|  |
| --- |
| **Data Dictionary – Drinking Water** |
| Provided by CDC’s Environmental Public Health Tracking Program |
| February 2023  CDC Lead: T.J. Pierce (pwc2@cdc.gov)  State Lead: Peter DiPippo ([peter.dipippo@health.ri.gov](mailto:peter.dipippo@health.ri.gov)) |  |

PURPOSE AND USE OF THIS DOCUMENT

This data set contributes to the Environmental Public Health Tracking Network. The EPHT cooperative agreement states that “by September 30, 2008 […all grantees must] track and make available core environmental health tracking measures on the State and National EPHT Network […including …] data/information on key water contaminants, as defined through the Content workgroup process.” The Content Workgroup Water Team identified initial contaminants of concern for the national EPHT program, identified nationally consistent data sources, and developed nationally consistent indicators and measures. This data set can be used to calculate the nationally consistent measures for the initial contaminants of concern.

This data set contains the information needed to calculate Environmental Public Health Tracking (EPHT) measures of contaminants in public water supply for arsenic, disinfection byproducts, nitrates, atrazine, di(2-ethylhexyl) phthalate (DEHP), radium, tetrachloroethene (tetrachloroethylene) (PCE), trichloroethene (trichloroethylene (TCE), and uranium. Data are derived from state Safe Drinking Water Act databases. The data set consists of two tables:

1. PWS Inventory. This file is required and contains descriptive and locational information about each public water system (PWS) with which water quality data is provided. This dataset should only include Community Water Systems (CWS) as defined and regulated by the Safe Drinking Water Act. It does not include Non-Transient Non-Community (NTNC) and Transient Non-Community water systems (TNC). There is one record for every year that a CWS was active, delivering drinking water to customers, and in which water quality data is complete. CWS that were once active and are currently inactive should be included if State's data support this scenario.

2. Drinking Water Quality Sampling Results. This file is required and contains one record for each community water system (CWS) for the mean and maximum concentrations per year of each of arsenic, disinfection byproducts (HAA5 and TTHM), nitrates, atrazine, di(2-ethylhexyl) phthalate (DEHP), radium, tetrachloroethene (tetrachloroethylene) (PCE), trichloroethene (trichloroethylene) (TCE), and uranium; and the mean concentrations per quarter of disinfection byproducts, nitrates and atrazine.

Page Break

DATA DICTIONARY

**PWS Inventory (all variables are required)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Header** |  |  |  |
| **Field Name/Schema Name** | **Field Description** | **Format** | **Allowed Values** |
| StateFIPSCode | State FIPS code | AN(2) | FIPS State Code |
| **Table Core** |  |  |  |
| **Schema Name** | **Field Description** | **Format** | **Allowed Values** |
| PWSIDNumber | PWS identifier | AN(9) | nine-character value consisting of the 2-letter state abbreviation followed by 7 numbers |
| YearAssociatedTo | Year data are associated to with regards to sampling results | Text(4) | YYYY. 1999 through latest complete year (e.g., 2021) |
| YearPulled | Year that these data were pulled from state records | Text(4) | YYYY. 1999 through latest year. |
| PWSName | Name of PWS | AN(40) | Any; “U” = Unknown; “NS” = Not submitted |
| PrincipalCountyServed FIPS | Principal county FIPS served by the CWS | AN(5) | Any; “U” = Unknown; “NS” = Not submitted |
| PrincipalCityFeatureID | Principal city, town or village Feature ID served by the CWS | N(10) | 9999999999;”-999” for Missing; “-888” for Not Submitted  Feature ID can be obtained from: <http://geonames.usgs.gov/domestic/download_data.htm> |
| TotalConnections | Number of residential service connections | N(7) | 1-9999999” |
| SystemPopulation | Permanent population uniquely served by the CWS | N(8) | 10-99999999 ” |
| PrimarySourceCode | Type of water source | AN(3) | GU = ground water under direct influence of surface water, GUP = purchased ground water under direct influence of surface water, GW = ground water, GWP = purchased ground water, SW = surface water, SWP = purchased surface water; “U” = Unknown; “NS” = Not submitted |
| Latitude | Latitude in NAD83 decimal degrees describing approximate center of retail service area of water system | N(10) | 00.0000000 to 90.000000;”-99.99” for Missing; “-88.88” for Not Submitted. |
| **Schema Name** | **Field Description** | **Format** | **Allowed Values** |
| Longitude | Longitude in NAD83 decimal degrees describing approximate center of retail service area of water system | N(11) | -180.000000 to 180.000000;”-999” for Missing; “-888” for Not Submitted. |
| LocationDerivationCode | Code describing how approximate latitude/longitude location was derived | AN(3) | SA = Service area polygon centroid; MFL = Mean of 1 or more facility locations that are expected to be proximate to service area extent; PCS = GNIS coordinates for Principal City Served; GSH = The geocoded address of water system headquarters; PNS - GNIS coordinates for Principal County Served; O= Other (e.g., zip code, etc.)   “-999” = Missing; “-888” = Not Submitted;  (See “Appendix A. Service Area Location Derivation Guidance of the How-To Guide” on EPHTN Share Point site for more information & guidance for deriving water system locations.) |

Page Break**Drinking Water Quality Sampling Results**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Header** |  |  |  |  |  |
| **Field Name/Schema Name** | **Field Description** | **Schema Group** | **Format** | **Allowed Values** | **Required (Y/N)** |
| StateFIPSCode | State FIPS Code | NA | AN(2) | FIPS State Code | Y |
| **Table Core** |  |  |  |  |  |
| **SchemaName** | **Field Description** | **Schema Group** | **Format** | **Allowed Values** | **Required (Y/N)** |
| PWSIDNumber | PWS identifier | NA | AN(9) | Nine-character value consisting of the 2-letter state abbreviation followed by 7 numbers | Y |
| Year | Year | NA | Text(4) | YYYY; 1999 through latest complete year (e.g., 2011) | Y |
| AnalyteCode | USEPA Analyte code for required constituents (arsenic, nitrate, TTHM, HAA5, atrazine, PCE, TCE, DEHP, radium, and uranium) | NA | N(4) | 1005=Arsenic; 2050=Atrazine; 2456=HAA5; 2950=TTHM; 2039=DEHP; 1040=Nitrate; 2987=PCE; 2984=TCE; 4010=Combined Radium 226 & 228; 4006=Uranium (see How-To-Guide for converting gross alpha particle activity to U in ug/L) | Y |
| ConcentrationUnits | The analyte-specific units of summary-level measures as reported in the Concentration field. Each analyte has a standard unit for this dataset. | NA | AN(6) | “ug/L” **allowed only for** (Arsenic, TTHM, HAA5, Atrazine, DEHP, PCE,TCE, uranium); “mg/L” **allowed only for** (Nitrate as nitrogen); “pCi/L” **allowed** **only for** (Radium) | Y |
| Concentration | Reported summary-level concentration. | NA | 6.4 | >0 for summary-level measure. | Y |
| DateSampled | Date last sampled (summary-level data) | NA | YYYY-MM-DD | A valid date from 1/1/1999 through December 31st of the latest complete year (e.g., 2011-12-31). | Y |
| **SchemaName** | **Field Description** | **Schema Group** | **Format** | **Allowed Values** | **Required (Y/N)** |
| AggregationType | Type of summary operation performed (i.e., mean or max) for summary-level data. | Summary  LevelGroup | AN(3) | “X” = Mean (for Annual and Quarterly data); “MX” = Maximum (For Annual Data **ONLY; DO NOT SUBMIT FOR QUARTERLY DATA** | Y |
| NumSamplingLocations | Number of compliance sampling locations from which summary-level records were derived. | Summary  LevelGroup | N(4) | 1-9999; “-888” for Not Submitted | Y |
| SummaryTimePeriod | Year or Quarter for summary-level data | Summary  LevelGroup | AN(10) | YYYY for annual summarized values; YYYY-Q for quarterly summarized values **Allowed Only  for Analyte Codes 2050, 2456, 2950 and 1040** | Y |
| NumSamples | Number of samples used in calculating the mean/max for a given analyte during a quarter or year. | Summary  LevelGroup | N(4) | 1-XXXX | Y |
| NumNonDetects | The number of samples that were non-detections for summary-level data. | Summary  LevelGroup | N(4) | 0-XXXX (XXXX must be no greater than NumSamples) | Y |

Page Break

Icon

Description automatically generated with medium confidenceCDC’S ENVIRONMENTAL PUBLIC HEALTH TRACKING PROGRAM

Contact us: [trackingsupport@cdc.gov](mailto:trackingsupport@cdc.gov)

Visit the Tracking Network today: [www.cdc.gov/ephtracking](http://www.cdc.gov/ephtracking)

Follow us on social media:

* Twitter (@CDC\_EPHTracking)
* Facebook (facebook.com/CDCEPHTracking)

faskfjlkajsdlfasdf