Content Domain: Drinking Water Quality

Recommendations for
Nationally Consistent Data and Measures

Part 2: Recommended Data Sets

Last Revised: July 6, 2012

Environmental Public Health Tracking
Drinking Water Quality
PUBLIC DRINKING WATER QUALITY DATASET:
PWS Inventory, Drinking Water Quality Sampling Results

Characteristic	Description
Data Sources	Safe Drinking Water Act/State Drinking Water Information System (SDWA/SDWIS),or SDWIS-like Data System
Purpose	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, 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. This dataset also accommodates sample-level data and includes one record for each compliance sample for the same analytes used in calculating summary concentrations. Some fields (e.g. NumSamplingStations) are only included for summary-level data observations and belong to the schema group "SummaryLevelGroup". Other fields (e.g. DetectionLimit) are only included for sample-level data observations and belong to the schema group "SampleLevelGroup". Sample-level observations are not required as part of the Spring 2012 CDC data submission; Associated optional fields for sample-level data are indicated as such in the Optionality column.
Restrictions	This is **not** a restricted access data set.

PWS Inventory

Header				
Field Name/SchemaName	**Field Description**	**Optionality**	**Format**	**Allowed Values**
StateFIPSCode	State FIPS code	Required	AN(2)	FIPS State Code
Table Core		**Optionality**		
SchemaName	**Field Description**		**Format**	**Allowed Values**
PWSIDNumber	PWS identifier	Required	AN(9)	nine character value consisting of the 2 letter state abbreviation followed by 7 numbers
YearAssociatedTo	Year that these data are associated with regards to sampling results	Required	Text(4)	YYYY. 1999 through latest complete year (e.g. 2011)
YearPulled	Year that these data were pulled from state records	Required	Text(4)	YYYY. 1999 through latest year.
PWSName	Name of PWS	Required	AN(40)	Any; “U” = Unknown; “NS” = Not submitted
PrincipalCountyServedFIPS	Principal county FIPS served by the CWS	Required	AN(5)	Any; “U” = Unknown; “NS” = Not submitted
PrincipalCityFeatureID	Principal city, town or village Feature ID served by the CWS	Required	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	Required	N(7)	0-9999999;”-999” for Missing; “-888” for Not Submitted
SystemPopulation	Permanent population uniquely served by the CWS	Required	N(8)	0-99999999; ”-999” for Missing; ; “-888” for Not Submitted
PrimarySourceCode	Type of source	Required	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
Note: LatitudeRange or LatitudeNS fields are mutually exclusive in the XML schema. One or the other is included as sub-elements of the <Latitude> parent element.				
LatitudeRange	Latitude in NAD83 decimal degrees describing approximate center of retail service area of water system	Only required if LatitudeNS is not used	N(10)	00.0000000 to 90.000000

LatitudeNS	If you do not have the latitude and plan to submit the value as Missing or Not Submitted, it must be submitted as LatitudeNS	Only required if LatitudeRange is not available	N(4)	”-99.99” for Missing; “-88.88” for Not Submitted.
Note: LongitudeRange or LongitudeNS fields are mutually exclusive in the XML schema. One or the other is included as sub-elements of the <Longitude> parent element.				
LongitudeRange	Longitude in NAD83 decimal degrees describing approximate center of retail service area of water system	Only required if LongitudeNS is not used	N(11)	-180.000000 to 180.000000
LongitudeNS	If you do not have the longitude and plan to submit the value as Missing or Not Submitted, it must be submitted as LongitudeNS.	Only required if LongitudeRange is not available	N(4)	”-999” for Missing; “-888” for Not Submitted.
LocationDerivationCode	Code describing how approximate latitude/longitude location was derived	Required	AN(4)	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.)

Drinking water quality sampling results

Header					
Field Name/SchemaName	**Field Description**	**Optionality**	**Schema Group**	**Format**	**Allowed Values**
StateFIPSCode	State FIPS code	Required	NA	AN(2)	FIPS State Code
Table Core					
SchemaName	**Field Description**	**Optionality**	**Schema Group**	**Format**	**Allowed Values**
PWSIDNumber	PWS identifier	Required	NA	AN(9)	Nine character value consisting of the 2 letter state abbreviation followed by 7 numbers
Year	Year	Required	NA	Text(4)	YYYY; 1999 through latest complete year (e.g. 2011)
AnalyteCode	USEPA Analyte code for required constituents (arsenic, nitrate, TTHM, HAA5, atrazine, PCE, TCE, DEHP, radium, and uranium)	Required	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)
ConcentrationUnits	The analyte-specific units of summary-level measures and individual sample values as reported in the Concentration and DetectionLimit fields. Each analyte has a standard unit for this dataset.	Required	NA	AN(6)	“ug/L” (Arsenic, TTHM, HAA5, Atrazine, DEHP, PCE,TCE, uranium); “mg/L” (Nitrate as nitrogen); “pCi/L” (Radium)
Concentration	Reported summary-level concentration or reported concentration of sample	Required	NA	6.4	>0 for summary-level measure or sample-level concentration; -888, if sample-level data and NonDetectFlag=1; -999= not reported
DateSampled	Date of sample (sample-level data) or Date last sampled (summary-level data)	Required	NA	YYYY-MM-DD	A valid date from 1/1/1999 through December 31st of the latest complete year (e.g. 2011-12-31).

SamplePointID	Sampling station identifier for sample-level records only.	Optional	SampleLevelGroup	AN(20)	Character ID for sampling station;"-999" for missing; "-888" for Not Submitted
DetectionLimit	Sample detection limit	Optional	SampleLevelGroup	3.6	>0, if NonDetectFlag=1; -888, if NonDetectFlag=0.
NonDetectFlag	Flag to indicate whether sample resulted in a detection or not	Optional	SampleLevelGroup	N(1)	1=Sample was a non-detect; 0=Sample was a detection
AggregationType	The type of summary operation performed (i.e. mean or max) for summary-level data.	Required	SummaryLevelGroup	AN(3)	"X" = Mean; "MX" = Maximum
NumSamplingLocations	Number of compliance sampling locations available from which summary-level records were derived.	Required	SummaryLevelGroup	N(4)	1-9999; "-888" for Not Submitted
SummaryTimePeriod	Year or Quarter for summary-level data	Required	SummaryLevelGroup	AN(10)	YYYY for annual summarized values; YYYY-Q for quarterly summarized values
NumSamples	The number of samples that were used in calculating the mean/max for a given analyte during a quarter or year.	Required	SummaryLevelGroup	N(4)	1-XXXX

NumNonDetects	The number of samples that were non-detections for summary-level data.	Required	SummaryLevelGroup	N(4)	0-XXXX (XXXX must be no greater than NumSamples)