copies of databases prior to data entry processes; file backups, etc.). The data review procedure is a QAQC procedure, and it is described above.

Data & Records Management: The only form created for this process is the data feedback form provided to external data providers after data review is complete (see above for details). See appendix 3 for the form.



Data Software: Software used for these procedures includes the Microsoft Office suite with specific need to utilize Access and Excel. The R statistics program and GIS software (recommended to use ESRI ArcGIS Pro) are also required.

References

MassDEP. 2025a. Standard Operating Procedure: Assessment Unit Creation. CN 618.0. Massachusetts Department of Environmental Protection, Bureau of Water Resources, Division of Watershed Management, Watershed Planning Program. Worcester, MA.

MassDEP. 2025b. Massachusetts Consolidated Assessment and Listing Methodology (CALM) Guidance Manual for the 2024 Reporting Cycle. CN 564.0, Massachusetts Department of Environmental Protection, Bureau of Water Resources, Division of Watershed Management, Watershed Planning Program. Worcester, MA.

Smith, R.F. 2025. External Data Guidebook. Massachusetts Department of Environmental Protection, Watershed Planning Program.



Appendices

Appendix 1: WPP uses the waterbody types listed in the 'MonitoringLocationTypeName' field of the EPA's WQX/WQP schema. The CALM guidance document (MassDEP 2025) lists the waterbodies used for assessment. The following is a list of potential aquatic habitats WPP may utilize in the geospatial database as part of the station definition (most are uncommon or not used):

- Atmosphere
- BEACH Program Site-Lake
- BEACH Program Site-Ocean
- BEACH Program Site-River/Stream
- Borehole
- Canal Drainage
- Channelized Stream
- Constructed Water Transport Structure
- Estuary
- Facility Industrial
- Facility Municipal Sewage (POTW)
- Facility Other
- Facility: Diversion
- Facility: Outfall
- Facility: Septic system
- Facility: Storm sewer
- Facility: Wastewater-treatment plant
- Facility: Wastewater sewer
- Facility: Water-distribution system
- Facility: Water-supply treatment plant
- Facility: Water-use establishment
- Lake
- Lake, Reservoir, Impoundment
- Mine/Mine Discharge Adit (Mine Entrance)
- Ocean
- Ocean: Coastal
- Other-Surface Water
- Pipe, Unspecified Source
- Pond
- Pond-Stormwater
- Reservoir
- River/Stream
- River/Stream Intermittent
- River/Stream Perennial
- Riverine Impoundment
- Spring
- Storm Sewer
- Stream
- Stream: Canal
- Stream: Ditch
- Stream: Tidal stream
- Subsurface: Groundwater drain
- Waste Sewer
- Well
- Well: Multiple wells
- Well: Test hole not completed as a well
- Wetland
- Wetland Estuarine-Emergent
- Wetland Estuarine-Forested
- Wetland Estuarine-Marsh
- Wetland Palustrine-Forested
- Wetland Palustrine-Shrub-Scrub
- Wetland Palustrine Pond



Appendix 2: Feedback form template used by WPP to provide feedback to external data providers. Forms are filled out manually by WPP staff after final review of data usability for assessment. The text below is formatted for display purposes in this SOP (i.e., blank spaces removed where text would be added by WPP staff).

Data Review Summary

MassDEP Review Date:

External Data Reviewed:

Parameters Reviewed:

External Contact Information:

Overall Evaluation/Summary

- ☐ All data considered usable by DEP for assessment purposes
- ☐ Some data considered usable by DEP for assessments
- ☐ No data considered usable by DEP for assessments
- ☐ All data considered usable by DEP for screening purposes only
- ☐ Some data considered usable by DEP for screening purposes only
- ☐ Some data usability inconclusive, due to lack of information

Minimum Requirements for Review

- QAPP approved for review period? ☐ Yes ☐ No ☐ Unknown
- SOPs available/sufficient? ☐ Yes ☐ No ☐ Other
- QA/QC data provided? ☐ Yes ☐ No ☐ Other
- Data is clear and organized and/or in template DEP EDD format ☐ Yes ☐ No ☐ Other

Potential improvements for data utility to MassDEP:

- | | | |
|--|--|---|
| <input type="checkbox"/> None | <input type="checkbox"/> Review sampling procedures with volunteers | <input type="checkbox"/> Provide complete SOPs |
| <input type="checkbox"/> Increase sample # | <input type="checkbox"/> Sample earlier in "am" for D.O. | <input type="checkbox"/> Improve attainment of DQOs |
| <input type="checkbox"/> Check equipment/calibration | <input type="checkbox"/> Provide QA/QC data in report | <input type="checkbox"/> Other: |
| <input type="checkbox"/> Include more QA data points | <input type="checkbox"/> Report some indication of success/failure with keeping to holding times | <input type="checkbox"/> Check blank water quality |

Conclusions (Parameter Specific):

Additional notes and potential improvements for data utility to DEP:

If you have any further questions regarding your data submittal and its use for CWA 305(b), 314 and 303(d) assessment and reporting purposes, feel free to reach out to Bob Smith (our External Monitoring and Data Coordinator) at Robert.F.Smith@mass.gov



Appendix 3: The EDB contains numerous non-numeric entries in the result field. These non-numeric entries must be removed and/or reconciled to proceed with quantitative analyses. The table below provides details on the meaning of and action taken for non-numeric entries in the EDB. Many entries match WPP internal data conventions. Reconciling entries most commonly is to account for measures below the detection limit or above the quantification limit.

Non-numeric result entry	Description/meaning	Data action
Values below match internal WPP data conventions		
--	No data (i.e., data not collected/not required/not calibrated for/otherwise not available)	Remove entry
##	Censored data (i.e., data that have been discarded for some reason; check qualifier symbol for cause(s))	Remove entry
**	Missing data (i.e., data that were intended to be collected but were not reported for any reason (e.g. site access issues, broken bottles, probe malfunction, etc.) other than no water)	Remove entry
^^	No water (i.e., a special case of missing data due to dry/no water conditions)	Remove entry
r	Data may not be representative of the waterbody, due to circumstances and/or conditions at the time of sampling (e.g., out-of-water, side channel, backwater, etc.)	Remove entry
Values below are unique to the EDB		
R	Unknown (believed to be same as "r")	Remove entry
-	Same as "--"	Remove entry
---	Same as "--"	Remove entry
N/A	"Not Assessed" (functionally the same as "not reported") and also used as "Not Applicable". Regardless, these are removals.	Remove entry
Na	"Not Assessed" (functionally the same as "not reported") and also used as "Not Applicable". Regardless, these are removals.	Remove entry
not reported	"Not reported"	Remove entry
nr	"Not reported"	Remove entry
NR	"Not reported"	Remove entry
NS	Unknown	Remove entry
X	Unknown	Remove entry
<	Below detection limit. May be matched with a numeric entry.	When matched with a numeric entry, the operator is removed and the numeric entry used. Without a numeric entry, the minimum detection value is entered.
TNTC	"To Numerous To Count", indicating above detection limit.	Use detection and quantification limit table to replace non-numeric entry with numeric entry if available.
MDL	"Method detection limit"	Use detection and quantification limit table to replace non-numeric entry with numeric entry if available.
nd	"Non-Detect". Same as less than method reporting limit.	Use detection and quantification limit table to replace non-



Non-numeric result entry	Description/meaning	Data action
		numeric entry with numeric entry if available.
ND	"Non-Detect". Same as less than method reporting limit.	Use detection and quantification limit table to replace non-numeric entry with numeric entry if available.
Not Detected	"Non-Detect". Same as less than method reporting limit.	Use detection and quantification limit table to replace non-numeric entry with numeric entry if available.
over rating	Above the stream flow rating curve. For measures of discharge.	Remove entry
Present Above Quantification Limit	Above the quantification limit of the method. May refer to discharge measure.	Use detection and quantification limit table to replace non-numeric entry with numeric entry if available. Measures of discharge are removed.
Present Below Quantification Limit	Present but below the quantification limit of the method. May refer to discharge measure.	Use detection and quantification limit table to replace non-numeric entry with numeric entry if available. Measures of discharge are removed.
,	Found within a numeric entry. For all situations where a comma is found in the result entry, the comma is followed by three integers indicating the symbol is meant to designate thousands place.	Symbol removed and the numeric entry used.
*	Found following numeric entry. Reason unknown.	Symbol removed the numeric entry used.