

Chemistry Data – Information Sheet

Author: Sonya Havens, Chris Hay
Last Update: February 8, 2023 (CH)

General

Chemistry data from the Experimental Lakes Area (ELA) has been collected across ELA lakes since May 1968. Water samples are collected and analyzed for a variety of parameters, for both specific research projects and long-term ecological research, and across manipulated lakes or enclosures and reference lakes.

The management of the chemistry dataset has periodically changed since its inception, as would be expected with technological advances over the years. Currently, all chemistry data undergo ETL (extract, transform, load) from source databases into a PostgreSQL master database. This represents a merging of older chemistry data 1968 through 2016 that were compiled into a Microsoft Access database and newer data 2017 to present which are managed in Sample Master (a Laboratory Information Management System).

The naming and formatting of fields is usually based on the DataStream Open Data Schema for Water Quality Data, which is itself based on the United States Environmental Protection Agency's "Water Quality eXchange" (WQX) (DataStream 2022; United States Environmental Protection Agency 2022).

List of master database fields (columns)

Alphabetically and broken into categories, here are the complete field names and data types as they are stored in the PostgreSQL master database:

Sample and Result Related

- activity_media_name (varchar)
- characteristic_name (varchar)
- characteristic_name_long (varchar)
- digital_object_identifier (varchar)
- field_comment (varchar)
- laboratory_name (varchar)
- laboratory_sample_id (varchar)
- method_speciation (varchar)
- precip_rain_gauge_mm (varchar)
- result_analytical_method_instrument (varchar)
- result_analytical_method_name (varchar)
- result_analytical_reference_method (varchar)
- result_comment (varchar)
- result_detection_quantitation_limit_measure (varchar)
- result_detection_quantitation_limit_type (varchar)

result_detection_quantitation_limit_unit (varchar)
result_sample_fraction (varchar)
result_status_id (varchar)
result_unit (varchar)
result_value (numeric)
activity_end_date (date)

Location

activity_depth_height_measure (numeric)
activity_depth_height_unit (varchar)
layer_collection_end_depth (numeric)
layer_collection_start_depth (numeric)
monitoring_location_name (varchar)
site_code (varchar)

Temporal

activity_end_time (time without time zone)
activity_start_date (date)
activity_start_time (time without time zone)
analysis_start_date (date)

Internal Data Management

dataset_name (varchar)
loader_version (varchar)
monitoring_location_id (uuid)
update_account (varchar)
update_date (date)

Data Dictionary (description for each field)

Each field has been categorized and listed within each category alphabetically. The field name is followed by a description, potential or example values, and any other important notes.

Sample and Result Related

- activity_media_name (varchar)
 - Describes what type of material was sampled.
 - Typically “Lake Water”, though other options include “Stream Water”, “Periphyton”, “Precipitation”, “Enclosure Water”, and more.
- characteristic_name (varchar)
 - A short name for the parameter (chemistry or other) that was tested (e.g., Ca, A254, PartP, etc.).
 - This field has a number of other fields that directly correspond based on the parameter – see the lookup table below for a complete list (but note that this has only been confirmed as complete and correct for 2017 to Present data).

characteristic_name	characteristic_name_long	method_speciation	result_sample_fraction	result_unit
A254	Specific UV Absorbance at 254 nm		Filterable	m ⁻¹
ALK	Alkalinity, total		Unfiltered	ueq/L

characteristic_name	characteristic_name_long	method_speciation	result_sample_fraction	result_unit
CA	Calcium	as Ca	Filterable	mg/L
CHLA	Chlorophyll a		Suspended	ug/L
CL	Chloride	as Cl	Filterable	mg/L
COLOUR	Colour		Filterable	
COND	Conductivity		Unfiltered	uS/cm
DIC	Inorganic carbon		Dissolved	umol/L
DOC	Organic carbon		Filterable	umol/L
E2:E3	Ratio of UV absorbance at 250 nm to 365 nm		Filterable	
E2:E4	Ratio of UV absorbance at 255 nm to 436 nm		Filterable	
FE	Iron	as Fe	Filterable	mg/L
K	Potassium	as K	Filterable	mg/L
MG	Magnesium	as Mg	Filterable	mg/L
MN	Manganese	as Mn	Filterable	mg/L
NA	Sodium	as Na	Filterable	mg/L
NH3	Ammonia	as N	Filterable	ug/L
NO2	Nitrite	as N	Filterable	ug/L
NO3	Nitrate	as N	Filterable	ug/L
O2	Dissolved oxygen (DO)	as O2	Dissolved	mg/L
PARTC	Carbon	as C	Non-Filterable (Particle)	ug/L
PARTFE	Iron	as Fe	Non-Filterable (Particle)	ug/L
PARTN	Nitrogen	as N	Non-Filterable (Particle)	ug/L
PARTP	Phosphorus	as P	Non-Filterable (Particle)	ug/L
PH	pH		Unfiltered	
S275-295	Log transformed spectral slope of UV absorbance from 275 nm to 295 nm		Filterable	
S350-400	Log transformed spectral slope of UV absorbance from 350 nm to 400 nm		Filterable	
SO4	Sulfate	as SO4	Filterable	mg/L
SR	Spectral ratio		Filterable	
SRSI	Silica, reactive	as Si	Dissolved	mg/L
SUVA	Specific UV Absorbance at 254 nm		Filterable	L/μmol·m
TDN	Total Nitrogen, mixed forms	as N	Dissolved	ug/L
TDP	Total Phosphorus, mixed forms	as P	Dissolved	ug/L
TDS	Total dissolved solids		Filterable	mg/L

characteristic_name	characteristic_name_long	method_speciation	result_sample_fraction	result_unit
TN	Total Nitrogen, mixed forms	as N	Unfiltered	ug/L
TP	Total Phosphorus, mixed forms	as P	Unfiltered	ug/L
TSS	Total suspended solids		Suspended	mg/L
TURBIDITY	Turbidity		Suspended	NTU

- **characteristic_name_long** (varchar)
 - A longer and more descriptive name for the parameter than “characteristic_name” provides. When possible, based on consistent nomenclature (DataStream 2022, United States Environmental Protection Agency 2022).
 - See the lookup table under “characteristic_name” for a complete list (for 2017 to Present)
- **digital_object_identifier** (varchar)
 - DOI associated with the data record, when the data themselves have been published to a repository.
 - Presently only “10.25976/r3kp-7m22” for certain stream chemistry data.
 - Future plans are to make more chemistry data available in public repositories, at which point DOIs will be noted for those records.
- **field_comment** (varchar)
 - Any kind of comment from when in the field, related to acquiring the sample (vs. result_comment, which is noted at a later stage).
 - May be an important caveat to analysis of the data (e.g., notes about the rain gauge overflowing, which affects accuracy of precip_rain_gauge_mm value).
- **laboratory_name** (varchar)
 - There are over 100 distinct values, and it appears the use of this field has not been consistent. It may be a name of an individual, institution, or location. In recent years it is typically bcooney – referring to an IISD-ELA Chemistry Research Assistant.
 - Note that although many samples are analyzed by IISD-ELA, samples are also shipped to external labs for some analyses (e.g., The University of Alberta Biogeochemical Analytical Service Laboratory).
- **laboratory_sample_id** (varchar)
 - A unique code for each sample (at least for 2017 to present data), though note that each sample is typically analyzed for multiple parameters.
 - Reflects the date the sample was logged in (e.g., 2022051101-01, format = yyyyymmdd##-##) (again, at least for 2017 to Present data).
- **method_speciation** (varchar)
 - Describes what the result_value actually represents for the characteristic_name.
 - For example, NO3 is actually NO3-N (2017 to Present)
 - See the lookup table under “characteristic_name” for a complete list (for 2017 to Present)
- **precip_rain_gauge_mm** (varchar)
 - For precipitation samples, the millimetre reading from the rain gauge collecting the precipitation.

- Is affected by heaviness of precipitation and length of time precipitation was collected (note the activity_start/end_date/time values).
- Overflow scenario:
 - Sometimes the rain gauge overflows and the true value is unknown – in these cases a ">" (greater-than) sign is used.
 - Sometimes, but not necessarily always, a related comment may be included in the field_comment field, describing the overflow.
 - Due to the potential ">" value, the data type for this field is varchar, not numeric.
 - Data users will need to decide what to do given this information.
 - This was discussed with Sonya Havens (IISD-ELA Research Chemist) 2022-05-31.
- result_analytical_method_instrument (varchar)
 - The instrument (tool) used to analyze the sample (e.g., Dionex DX-600 Ion Chromatography, Fisher Scientific Accumet pH meter, Cole-Parmer 2100 Spectrophotometer, et.c.)
 - Instrumentation cannot be assumed to be the same within certain time periods or characteristic_name values. For example, it is possible that two different instruments may be used in a given year, if one breaks down and an alternate is needed. The instrument used is recorded for every sample or test, so be aware that it may change from record to record (2017 to present).
- result_analytical_method_name (varchar)
 - The short name for the analytical method used (e.g., ELA-PartP, UA-CISO4, etc.).
 - The prefix is the specific institution associated with the method (e.g., ELA = Experimental Lakes Area, UA = University of Alberta).
- result_analytical_reference_method (varchar)
 - When users of IISD-ELA Chemistry data publish research based on these data, they need to cite how the analyses were conducted. Currently, the only way to cite our methods is to cite "The Chemical Analysis of Fresh Water, 1977, Second Edition" (Stainton et al. 1977).
 - "The Chemical Analysis of Fresh Water, Third Edition" is in the process of being published (potentially some time in 2022 or 2023). The intention is that this will be updated and published on a more frequent basis. The 'Third Edition' is already outdated before it has even been published, because instruments have changed for some analyses in 2022 (i.e., TDP, alkalinity, pH, conductivity, and turbidity).
 - The edition cited depends on the date the sample was collected (activity_start_date plus one year – typically samples are analyzed a year after collection – this is imperfect, but due to the actual analysis date field having some issues, this was the best alternative at the time)
 - 1968 – 1976
 - No citation provided (unknown)
 - 1977 – 2014
 - The Chemical Analysis of Fresh Water - Second Edition, 1977
 - 2015 – 2021
 - The Chemical Analysis of Fresh Water - Third Edition, 202X
 - 2022 - 2022
 - The Chemical Analysis of Fresh Water - Third Edition, 202X (except for TDP, alkalinity, pH, conductivity, and turbidity)

- This will need to be reviewed each year to determine how to cite the newest data
- result_comment (varchar)
 - Notes about the result – rarely used, seemingly noted at later analysis or quality control stages (vs. field_comment, which was collected while in the field).
 - E.g., “Trend analysis shows the result to be an outlier and possibly erroneous.” Or “This value is the average of two instrument duplicate measurements which were not within 5% (281 vs 265 ug/L). The logic is that assuming one of these is the true value, the average value will be within 3% of it.”
- result_detection_quantitation_limit_measure (varchar)
 - The result detection limit value (note the associated unit is in the field “result_detection_quantitation_limit_measure”)
 - 1968 through 2016
 - We are lacking documentation on how detection limits were calculated, and it is unlikely we will be able to obtain any more information in the future.
 - 2017 to Present
 - Initially average detection limits were calculated and applied as annual averages per parameter (characteristic_name). However we are working to switch these to individual detection limits per analytical run (aiming to be changed some time in 2023). Given the finer scale of analytical run detection limits, the data user can average per year or summarize whichever way they deem fit.
 - Whether a result value is above or below detection is determined at the time of data entry.
- result_detection_quantitation_limit_type (varchar)
 - For all IISD-ELA Chemistry data (at least 2017 to Present), it is always “Method Detection Level” (United States Environmental Protection Agency 2017; i.e., “Limit of Detection”).
 - That is, not the “Limit of Quantitation” which is 3.3 times the “Limit of Detection”.
- result_detection_quantitation_limit_unit (varchar)
 - A field for the unit associated with the “result_detection_quantitation_limit_measure”
 - Always identical to the result_unit value (at least for 2017 to Present), as discussed with Sonya Havens (IISD-ELA Research Chemist) 2022-05-06.
 - See the lookup table under “characteristic_name” for a complete list (for 2017 to Present)
- result_sample_fraction (varchar)
 - Fraction of sample associated with result (United States Environmental Protection Agency 2017).
 - See the lookup table under “characteristic_name” for a complete list (for 2017 to Present).
 - Currently used options: Dissolved, Filterable, Non-Filterable (Particle), Suspended, Unfiltered (or null)
- result_status_id (varchar)
 - Provides a level of quality control that the record has undergone (thus, “trust” in the data). Note that this only applies for data 2017 to Present. There are three levels.
 - Options: Preliminary, Accepted, Final, Not noted, or null
 - Preliminary: Samples that have passed the analytical quality control parameters (samples that failed are not entered into the database). Quality control parameters for each analysis will be available in ‘The Chemical Analysis of Fresh

Water - Third Edition'. This step is benchmark quality control (e.g., blanks, duplicates, etc.). Internally this corresponds with level result_status_id 1 in Sample Master and "Entered".

- Accepted: Samples that have undergone a second level of quality control. That is, records that have passed trend analysis. Trend analysis is conducted for each parameter at each site to look for outliers and then investigate the outliers. This step is done internally using an R script for efficiency. For more information, see The Chemical Analysis of Fresh Water - Third Edition'. Internally this corresponds with level result_status_id 2 in Sample Master and "Validated".
- Final: Samples that have undergone external peer-review. Currently IISD-ELA has not yet defined a process to evaluate whether a record has been peer-reviewed, but there are plans to do this and use this designation in the future. Internally this corresponds with level result_status_id 3 in Sample Master and "Approved".
- Not noted or null: Not noted, for all pre-2017 records. Records with result_status_id 0 or null within 2017 to present are not written to the database as the result value has not been entered yet (though other parameters for a record may be filled out ahead of time in Sample Master).
- Note that IISD-ELA's quality control management extends beyond the factors considered for this field. For example, IISD-ELA carries out benchtop quality control practices, which are stipulated in each test's standard operating procedure.
- result_unit (varchar)
 - The unit of measure associated with the "result_value"
 - See the lookup table under "characteristic_name" for a complete list (for 2017 to Present)
- result_value (numeric)
 - The actual measured number for the parameter – the result (for associated unit, see "result_unit")

Location

- activity_depth_height_measure (numeric)
 - The depth the sample was collected from the surface of the lake, when the sample is collected at a specific numeric level (e.g., 5; i.e., "lake profile sampling" described in the lake sampling & field observation information sheet)
 - For the associated unit, see "activity_depth_height_unit"
 - If the sample represents a lake stratification layer (e.g., epilimnion; i.e., "integrated water sampling" from the lake sampling & field observation information sheet), this field is left as null and the layer (e.g. "epi") is indicated in "activity_depth_height_unit"
- activity_depth_height_unit (varchar)
 - The unit associated with the depth value ("activity_depth_height_measure"), e.g., "m" for metres, if the depth value was a specific numeric level ("lake profile sampling" described in the lake sampling & field observation information sheet)
 - If the sample is not from a numeric depth, but from a layer of the lake ("integrated water sampling" from the lake sampling & field observation information sheet), this field will note the stratification layer (epilimnion, metalimnion, hypolimnion) or a "surface grab" if the sample was collected from reaching an arm into the water
 - Options (2017 to Present): epi, meta, hypo, m, surface grab

- Records pre-2017 may contain other unusual units (data require cleanup)
- layer_collection_end_depth (numeric)
 - Along with “layer_collection_start_depth”, this field is only used for 2017 to Present data and only when the sample was collected from a lake stratification layer (e.g., epilimnion; “integrated water sampling” from the lake sampling & field observation information sheet) to specify the range of the layer that the sample was collected within
 - The unit is always metres
 - E.g., For a record with layer_collection_start_depth = 1.50, layer_collection_end_depth = 5.00, and activity_depth_height_unit = epi, this indicates an integrated sample from an epilimnion layer defined as 1.5m to 5m depth from the surface of the lake
- layer_collection_start_depth (numeric)
 - Along with “layer_collection_end_depth”, this field is only used for 2017 to Present data and only when the sample was collected from a lake stratification layer (e.g., epilimnion; “integrated water sampling” from the lake sampling & field observation information sheet) to specify the range of the layer that the sample was collected within
 - The unit is always metres
 - E.g., For a record with layer_collection_start_depth = 1.50, layer_collection_end_depth = 5.00, and activity_depth_height_unit = epi, this indicates an integrated sample from an epilimnion layer defined as 1.5m to 5m depth from the surface of the lake
- monitoring_location_name (varchar)
 - Describes the geographic location the sample was collected, through combining three parts: location, sublocation, station
 - E.g., 239 LA CB, which means watershed for lake 239, LA indicating the lake area in the watershed, and CB indicating the centre buoy in the lake
 - For more information, see monitoring location documentation
- site_code (varchar)
 - The original site code from the study, which may differ from the standardized monitoring_location_name designation.
 - E.g., “Control R1” vs. “114 ENCLP Metformin-01”

Temporal

These fields are self explanatory. Note the distinction between “activity” (i.e., collecting the sample in the field) vs. “analysis” (laboratory measurement of sample). Sometimes only dates are provided and not times if that degree of fine scale temporal measurement would be unjustified. Sometimes only the start value is provided, and not the end, where it is assumed the event is a short moment or the length of time is not important (e.g. collecting a lake sample on a certain date, v.s. precipitation collected in a rain gauge where the length of time is more important).

Note that currently “analysis_start_date” is facing some data quality uncertainties, so should be used with caution. Work is underway to correct this.

- activity_end_date (date)
- activity_end_time (time without time zone)
- activity_start_date (date)
- activity_start_time (time without time zone)
- analysis_start_date (date)

Internal Data Management

- dataset_name (varchar)
- loader_version (varchar)
- monitoring_location_id (uuid)
- update_account (varchar)
- update_date (date)

References

DataStream. 2022. DataStream Open Data Schema for Water Quality Data – JSON Schema.

<https://github.com/datastreamapp/schema>

Stainton, M.P., M.J. Capel, and F.A.J. Armstrong. 1977. The Chemical Analysis of Fresh Water, 2nd ed. Fish. Mar. Serv. Misc. Spec. Publ. 25:116 p.

<https://waves-vagues.dfo-mpo.gc.ca/Library/110147.pdf>

Havens, S.M., M.P. Stainton, and B.C. Cooney. (In review). The Chemical Analysis of Fresh Water – Third Edition.

United States Environmental Protection Agency. 2017. WQXWeb Template Dictionary v3.xls (Sheet: WQX Web Allowable Values) from wqxweb3_physical_chemical_package_0.zip on page Water Quality Exchange Web Template Files. Last saved 2017-07-06.

<https://www.epa.gov/waterdata/water-quality-exchange-web-template-files>

United States Environmental Protection Agency. 2022. Characteristic.CSV. from WQX Domain Value Services and Downloads. Last modified 2022-05-31.

<https://www.epa.gov/waterdata/storage-and-retrieval-and-water-quality-exchange-domain-services-and-downloads>