[Running the Data Extract Query 2](#_Toc194494624)

[Water Data Extract 5](#_Toc194494625)

[Post Processing of Water Data Extracts 9](#_Toc194494626)

[Air Data Extract 9](#_Toc194494627)

# Running the Data Extract Query

The data conversion scripts are in GithHub here: <https://github.com/bcgov/nr-enmods-dar/tree/documentation/data%20conversion>

Note that this is currently in a branch. This will ultimately be merged into the documentation folder in the main branch.

You can run the query in your SQL IDE of choice, this documentation assumes the IDE used is SQL Developer.

To run the query:

1. Open a new SQL Window

A screenshot of a computer

AI-generated content may be incorrect.

1. Copy and paste the “data conversion.sql” script into the new SQL Script window.
2. The query is created in such a way that there are two common table expressions (CTE) (“core data” and “sample data”) that are referenced by the queries to extract water, air, etc. The “core data” CTE is the query used to retrieve all sample and related results data. The “sample data” is a query used to just retrieve the sample data. A CTE was used so that the same query could be re-used for water, air, etc.

To reuse the CTEs, the water and air queries are both included in the single query. If you want to just run one query, then comment out the other query.

1. To run the query, press the green “play button”

A screenshot of a computer

AI-generated content may be incorrect.

1. The queries can take a long time to run (around 10 minutes). Exporting all results to a CSV file takes much longer (up to 3 hours for the water extract, less than half an hour for the air extract). To extract the results to a CSV file, right click in the results section and click “Export…”

A screenshot of a computer

AI-generated content may be incorrect.

Ensure that the format is “csv”, give it a path to save the export to, and click next to save the file.

A screenshot of a computer

AI-generated content may be incorrect.

# Water Data Extract

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Source Table.Column or Logic** |
| **Observation ID** | Unique identifier for each observation. Not yet implemented (currently blank). | Always Blank |
| **Ministry Contact** | The name of the ministry staff who oversaw or initiated the sampling event. | print.staffs.first\_name || ' ' || print.staffs.last\_name |
| **Sampling Agency** | The agency responsible for collecting the sample. | ems\_client\_locations.id || ' - ' || ems\_client\_locations.name |
| **Project** | Hardcoded value identifying the project. | 'BCLMN' |
| **Work Order Number** | The requisition or work order under which the sample was collected. | ems\_samples.requisition\_id |
| **Location ID** | Identifier for the monitoring location. | ems\_samples.mon\_locn\_id |
| **Field Visit Start Time** | Earliest start time of the sampling activity at a location on a given day, formatted in ISO8601 with -08:00 offset. | MIN(collection\_start\_date) from ems\_samples using an inline subquery to get the earliest start and latest end times for each MON\_LOCN\_ID per day |
| **Field Visit End Time** | Latest end time of the sampling activity at a location on a given day, unless equal to start time (then null). | MAX(collection\_end\_date) from ems\_samples (inline subquery eal.latestendtime)  if earlieststarttime = latestendtime, don't display the end date and time |
| **Field Visit Participants** | Name or identifier of the individual who performed the sampling. | ems\_samples.sampler |
| **Field Visit Comments** | General comments about the field visit. | ems\_samples.field\_comment AS Activity Comments |
| **Activity Comments** | Duplicate of field comment for internal tracking. | ems\_samples.field\_comment |
| **Field Filtered** | Indicates whether the sample was filtered in the field (always null in current dataset). | Always Blank |
| **Field Filtered Comment** | Comment on field filtering if applicable (always null in current dataset). | Always Blank |
| **Field Preservative** | Describes the chemical or process used to preserve the sample in the field. | ems\_preservatives.description |
| **Field Device ID** | Unique identifier for the field instrument (left blank). | Always Blank |
| **Field Device Type** | Describes the field device or method used (mapped from parameter and method). | OBSERVED\_PROPERTIES\_FOR\_ETL.Device\_Type |
| **Sampling Context Tag** | Additional context for sampling event (left blank). | Always Blank |
| **Collection Method** | Describes how the sample was collected (e.g., Grab, Composite). Complex logic based on ems\_collection\_methods.code. | ems\_collection\_methods.code |
| **Medium** | The environmental medium sampled (e.g., WATER). | ems.ems\_locn\_state\_descriptor\_export\_map.enmods\_medium |
| **Depth Upper** | Upper boundary of the sample depth, in meters. | ems\_samples.depth\_upper |
| **Depth Lower** | Lower boundary of the sample depth, in meters. | ems\_samples.depth\_lower |
| **Depth Unit** | Unit of measure for depth (always 'metre' if upper/lower present). | 'metre' if depth exists else NULL |
| **Observed DateTime** | Date and time the sample was collected, formatted with offset. | ems\_samples.collection\_start\_date  If there's a duplicate, increment the observed date by one minute for each instance of the duplicate. |
| **Observed Date Time End** | Date and time sampling ended, null if same as start. | ems\_samples.collection\_end\_date |
| **Observed Property ID** | Mapped combination of parameter, method, and unit for standardized reporting. | OBSERVED\_PROPERTIES\_FOR\_ETL.NewNameID |
| **Result Value** | Numeric result from field or lab analysis. | ems\_results.result\_numeric |
| **Method Detection Limit** | Lowest concentration reliably detectable; converted to target unit if needed. | ems\_results.method\_detect\_limit or ems\_parm\_dicts.METHOD\_DETECT\_LIMIT |
| **Method Reporting Limit** | Minimum reportable value (not provided, left blank). | NULL |
| **Result Unit** | Unit for result value (e.g., mg/L). | AQS\_UNITS\_TEMP.AQS\_NAME\_ON\_IMPORT |
| **Detection Condition** | Indicates result is below detection limit if result\_letter = '<'. | ems\_results.result\_letter |
| **Limit Type** | Type of limit applied (not used in EMS, left blank). | NULL |
| **Fraction** | Sample fraction type (e.g., total, dissolved), from mapping table. | OBSERVED\_PROPERTIES\_FOR\_ETL.Fraction |
| **Data Classification** | Category of result: FIELD\_RESULT, LAB, QC, etc. | OBSERVED\_PROPERTIES\_FOR\_ETL.Classification |
| **Source of Rounded Value** | Not used in EMS, left blank. | NULL |
| **Rounded Value** | Rounded version of result (not implemented, left blank). | NULL |
| **Rounding Specification** | Specification used to round values (left blank). | NULL |
| **Analyzing Agency** | Agency that performed the lab analysis. | ems\_client\_locations.short\_name |
| **Analysis Method** | Code for the analytical method used to produce the result. | ems\_results.anal\_method\_cd |
| **Analyzed Date Time** | Date sample was analyzed in lab, defaults to observed date if null. | ems\_results.analytical\_date or ems\_samples.collection\_start\_date |
| **Result Status** | Always set to 'Preliminary'. | 'Preliminary' |
| **Result Grade** | Always set to 'Ungraded'. | 'Ungraded' |
| **Activity ID** | Not currently used (left blank). | NULL |
| **Activity Name** | Unique identifier of the sample activity. | ems\_samples.id |
| **Tissue Type** | Type of tissue sampled (blank for water). | ems\_tissue\_types.description |
| **Lab Arrival Temperature** | Temperature at which the sample arrived at the lab. | ems\_samples.lab\_arrival\_temperature |
| **Specimen Name** | Name assigned to biological specimen, or blank if not applicable. | OBSERVED\_PROPERTIES\_FOR\_ETL.OP\_Group.  To introduce uniqueness to each record, -n (where n is the incremented duplicate number starting at 2) is appended to the specimen name in cases where there are multiple records with the same:   * Location ID * Field Visit Start Time * Medium * Depth Upper * Activity Name * Specimen Name * Data Classification * QC Type * Observed Property ID   For example, if there are three rows with the same values as above, then the specimen name would be  {specimen name}  {specimen name}-2  {specimen name}-3  If the data classification is LAB or SURROGATE\_RESULT, and the specimen name is blank, then set the specimen name to "Activity Name" (which is the same as the smpl.id column) |
| **Lab Quality Flag** | Indicator for lab quality issue (left blank). | NULL |
| **Lab Arrival Date and Time** | Timestamp when the sample arrived at the lab. | ems\_samples.lab\_arrival\_date |
| **Lab Prepared DateTime** | Timestamp when the lab prepared the sample (left blank). | NULL |
| **Lab Sample ID** | Lab-generated ID to track the sample. | ems\_results.lab\_sample\_id |
| **Lab Dilution Factor** | Dilution factor used in lab (left blank). | NULL |
| **Lab Comment** | Comment field associated with lab work. | ems\_samples.lab\_comment |
| **Lab Batch ID** | Identifier for the lab batch in which the sample was processed. | ems\_results.lab\_batch\_id |
| **QC Type** | Classifies sample as Replicate, Blank, or Spike based on sample class description. | ems\_sample\_classes.description |
| **QC Source Activity Name** | Name of the activity that the QC sample was based on (left blank). | NULL |
| **Composite Stat** | Statistical method used for composites (not used for lakes, left blank). | NULL |

# Post Processing of Water Data Extracts

The data extract for water is roughly 10GB in size, and thus cannot be opened easily. As such, the following commands are executed (using terminal on a Mac) against the data extract to split the file into multiple files, each with roughly 700,000 records.

To split the file into multiple files:

*split -l 699999 -d {file\_name}.csv split\_part\_*

To add the header row to each file:

*header=$(head -n 1 {file\_name}.csv)*

*for file in split\_part\_\*; do*

*(echo "$header" && cat "$file") > "water\_$file"*

*rm "$file"*

*done*

Note that the first file will have a duplicate header as a result of this, you can manually delete the duplicate header after.

# Air Data Extract

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Source Table.Column or Logic** |
| **Observation ID** | Unique identifier for each observation (placeholder, currently blank). | N/A |
| **Ministry Contact** | Name of the person from the ministry responsible for the sample. | print.staffs.first\_name || ' ' || print.staffs.last\_name |
| **Sampling Agency** | Agency responsible for collecting the sample. | ems\_client\_locations.id || ' - ' || ems\_client\_locations.name |
| **Project** | Fixed project name for this dataset. | 'BCLMN' |
| **Work Order Number** | ID that links this sample to a specific work order or requisition. | ems\_samples.requisition\_id |
| **Location ID** | Monitoring location ID where the sample was collected. | ems\_samples.mon\_locn\_id |
| **Field Visit Start Time** | Start time of the sampling visit, formatted with timezone offset. | MIN(collection\_start\_date) via eal.earlieststarttime |
| **Field Visit End Time** | End time of the sampling visit, omitted if same as start time. | MAX(collection\_end\_date) via eal.latestendtime  Field Visit End Time will be null if it equals Field Visit Start Time. |
| **Field Visit Participants** | Individuals or team who collected the sample. | ems\_samples.sampler |
| **Field Visit Comments** | Comments regarding the field visit. | ems\_samples.field\_comment |
| **Activity Comments** | Duplicate of field visit comments, used internally. | ems\_samples.field\_comment |
| **Field Filtered** | Indicates if field filtration occurred (always null in EMS). | NULL |
| **Field Filtered Comment** | Comment about field filtration (always null in EMS). | NULL |
| **Field Preservative** | Substance used to preserve the sample on site. | ems\_preservatives.description |
| **Field Device ID** | ID of the field instrument (not captured, always null). | NULL |
| **Field Device Type** | Device type or method, mapped via OBSERVED\_PROPERTIES\_FOR\_ETL. | OBSERVED\_PROPERTIES\_FOR\_ETL.Device\_Type |
| **Sampling Context Tag** | Extra tag for context (always null). | NULL |
| **Collection Method** | Sampling method determined from ems\_collection\_methods.code with extensive mapping logic. | ems\_collection\_methods.code |
| **Medium** | Type of medium sampled (e.g., AIR, WATER). | ems\_locn\_state\_descriptor\_export\_map.enmods\_medium |
| **Depth Upper** | Upper depth in meters (if applicable). | ems\_samples.depth\_upper |
| **Depth Lower** | Lower depth in meters (if applicable). | ems\_samples.depth\_lower |
| **Depth Unit** | Unit for depth ('metre' when upper/lower depth provided). | Set to ‘metre’ if either Depth Upper or Depth Lower is present; otherwise NULL. |
| **Observed DateTime** | Date and time sampling began (ISO format with offset). | ems\_samples.collection\_start\_date  If there's a duplicate, increment the observed date by one minute for each instance of the duplicate. |
| **Observed Date Time End** | End time of observation, null if same as start. | ems\_samples.collection\_end\_date |
| **Observed Property ID** | Derived property ID from parameter, method, and unit mapping or labeled values for air. | OBSERVED\_PROPERTIES\_FOR\_ETL.NewNameID or hardcoded |
| **Result Value** | Measured result from sample or field observation. | ems\_results.result\_numeric or smpl.flow, smpl.filter\_size |
| **Method Detection Limit** | Limit of detection, unit converted if needed. | ems\_results.method\_detect\_limit, ems\_parm\_dicts.METHOD\_DETECT\_LIMIT |
| **Method Reporting Limit** | Minimum value to report (blank). | NULL |
| **Result Unit** | Unit for result values (e.g., ug/m3, um). | AQS\_UNITS\_TEMP.AQS\_NAME\_ON\_IMPORT or smpl.flow\_unit\_cd or hardcoded |
| **Detection Condition** | Flag indicating if result is below detection threshold. | ems\_results.result\_letter |
| **Limit Type** | Left blank (not tracked in EMS). | NULL |
| **Fraction** | Part of sample measured (e.g., total, dissolved). | OBSERVED\_PROPERTIES\_FOR\_ETL.Fraction |
| **Data Classification** | Indicates data source: LAB, FIELD\_RESULT, ACTIVITY\_RESULT. | OBSERVED\_PROPERTIES\_FOR\_ETL.Classification or hardcoded |
| **Source of Rounded Value** | Blank (not used in EMS). | NULL |
| **Rounded Value** | Blank (not used in EMS). | NULL |
| **Rounding Specification** | Blank (not used in EMS). | NULL |
| **Analyzing Agency** | Agency that analyzed the sample. | ems\_client\_locations.short\_name |
| **Analysis Method** | Lab method code (null for field/volume observations). | ems\_results.anal\_method\_cd |
| **Analyzed Date Time** | Date when sample was analyzed, or fallback to observed date. | ems\_results.analytical\_date or smpl.collection\_start\_date |
| **Result Status** | Set to 'Preliminary'. | 'Preliminary' |
| **Result Grade** | Set to 'Ungraded'. | 'Ungraded' |
| **Activity ID** | Blank (not extracted from EMS). | NULL |
| **Activity Name** | Sample ID representing the activity. | [ems\_samples.id](http://ems_samples.id) |
| **Tissue Type** | Type of biological tissue sampled, if applicable. | ems\_tissue\_types.description |
| **Lab Arrival Temperature** | Temperature at lab intake. | ems\_samples.lab\_arrival\_temperature |
| **Specimen Name** | Group/specimen label for biological or field classification. | OBSERVED\_PROPERTIES\_FOR\_ETL.OP\_Group  To introduce uniqueness to each record, -n (where n is the incremented duplicate number starting at 2) is appended to the specimen name in cases where there are multiple records with the same:   * Location ID * Field Visit Start Time * Medium * Depth Upper * Activity Name * Specimen Name * Data Classification * QC Type * Observed Property ID   For example, if there are three rows with the same values as above, then the specimen name would be  {specimen name}  {specimen name}-2  {specimen name}-3  If the data classification is LAB or SURROGATE\_RESULT, and the specimen name is blank, then set the specimen name to "Activity Name" (which is the same as the smpl.id column) |
| **Lab Quality Flag** | Quality control flags from lab (not implemented). | NULL |
| **Lab Arrival Date and Time** | Timestamp for when the lab received the sample. | ems\_samples.lab\_arrival\_date |
| **Lab Prepared DateTime** | Blank (not tracked). | NULL |
| **Lab Sample ID** | ID used to identify sample in the lab. | ems\_results.lab\_sample\_id |
| **Lab Dilution Factor** | Blank (not captured in EMS). | NULL |
| **Lab Comment** | Lab notes or comments about the sample. | ems\_samples.lab\_comment |
| **Lab Batch ID** | Identifier for the lab batch. | ems\_results.lab\_batch\_id |
| **QC Type** | Sample classification: Replicate, Blank, Spike, or empty. | ems\_sample\_classes.description |
| **QC Source Activity Name** | Blank (not linked). | NULL |
| **Composite Stat** | Blank, not tracked for air samples. | NULL |
| **Observed Property ID** | Hardcoded as 'Air Volume (vol.)'. (Used in: Air Flow Volume section). | 'Air Volume (vol.)' |
| **Result Value** | Flow volume from the air sample. (Used in: Air Flow Volume section). | ems\_samples.flow |
| **Result Unit** | Flow volume unit from measurement units. (Used in: Air Flow Volume section). | ems\_measurment\_units.short\_name |
| **Observed Property ID** | Hardcoded as 'Air Filter Size (len.)'. (Used in: Air Filter Size section). | 'Air Filter Size (len.)' |
| **Result Value** | Air filter size from the sample. (Used in: Air Filter Size section). | ems\_samples.filter\_size |
| **Result Unit** | Hardcoded unit 'um' for filter size. (Used in: Air Filter Size section). | 'um' |

# Continuous Data

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Source Table.Column or Logic** |
| **Observation ID** | Unique identifier for each observation (placeholder, currently blank). | N/A |
| **Ministry Contact** | Name of the person from the ministry responsible for the sample. | print.staffs.first\_name || ' ' || print.staffs.last\_name |
| **Sampling Agency** | Agency responsible for collecting the sample. | ems\_client\_locations.id || ' - ' || ems\_client\_locations.name |
| **Project** | Fixed project name for this dataset. | 'BCLMN' |
| **Work Order Number** | ID that links this sample to a specific work order or requisition. | ems\_samples.requisition\_id |
| **Location ID** | Monitoring location ID where the sample was collected. | ems\_samples.mon\_locn\_id |
| **Field Visit Start Time** | Start time of the sampling visit, formatted with timezone offset. | MIN(collection\_start\_date) via eal.earlieststarttime |
| **Field Visit End Time** | End time of the sampling visit, omitted if same as start time. | MAX(collection\_end\_date) via eal.latestendtime  Field Visit End Time will be null if it equals Field Visit Start Time. |
| **Field Visit Participants** | Individuals or team who collected the sample. | ems\_samples.sampler |
| **Field Visit Comments** | Comments regarding the field visit. | ems\_samples.field\_comment |
| **Activity Comments** | Duplicate of field visit comments, used internally. | ems\_samples.field\_comment |
| **Field Filtered** | Indicates if field filtration occurred (always null in EMS). | NULL |
| **Field Filtered Comment** | Comment about field filtration (always null in EMS). | NULL |
| **Field Preservative** | Substance used to preserve the sample on site. | ems\_preservatives.description |
| **Field Device ID** | ID of the field instrument (not captured, always null). | NULL |
| **Field Device Type** | Device type or method, mapped via OBSERVED\_PROPERTIES\_FOR\_ETL. | OBSERVED\_PROPERTIES\_FOR\_ETL.Device\_Type |
| **Sampling Context Tag** | Extra tag for context (always null). | NULL |
| **Collection Method** | Sampling method determined from ems\_collection\_methods.code with extensive mapping logic. | ems\_collection\_methods.code |
| **Medium** | Type of medium sampled (e.g., AIR, WATER). | ems\_locn\_state\_descriptor\_export\_map.enmods\_medium |
| **Depth Upper** | Upper depth in meters (if applicable). | ems\_samples.depth\_upper |
| **Depth Lower** | Lower depth in meters (if applicable). | ems\_samples.depth\_lower |
| **Depth Unit** | Unit for depth ('metre' when upper/lower depth provided). | Set to ‘metre’ if either Depth Upper or Depth Lower is present; otherwise NULL. |
| **Observed DateTime** | Date and time sampling began (ISO format with offset). | ems\_samples.collection\_start\_date  If there's a duplicate, increment the observed date by one minute for each instance of the duplicate. |
| **Observed Date Time End** | End time of observation, null if same as start. | ems\_samples.collection\_end\_date |
| **Observed Property ID** | Derived property ID from parameter, method, and unit | OBSERVED\_PROPERTIES\_FOR\_ETL.NewNameID |
| **Result Value** | Measured result from sample or field observation. | ems\_results.result\_numeric or smpl.flow, smpl.filter\_size |
| **Method Detection Limit** | Limit of detection, unit converted if needed. | ems\_results.method\_detect\_limit, ems\_parm\_dicts.METHOD\_DETECT\_LIMIT |
| **Method Reporting Limit** | Minimum value to report (blank). | NULL |
| **Result Unit** | Unit for result values (e.g., ug/m3, um). | AQS\_UNITS\_TEMP.AQS\_NAME\_ON\_IMPORT or smpl.flow\_unit\_cd or hardcoded |
| **Detection Condition** | Flag indicating if result is below detection threshold. | ems\_results.result\_letter |
| **Limit Type** | Left blank (not tracked in EMS). | NULL |
| **Fraction** | Part of sample measured (e.g., total, dissolved). | OBSERVED\_PROPERTIES\_FOR\_ETL.Fraction |
| **Data Classification** | Indicates data source: LAB, FIELD\_RESULT, ACTIVITY\_RESULT. | OBSERVED\_PROPERTIES\_FOR\_ETL.Classification or hardcoded |
| **Source of Rounded Value** | Blank (not used in EMS). | NULL |
| **Rounded Value** | Blank (not used in EMS). | NULL |
| **Rounding Specification** | Blank (not used in EMS). | NULL |
| **Analyzing Agency** | Agency that analyzed the sample. | ems\_client\_locations.short\_name |
| **Analysis Method** | Lab method code (null for field/volume observations). | ems\_results.anal\_method\_cd |
| **Analyzed Date Time** | Date when sample was analyzed, or fallback to observed date. | ems\_results.analytical\_date or smpl.collection\_start\_date |
| **Result Status** | Set to 'Preliminary'. | 'Preliminary' |
| **Result Grade** | Set to 'Ungraded'. | 'Ungraded' |
| **Activity ID** | Blank (not extracted from EMS). | NULL |
| **Activity Name** | Sample ID representing the activity. | [ems\_samples.id](http://ems_samples.id) |
| **Tissue Type** | Type of biological tissue sampled, if applicable. | ems\_tissue\_types.description |
| **Lab Arrival Temperature** | Temperature at lab intake. | ems\_samples.lab\_arrival\_temperature |
| **Specimen Name** | Group/specimen label for biological or field classification. | OBSERVED\_PROPERTIES\_FOR\_ETL.OP\_Group  To introduce uniqueness to each record, -n (where n is the incremented duplicate number starting at 2) is appended to the specimen name in cases where there are multiple records with the same:   * Location ID * Field Visit Start Time * Medium * Depth Upper * Activity Name * Specimen Name * Data Classification * QC Type * Observed Property ID   For example, if there are three rows with the same values as above, then the specimen name would be  {specimen name}  {specimen name}-2  {specimen name}-3  If the data classification is LAB or SURROGATE\_RESULT, and the specimen name is blank, then set the specimen name to "Activity Name" (which is the same as the smpl.id column) |
| **Lab Quality Flag** | Quality control flags from lab (not implemented). | NULL |
| **Lab Arrival Date and Time** | Timestamp for when the lab received the sample. | ems\_samples.lab\_arrival\_date |
| **Lab Prepared DateTime** | Blank (not tracked). | NULL |
| **Lab Sample ID** | ID used to identify sample in the lab. | ems\_results.lab\_sample\_id |
| **Lab Dilution Factor** | Blank (not captured in EMS). | NULL |
| **Lab Comment** | Lab notes or comments about the sample. | ems\_samples.lab\_comment |
| **Lab Batch ID** | Identifier for the lab batch. | ems\_results.lab\_batch\_id |
| **QC Type** | Sample classification: Replicate, Blank, Spike, or empty. | ems\_sample\_classes.description |
| **QC Source Activity Name** | Blank (not linked). | NULL |
| **Composite Stat** | Blank, not tracked for air samples. | Composite Stat will be populated with ‘Mean’, ‘Maximum’, or ‘Minimum’ based on the source value (CONTINUOUS\_AVERAGE, CONTINUOUS\_MAXIMUM, CONTINUOUS\_MINIMUM). Otherwise, it will be blank.” |