The Data Management and Quality Assurance/Quality Control Process for EPA's Fourth Six-Year Review's Microbial and Disinfection Byproduct Preliminary Datasets

Office of Water (4607M)
EPA- 810-R-22-001
August 2022

Disclaimer

This document describes the Microbial and Disinfection Byproducts (MDBP) compliance monitoring data and treatment technique information that was collected for EPA's fourth Six-Year Review (SYR4). The purpose of the Six-Year Review (SYR) is to evaluate current information for regulated contaminants to determine if there is new information to support a regulatory revision that will improve or strengthen public health protection. The SYR4's MDBP data files are being preliminarily released ahead of the publication of SYR4 results for the purpose of MDBP rule revisions analyses. For more information on the Potential Revisions of the MDBP Rules see EPA's webpage https://www.epa.gov/dwsixyearreview/potential-revisions-microbial-and-disinfection-byproducts-rules. The data files released in July 2022 are believed to be fully accurate. Should errors or other data quality issues be identified between July 2022 and the date for the final release of SYR4, EPA may elect to update the MDBP data files – i.e., at any time up until the completion of SYR4.

Executive Summary

The 1996 Amendments to the Safe Drinking Water Act (SDWA) require that the Environmental Protection Agency (EPA) "shall, at least once every six years, review and revise, as appropriate, each National Primary Drinking Water Regulation (NPDWR)." The NPDWRs are often referred to as the national drinking water contaminant regulations or drinking water standards. The purpose of the review, called the Six-Year Review (SYR), is to evaluate current information for regulated contaminants to determine if there is new information on health effects, treatment technologies, analytical methods, occurrence and exposure, implementation and/or other factors that provides a health or technical basis to support a regulatory revision that will improve or strengthen public health protection. To support each of Six-Year Review processes (including fourth Six-Year Review, SYR4, the EPA issues an Information Collection Request (ICR) to the States and primacy agencies to collect the recent data information that public water systems (PWSs) have submitted per requirements of NPDWRs. The data is voluntarily submitted and typically consist of the compliance monitoring records and the records related to treatment technique requirements, usually covering a period of about six years for every cycle. For more information on the SYR4 ICR see EPA's website: https://www.epa.gov/dwsixyearreview/six-year-review-4-drinking-water-standards-information-collection-request)

As a result of EPA's third Six-Year Review (SYR3) of NPDWRs that was published in 2017 (https://www.epa.gov/dwsixyearreview/six-year-review-3-drinking-water-standards), EPA identified eight contaminants covered by the Microbial and Disinfection Byproducts (MDBP) rules as candidates for revision. The eight contaminants include: Chlorite, *Cryptosporidium*, Haloacetic acids, Heterotrophic bacteria, *Giardia lamblia*, *Legionella*, Total Trihalomethanes, and viruses. The eight contaminants are included in the following MDBP rules: Stage 1 and Stage 2 Disinfectants and Disinfection Byproducts Rules, Surface Water Treatment Rules, Interim Enhanced Surface Water Treatment Rule, and Long-Term 1 Enhanced Surface Water Treatment Rule. As a follow-on to SYR3, EPA is conducting analyses to further evaluate the eight NPDWRs for potential regulatory revisions under the potential MDBP Rule Revisions effort (https://www.epa.gov/dwsixyearreview/potential-revisions-microbial-and-disinfection-byproducts-rules). To help support the ongoing considerations of the potential MDBP Rule Revisions and related analyses, EPA is posting the SYR4 ICR data files pertaining to MDBP rules prior to the publication of SYR4 results. The SYR4 ICR data records not pertaining to MDBP rules will be available along with the SYR4 results, expected in 2023.

Since the data recording, managing practices and resultant data records can vary among individual states and primacy agencies, upon receipt of the data files for SYR, EPA conducts a Quality Assurance/Quality Control (QA/QC) Process to normalize the data records for analyses at a national level (including characterization of national occurrence baselines of regulated contaminants). This document describes the QA/QC process for the posted MDBP data files contained in the SYR4 ICR dataset for the potential MDBP Rule Revisions. This document describes the overall QA/QC process that was applied to all SYR4 ICR data as well as the QA/QC process applied specifically to the MDBP data files.

The document also contains a User Guide for downloading and importing the MDBP data from the EPA website (https://www.epa.gov/dwsixyearreview/microbial-and-disinfection-byproduct-data-files-2012-2019-epas-fourth-six-year).

Contents

List of Exhibits

Appendices

Acronyms

CAS	Chemical Abstracts Service
CO	Confirmation
CWS	Community Water System
DBP	Disinfection Byproduct
DBPR	Disinfection Byproduct Rule
D/DBPR	Disinfectants and Disinfection Byproducts Rule
EC	*Escherichia coli (E. coli)*
eDWR	Electronic Drinking Water Report
EPA	Environmental Protection Agency (United States)
FBRR	Filter Backwash Recycling Rule
FC	Fecal Coliforms
GW	Ground Water
GWR	Ground Water Rule
GWUDI	Ground Water Under Direct Influence (of Surface Water)
HAA	Haloacetic Acids
HPC	Heterotrophic Plate Count
IESWTR	Interim Enhanced Surface Water Rule
ICR	Information Collection Request
LT1ESWTR	Long-Term 1 Enhanced Surface Water Treatment Rule
LT2ESWTR	Long-Term 2 Enhanced Surface Water Treatment Rule
MCL	Maximum Contaminant Level
MDBP	Microbial and Disinfection Byproducts
MDL	Method Detection Limit
mg/L	Milligrams per Liter
MOR	Monthly Operating Report
MR	Maximum Residence
MRDL	Maximum Disinfectant Residual Level
MRL	Minimum Reporting Level
MS	Microsoft
NCOD	National Contaminant Occurrence Database
ND	Non-detect or Non-detection
NPDWR	National Primary Drinking Water Regulation
NTNCWS	Non-Transient Non-Community Water System
OMB	Office of Management and Budget
PWS	Public Water System
PWSID	Public Water System Identification Number
QA	Quality Assurance
QC	Quality Control
RP	Repeat
RT	Routine
RTCR	Revised Total Coliform Rule
SDWA	Safe Drinking Water Act
SDWIS/Fed	Safe Drinking Water Information System / Federal Version

SDWIS/State	Safe Drinking Water Information System/State Version
SW	Surface Water
SWP	Purchased Surface Water
SWTR	Surface Water Treatment Rule
SYR	Six-Year Review
SYR3	Third Six-Year Review
SYR4	Fourth Six-Year Review
TC	Total Coliform
TCR	Total Coliform Rule
TG	Triggered
TNCWS	Transient Non-Community Water System
TOC	Total Organic Carbon
TTHM	Total Trihalomethanes
USEPA	United States Environmental Protection Agency
µg/L	Micrograms per Liter

Chapter 1 Introduction

This document describes the Quality Assurance/Quality Control (QA/QC) process applied to the Microbial and Disinfection Byproduct (MDBP) data that was collected as a part of the fourth Six-Year Review (Six-Year Review 4 or SYR4) of National Primary Drinking Water Regulations (NPDWRs). The purpose of the Six-Year Review (SYR) is to evaluate current information for regulated contaminants to determine if there is new information to support a regulatory revision that will improve or strengthen public health protection. This document describes how this data were requested, obtained, received, evaluated and formatted (when necessary). This document also describes data quality issues and modifications to the data to make it consistent throughout and usable for analyses. The SYR4 MDBP data files are being released separately of SYR4 publication for the purpose of MDBP rulemaking revisions analyses.

The SYR4 compliance monitoring data and treatment technique information were provided to EPA voluntarily by primacy agencies via the SYR4 Information Collection Request (ICR) process. EPA received data from 59 primacy agencies (46 states plus territories, Washington, D.C., and Tribes).

The SYR4 ICR data were received from primacy agencies in a variety of formats and data structures and required restructuring to a uniform format for the purpose of conducting contaminant occurrence analyses.

This document describes the MDBP compliance monitoring data and treatment technique information requested and received for SYR4, and provides an overview of the data management, and the QA/QC efforts used to prepare the MDBP datasets.

Chapter 2 Data Acquisition

To obtain national compliance monitoring data and treatment technique information used in support of SYR4, EPA conducted a data call-in from the states, through the National Compliance Monitoring Information Collection Request (ICR) Dataset for the fourth Six-Year Review (or "SYR4 ICR dataset"). For more information on the process undertaken to request the voluntary submission of compliance monitoring data and treatment technique information from primacy agencies, see the fourth Six-Year Review ICR (84 FR 58381, USEPA, 2019).

EPA contacted each primacy agency via a letter requesting the voluntary submission of their compliance monitoring data and treatment technique information for all NPDWRs and related parameters that were collected between January 2012 and December 2019.

EPA requested only information stored electronically (no paper records) and that represented routine compliance monitoring data and treatment technique information. Exhibit 1 shows the regulated contaminants for Stage 1 and Stage 2 Disinfectants and Disinfection Byproducts Rules DBP Rules (D/DBPRs) and Surface Water Treatment Rules (SWTRs) for which EPA requested data, and Exhibit 2 shows the requested data elements (e.g., columns or fields) for each sample result. Note that there were cases where EPA did not receive any data on the data elements and/or analytes requested (these cases were at both the state and system level).

Exhibit 1: List of Microbial and Disinfection Byproducts Contaminants/Parameters Identified in SYR4 ICR for which Data Were Requested from States

Disinfectants and Disinfection Byproducts Rules (D/DBPRs)		
Total Trihalomethanes (TTHMs):	Haloacetic Acids 5 (HAA5):	Bromate
Chloroform	Monochloroacetic acid	Chlorite*
Bromodichloromethane	Dichloroacetic acid	Chlorine*
Dibromochloromethane	Trichloroacetic acid	Chloramines*
Bromoform	Bromoacetic acid	Chlorine dioxide
	Dibromoacetic acid	
Total Coliform Rule (TCR) and Revised Total Coliform Rule (RTCR)		
Total coliforms	Fecal coliforms	*Escherichia coli (E. coli)*
Surface Water Treatment Rules (SWTRs)		
Chlorine**	*Cryptosporidium* ***	Heterotrophic Plate Count (HPC)
Chloramines**	*Giardia lamblia*	
Filter Backwash Recycling Rule (FBRR)		
No specific occurrence data collected.		

*As a maximum disinfectant residual level (MDRL). Chlorine and chloramines are reported as free chlorine and total chlorine, respectively.
** As a minimum disinfectant residual level. Chlorine and chloramines are reported as free chlorine and total chlorine, respectively.
***The monitoring data from Round 2 under Long- Term 2 Enhanced Surface Water Treatment Rule (LT2), is being reviewed and will be available along with the SYR4 results.

Exhibit 2: Data Elements Requested by EPA for the Fourth Six-Year Review[1]

Data Category	Description
System-Specific Information	
Public Water System Identification Number (PWSID)	The code used to identify each PWS. The code begins with the standard 2-character postal state abbreviation or Region code; the remaining 7 numbers are unique to each PWS in the state.
System Name	Name of the PWS.
Federal Public Water System Type Code	A code to identify whether a system is: • Community Water System; • Non-transient Non-community Water System; or • Transient Non-community Water System.
Population Served	Highest average daily number of people served by a PWS, when in operation.
Federal Source Water Type	Type of water at the source. Source water type can be: • Ground water; or • Surface water; or • Ground water under the direct influence of surface water (GWUDI) (**Note**: Some States may not distinguish GWUDI from surface water sources. In those States, a GWUDI source should be reported as a surface water source type.)
Treatment Information	
Water System Facility	System facility data, including: treatment plant identification number, treatment plant information, treatment unit process/objectives, facility flow, treatment train (train or flow of water through treatment units within the treatment plant).
Filtration Type	Information relating to system filtration, including: filtration status, types of filtration (e.g., unfiltered, conventional filtration, and other permitted values).
Treatment Technique Information	Information pertaining to treatment processes. Types of treatment technique information including: disinfectants used and their doses for primary and secondary disinfection, coagulant/coagulant aid type and dose, disinfectant concentration, disinfection profile/benchmark data, log of viral inactivation/removal, contact time, contact value, pH, temperature.
Filter Backwash Information	Information about filter backwash that is returned to the treatment plant influent (e.g., information on: recycle/schematic status, alternative return location, corrective action requirements, and recycle flows and frequency).
Sample-Specific Information	
Sampling Point Identification Code	A sampling point identifier established by the state, unique within each applicable facility, for each applicable sampling location (e.g., entry point to the distribution system). This information enables occurrence assessments that address intra-system variability.
Sample Identification Number	Identifier assigned by state or the laboratory that uniquely identifies a sample.
Sample Collection Date	Date the sample is collected, including month, day, and year.

Data Category	Description
Sample Type	Indicates why the sample is being collected (e.g., compliance, routine, repeat, confirmation, additional routine samples, duplicate, special, special duplicate, etc.).
Sample Analysis Type Code	Code for type of water sample collected. • Raw (Untreated) water sample • Finished (Treated) water sample *For TCR Repeats only; indicator of sampling location relative to sample point where positive sample was originally collected:* • Upstream • Downstream • Original
Contaminant	Contaminant name, 4-digit SDWIS contaminant identification number, or Chemical Abstracts Service (CAS) Registry Number for which the sample is being analyzed.
Sample Analytical Result - Sign	The sign indicates whether the sample analytical result was: • (<) "less than" means the contaminant was not detected or was detected at a level "less than" the minimum reporting level (MRL). • (=) "equal to" means the contaminant was detected at a level "equal to" the value reported in "Sample Analytical Result - Value." • (+) "positive result" *(For RTCR data, only positive E. coli result sign to be included.)*
Sample Analytical Result - Value	Actual numeric (decimal) value of the analysis for the chemical results, or the MRL if the analytical result is less than the contaminant's MRL. *(For the TCR and RTCR, TC and E. coli will indicate presence/absence, and positive E. coli will have numeric results.)*
Sample Analytical Result - Unit of Measure	Unit of measurement for the analytical results reported (usually expressed in either µg/L or mg/L for chemicals; or pCi/l or mrem/yr for radiological contaminants). *(Not required for TCR and RTCR data)*
Sample Analytical Method Number	EPA identification number of the analytical method used to analyze the sample for a given contaminant.
Source Water Monitoring Information	Total organic carbon (TOC), including percent TOC removal, TOC removal summary, pH, alkalinity, monitoring data entered as individual results or included in DBP (or monthly operating report) summary records, alternative compliance criteria, results from round 2 monitoring under LT2 ESWTR (including *Cryptosporidium, E. coli*, turbidity, or state-approved alternate indicators).
Sample Summary Reports	Sample summaries for DBPRs, SWTRs, RTCR, GWR corrective actions, and the Lead and Copper Rule (LCR) associated with analytical result records. Values used for compliance determination [e.g., turbidity (combined effluent/individual effluent), disinfectant residual levels in treatment plant and distribution system, treatment technique information, HPC, etc.]

[1] These are the data elements requested in the SYR4 ICR. Note that the "Data Category" and "Description" Columns were intentionally descriptive rather than prescriptive. This allowed the states that do not use SDWIS/State flexibility to provide as much information as possible. EPA accepted all data "as is" without prescribing structure or format.

About 78 percent of all states currently store and manage at least portions of their compliance monitoring data and/or treatment technique information in the Safe Drinking Water Information System/State Version (SDWIS/State). EPA developed SDWIS/State in collaboration with state primacy agencies to manage drinking water information and provide a common structure for the

development of reusable components and shared applications. The SDWIS/State structure is flexible enough to support the most complex primacy agency program implementation while maintaining a common core of data elements required for reporting to SDWIS/Fed. In an attempt to make the SYR4 data submittal process as easy for states as possible, EPA developed a SDWIS/State Extract Tool (also referred to as "extraction tool" throughout this document), which enabled to run a customized query to pull the requested data from a SDWIS/State database maintained by those states. All of the primacy agencies using SDWIS/State that submitted data to EPA for SYR4 used the extraction tool to extract and compile the EPA-requested compliance monitoring and treatment technique data.

SDWIS/State supports the eDWR (Electronic Drinking Water Report) XML Schema used by laboratories throughout the nation to electronically report sample analytical results as structured data to SDWIS/State. As a result, primacy agencies receive high quality data from laboratories that is batch-processed into SDWIS/State rather than manually entered. Consequently, states have a substantial amount of high-quality structured data available in SDWIS/State. In all, for SYR4, 46 states and 13 other primacy agencies provided compliance monitoring data and treatment technique information that included parametric records. The seven states/primacy agencies that did not provide any SYR4 data were Georgia, Michigan, Mississippi, New Mexico, Guam, Puerto Rico, and U.S. Virgin Islands.

Exhibit 3 lists the states that did submit SYR4 data and indicates whether or not they used the extraction tool. Thirty-five states, Washington D.C, and six regional tribal entities used the extraction tool to extract all or some of their data; therefore, those datasets were all submitted in a similar format. The 17 states/entities not using SDWIS/State submitted their compliance monitoring data and treatment technique information "as is," resulting in a variety of formats, including dBase, MS Excel, XML, MS Access, and comma-delimited. With the exception of two states whose data were downloaded from their publicly available website (California and Florida), all states submitted their data over the Internet via EPA's Central Data Exchange.

Exhibit 3: Summary of States and Other Entities that Provided Compliance Monitoring Data and Treatment Technique Information for SYR4

	State/Entity Name		
States/Tribes that DID use the SDWIS/State Extract Tool	Alabama Alaska Arizona Arkansas Connecticut Delaware Hawaii Idaho Illinois Indiana Iowa Kansas Kentucky Louisiana	Maine Maryland Missouri Montana Nebraska Nevada New Jersey New York North Carolina North Dakota Ohio Oklahoma Oregon Region 4 tribes	Region 5 tribes Region 6 tribes Region 7 tribes Region 8 tribes Region 10 tribes Rhode Island South Carolina Texas Utah Vermont Virginia Washington D.C West Virginia Wyoming
States/Tribes that DID NOT use the SDWIS/State Extract Tool	American Samoa California[1] Colorado Commonwealth of the Northern Mariana Islands Florida[1] Massachusetts	Minnesota Navajo Nation New Hampshire Pennsylvania Region 1 tribes Region 2 tribes	Region 9 tribes South Dakota Tennessee Washington Wisconsin
States/Tribes that DID NOT submit any SYR4 data	Georgia Guam Michigan	Mississippi New Mexico	Puerto Rico U.S. Virgin Islands

[1] CA and FL compliance monitoring and treatment technique information was extracted from a publicly available website

Chapter 3 Data Management

This section provides descriptions of the data management tasks that were used to prepare the SYR4 datasets for QA/QC review. The SDWIS/State Extract Tool pulled the SDWIS/State data into Microsoft Access. Data from states that did not use the SDWIS/State Extraction tool were restructured into a format similar to the SDWIS/State Extraction tool's output. The two groups of datasets (the extract states and the non-extract states (referred to for the remainder of this document as the "SDWIS states" and the "non-SDWIS states," respectively) were managed separately, ultimately getting all datasets into the same format.

A status documentation file was maintained that included information for each state. Specifically, the status documentation described the state datasets received as well as the date received, file type, whether the extraction tool was used and the date range of the data. The status documentation also described any state-specific notes, issues or concerns. Upon receipt of each state dataset, EPA created state-specific directories for each raw dataset. Original datasets were saved and maintained exactly as received and stored in EPA database. Any subsequent changes to a state's dataset were made to a copy of the original dataset and all changes were documented.

3.1 Review of SYR4 Dataset Content

Similar to prior rounds of the Six-Year Review, the first assessment of the submitted SYR4 datasets sought to verify that all of the necessary data elements were included in each state dataset. This review included a comparison of the data elements requested in the state letter, specifically those necessary for the SYR4 analyses, to the entire list of data elements included in each state's dataset. Although data dictionaries were not necessary for the review of data from the SDWIS states, these files (and any other available supporting information provided by the states) were useful interpreting the data submitted by the non-SDWIS states. Supporting information included descriptions of the sampling efforts provided in emails from the state, additional information on acronym definitions, and more.

Data dictionaries and supporting information were reviewed for definitions of the various data elements, row and column headings, codes, and acronyms. If fields were missing or not recognizable, EPA included a question to the state in their "flagged record report" email. "Flagged record reports" were detailed reports sent via email to each state that identified records of potential data quality concern. In addition, questions on data completeness, statewide waivers, and any other unique factors within the state's dataset were included. In addition, many of the non-SDWIS states submitted datasets with more data elements than necessary. In those cases, EPA determined which data elements were and were not specific to the SYR4 data request.

EPA also confirmed that all of the requested contaminants from the SYR4 ICR were included in each state dataset. As a first step for the non-SDWIS states, EPA reviewed the CHEMIDs (i.e., four-digit SDWIS codes) and/or contaminant names within each state's dataset. Many states included only CHEMIDs or contaminant names. A few other states only included CAS numbers or state-specific codes. EPA populated missing information using a variety of sources including a list of SDWIS codes from the SDWIS/Fed database as well as the ChemIDPlus website (if only

CAS numbers were included). Nine of the non-SDWIS states submitted at least some data for a contaminant or contaminants for which a four-digit SDWIS code could not be determined. Other times, the state appeared to be using an incorrect four-digit SDWIS code for a particular contaminant. EPA compiled a list of questions for states related to issues such as missing contaminants or undetermined CHEMIDs to be included in the "flagged record reports." States were asked questions such as if there was a statewide waiver for missing contaminants, if certain contaminant data were stored in a separate database, or if there had been a typographical error with a particular CHEMID.

Sample collection dates were reviewed to ensure that there were not any inconsistent dates reported (e.g., data from the year 1900). If there were suspicious/incorrect sample collection dates included, EPA tried to use other data elements to provide insight on the correct date (e.g., "analyzed date"). If the correct date could not be determined, EPA included a question for the state in its "flagged record report" and either states followed up with EPA or EPA followed up with states.

3.2 Restructuring Non-SDWIS State Data

Datasets received from the non-SDWIS states were restructured through a series of Microsoft (MS) Access queries into a format similar to the data structure of the data from the SDWIS states to allow for the construction of a unified database for the SYR4 national contaminant occurrence analyses. As a first step in this process, EPA identified the data structure of each non-SDWIS state dataset to plan the best method for conversion to the final database structure.

Prior to populating the SYR4 ICR database, EPA standardized the data reported by each non-SDWIS state to reflect the appropriate SDWIS codes. For example, in the source water type field (i.e., "D_FED_PRIM_SRC_CD"), all instances of "surface water" or "S" were changed to "SW." In the system type field (i.e., "D_PWS_FED_TYPE_CD"), all instances of "CWS" or "community" were changed to "C" for community water systems. All PWSIDs had to be put in the federal format of the two-character postal state abbreviation or Region code followed by a seven-digit number, unique to each PWS in the state.

After the various state-specific formatting and transformations were completed, EPA imported all non-SDWIS datasets into Access to ultimately merge with the SDWIS/State data sets in Oracle, a database storing all SYR4 data. In some cases, EPA imported only the data elements identified as essential to the occurrence analysis. Upon completion, EPA compared all transformed state datasets to the original datasets to ensure all data were accurately converted. Furthermore, EPA saved a record of the procedures used to map the state datasets to the SYR4 ICR database. All queries were created and saved in Access to document the transformation, ensuring that this process is reproducible.

3.3 Establishing Consistent Data Fields for Analytical Results (SDWIS and Non-SDWIS States)

EPA structured the sample analytical result sign, sample analytical result value, and sample analytical result unit of measure into a consistent format to prepare the data for occurrence

analysis. EPA conducted this step prior to reviewing the data for potential outliers. Many of the state datasets included analytical results signs (e.g., "<" for non- detections or "=" for detections), detection limits and analytical results data in multiple fields. EPA added a "DETECT" field to the SYR4 ICR dataset to identify the results sign and to more easily conduct analyses. Wherever the analytical result was greater than zero and the result sign indicated a detection, then DETECT was set equal to 1, representing a detection. When the analytical result was equal to zero and/or the result sign indicated a non-detection, then DETECT was set equal to 0 (i.e., a non-detect).

EPA received data with various units of measure. It was important that all data for each individual contaminant be expressed in a single unit to facilitate analysis. For this analysis, EPA converted all data for trihalomethanes (THMs) and haloacetic acids (HAAs) to µg/L. All records with missing or unusual units in the SYR4 ICR dataset were sent back to states for input as part of flagged reports mentioned earlier.

Chapter 4 Data Quality Assurance and Quality Control

After EPA converted the state datasets into a consistent format, a significant effort was undertaken to ensure the quality of the data submitted. Data quality, completeness, and representativeness were key considerations for the dataset. Given the size, scope, and variety of formats of the datasets received from the states, EPA conducted extensive data management and QA/QC evaluation on the data to be included in the SYR4 ICR dataset. This QA/QC evaluation involved the assessment of data ranging in quality across the different contaminants and different states. This chapter includes a summary description of the QA/QC measures that were conducted on the state datasets for all SYR4 data which includes the MDBP data.

4.1 Quality Assurance Measures Applied to All Contaminants

Before analyzing contaminant occurrence, EPA performed a rigorous QA/QC evaluation of the data from each state (for both SDWIS and non-SDWIS state users). When necessary, EPA sent emails to states, asking specific questions about its dataset. Question topics included descriptions of non-intuitive data element names, definitions of field headings, or non-standard codes that were not described in any documentation files from the state. EPA also confirmed that all of the requested contaminants were included in each state dataset. When a state was missing data for any of the contaminants, EPA asked the state to identify the reason for the omission, such as a state-wide waiver of the requirement to monitor for the contaminant(s). Information provided by states was documented and kept as a record.

Exhibit 4 lists the system types that are required to sample for the MDBP contaminants. All data that passed the QA/QC process from these systems were included in the SYR4 datasets. Data from systems that were not required to sample for a given contaminant were excluded from the SYR4 datasets.

Exhibit 4: Contaminant Group Monitoring Requirements

Contaminant Group	System Types Required to Sample (sample data included in analyses)	System Types Not Required to Sample (sample data excluded from analyses)
Disinfection Byproducts and disinfectant residuals	Stage 1 and Stage 2 DBP Rules: All community water systems and non-transient noncommunity water systems that add a disinfectant other than ultraviolet (UV) light or deliver disinfected water, and transient non-community water systems that add chlorine dioxide.	Community water systems and non-transient noncommunity water systems that do not add a disinfectant other than UV light, as well as transient non-community water systems that add a disinfectant other than chlorine dioxide.
Microbial Contaminants and disinfectant residuals	Groundwater Rule (GWR): The GWR applies to all public water systems that use ground water, including consecutive systems, except that it does not apply to PWSs that combine all of their ground water with surface water or with ground water under the direct influence of surface water prior to treatment. Surface Water Treatment Rules (SWTRs): The SWTRs apply to all public water systems that use surface water or ground water under direct influence of surface water. Revised Total Coliform Rule (RTCR): The RTCR applies to all public water systems.	None.

EPA created several automated data QA checks within the SYR4 ICR dataset. These QA checks identified (or "flagged") records of potential data quality concerns. EPA sent out a detailed report to each state describing their flagged records called a "flagged records report." These reports included the counts of flagged records by category, as well as specific questions related to each of these categories. In addition, an attachment identified the specific records that were flagged. EPA requested that each state provide the appropriate disposition (delete, make corrections, etc.) of these flagged records. EPA documented all changes made to the compliance monitoring data and suggested to the states that they make corrections in their data system as well, if appropriate. To resolve data quality issues that required significant corrections to the raw data, such as identifying outliers or identifying and changing incorrect units, consultations with state data management staff were conducted or attempted before data corrections were completed.

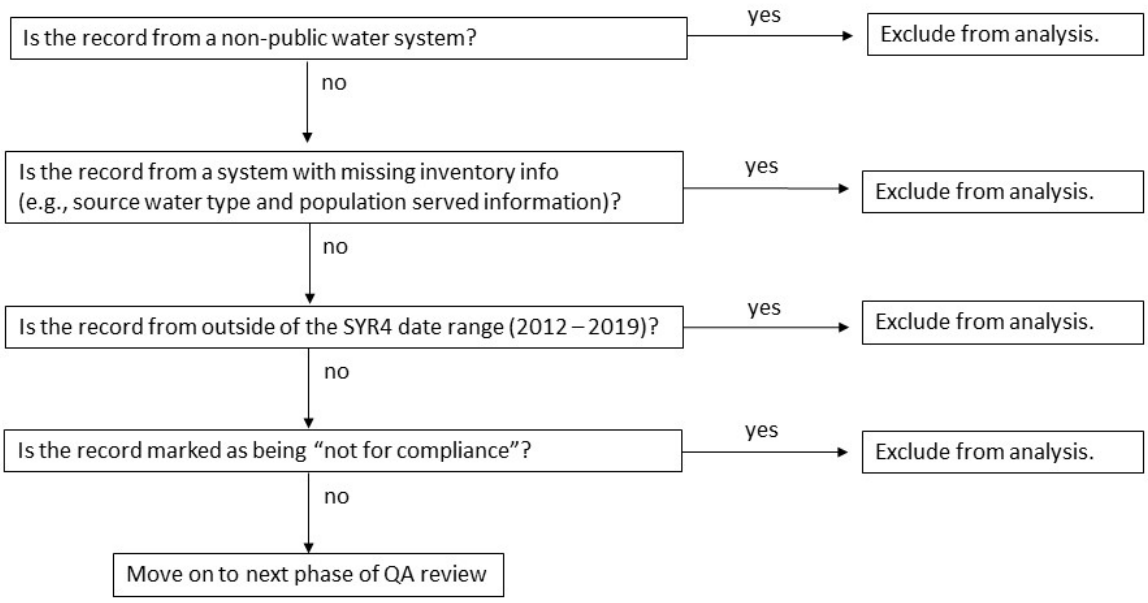
The following sections provide a description of the various QA measures applied to the entire SYR4 dataset that were used to identify records of potential data quality concern. For all flagged records, input from states was always considered as the initial criteria in deciding on the appropriate action or decision to include or exclude the record from analysis. When states did not provide a response or action, EPA used best professional judgement on whether to include or exclude the data in question. When a determination was made to exclude records from the occurrence analyses, a code was added to the "transaction table" in the database to indicate that

the record should not be included in the analyses. This code could be changed if EPA were to revise their decision about excluding/including particular records for occurrence analyses.

Section 4.1.1 through 4.1.5 describe the QA measures that were applied to the entire database (i.e., were relevant to all regulated contaminant monitoring data in the SYR4 ICR dataset).

Exhibit 5 provides a visual for the overall flow of the QA/QC process for QA measures applied to all SYR4 contaminants. Additional QA/QC measures applied to specified groups of contaminants are included in Chapter 5 (DBPs and DBP related parameters) and Chapter 6 (microbial contaminants).

Exhibit 5: Flow Chart of QA Measures Applied to All SYR4 Contaminants



4.1.1 Non-Public Water Systems

Some primacy agencies require water systems that do not meet the criteria to be classified as public water systems to submit sample results that are "routine" or "for compliance." The primacy agency's information system usually identifies these water systems as "non-public" or uses another method to differentiate them from public water systems. All records from non-public water systems were excluded. The records that were included were from systems that classify as PWSs by definition, or systems that identify as a PWS, e.g., wholesale systems.

4.1.2 Systems with Missing Inventory Data

For some of the non-SDWIS states, there were systems for which the inventory information was missing (e.g., no source water type or no population served). When inventory data were incomplete or missing, the missing data were populated with data from the SDWIS/Fed data from the fourth quarter of December 2019. All cases where SDWIS/Fed data were used to

populate inventory data fields in the state's dataset were documented. Note that inventory information may differ for a given system over time so the SDWIS (2019) fourth quarter data may not fully match the actual inventory information at the time of sampling. All records from systems whose inventory data were still missing after filling gaps with SDWIS/Fed were excluded from the datasets.

4.1.3 Sample Results Collected Outside of the Date Range

The SYR4 ICR requested compliance monitoring data and treatment technique information from January 1, 2012 through December 31, 2019. The extraction tool only pulled sample results from this time period. However, some non-SDWIS states submitted sample results from outside of this date range; all sample results collected outside of the date range were excluded from the datasets.

4.1.4 Non-Compliance

In some cases, water systems may submit sample results that are not used to determine compliance with NPDWRs. States that use information systems with automated compliance determination functions often use indicators to differentiate these sample results such as the "compliance purpose indicator code" or something similar. While the extraction tool only pulled compliance sample results, some non-compliance sample results were present in data from the non-SDWIS states. There were a few non-SDWIS states for which EPA asked for more details on how to accurately identify the sample results that were "for compliance." Three non-SDWIS states (California, Colorado and Minnesota) did not make a designation as to whether their data were for compliance. For all occurrence datasets, EPA assumed that all data from these three states were for compliance and included in the datasets. All sample results flagged as "not for compliance" were excluded from the dataset.

4.1.5 Uniform System Inventory Information

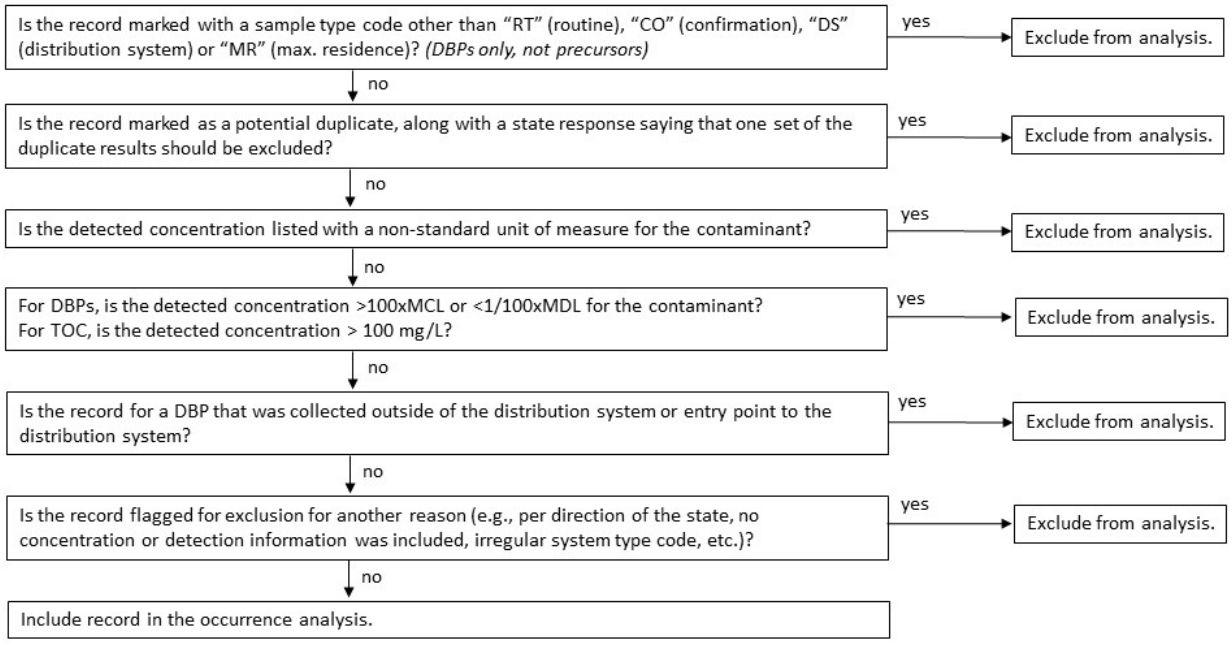
For analyses, each system must have a single source water type and population-served designation to define each system in a unique source water type/population size strata. Systems using both ground water and surface water, and systems using ground water under direct influence of surface water, were considered surface water systems to include in datasets (note, the number of systems that use different sources, disconnected from one another, are unknown. This methodology used to designate source may underestimate the number of ground water systems and overestimate the number of surface water systems). Systems with more than one specified value of population served were assigned the population served value that occurred most frequently within those years of data collected.

Chapter 5 Quality Assurance Measures Applied to Disinfection Byproducts and Disinfection Byproduct Related Parameters

In addition to the QA measures described in Chapter 4 that were applied to all contaminants, there were several additional contaminant-specific QA measures applied to particular contaminant data. In this way QA measures applied to DBP data will differ from those QA measures applied to microbial contaminant data. The QA measures applied to DBPs and DBP related parameters are described in this chapter.

Exhibit 6 presents a flow chart of these additional QA measures for DBPs and DBP related-parameters.

Exhibit 6: Flow Chart of Additional QA Measures Specific to DBPs and DBP Related Parameters



After applying the various QA measures to nearly 12 million SYR4 ICR records for the DBPs and DBP related parameters, 96 percent of the records from 58 states and primacy agencies remained in the final dataset. Exhibits 7 documents the specific counts of DBP records included and excluded in each QA step. Exhibit 7 includes records for the following DBP contaminants: TTHM, bromoform, chloroform, dibromochloromethane, bromodichloromethane, HAA5, dibromoacetic acid, dichloroacetic acid, bromoacetic acid, monochloroacetic acid, trichloroacetic acid, bromate, chlorite and DBP Related Parameters: pH, alkalinity, and total organic carbon (TOC).

Exhibit 7: Summary of the Count of Analytical Sample Results Removed via the QA Measures Applied to DBP Rule Contaminants[1]

QA Step	Count of Records	
	Included	Excluded
Original number of analytical sample results	11,755,299	
Step 1: Removal of analytical sample results from non-public water systems.	11,754,859	440
Step 2: Removal of data from systems with missing source water type and/or population served information.	11,748,860	5,999
Step 3: Removal of data with a sample collection date outside of the Six-Year 4 date range of 2012 - 2019.	11,717,184	31,676
Step 4: Removal of data marked as being "not for compliance."	11,700,871	16,313
Step 5: Removal of DBP data with sample type code other than "RT" (routine), "CO" (confirmation), "DS" (distribution system), or "MR" (max. residence).	11,671,157	29,714
Step 6: Removal of records marked as potential duplicates, along with a state response saying that one set of the duplicate results should be excluded.	11,652,715	18,442
Step 7: Removal of DBP data with detected concentrations with non-standard/blank unit of measure for the contaminant.	11,651,996	719
Step 8: Removal of detected concentrations greater than 100*MCL or less than 1/100*MDL for the contaminant. For TOC, removal of detections >100xMCL.	11,651,791	205
Step 9: Removal of DBP records sampled outside of the distribution system or entry point to the distribution system.	11,229,596	422,195
Step 10: Removal of records with no data/results	11,229,589	7
Step 11: Removal of records with irregular system type codes (specific to State of PA where unknown system type codes were included)	11,228,599	990
Final number of records	11,228,599	
Percent Included	96%	

[1] This table includes records for the following contaminants: TTHM, bromoform, chloroform, dibromochloromethane, bromodichloromethane, HAA5, dibromoacetic acid, dichloroacetic acid, bromoacetic acid, monochloroacetic acid, trichloroacetic acid, bromate, chlorite, pH, alkalinity, and total organic carbon.

5.1 Non-Routine Samples

Some primacy agencies have regulations that are more stringent than the NPDWRs and require water systems to submit more sample results than federally required. Primacy agencies also may require laboratories to report all sample results from water systems including results from contaminants that are not regulated. Usually, non-routine sample results that are specifically listed as “special request” in the database are also identified as being “non-compliance” samples. Most other types of non-routine sample results, such as confirmation, repeat or maximum residence time sample results are considered as “for compliance.” While the extraction tool excluded sample results that were “not for compliance,” some “special” sample results that were marked as being “for compliance” were included in the data extracted from SDWIS states. In addition, “non-routine / not for compliance” results were present in data from the non-SDWIS

states. All DBP results that were marked as routine ("RT"), confirmation ("CO"), or maximum residence ("MR") were included in the DBP dataset.

5.2 Duplicate Records

In the analysis of DBPs and DBP related parameters data, potential duplicates were identified as all detection records with the same PWSID, Sample Point ID, analyte, sample collection date, and concentration. All records identified as potential duplicates were retained in the occurrence dataset unless the state responded to indicate that records were indeed duplicates and should be excluded from the dataset.

5.3 Units of Measure

EPA identified all detection records for the DBPs, TOC, and alkalinity where the units of measure reported were not one of the standard units used for the particular contaminant (i.e., not equal to "mg/L" or "µg/L"). For example, a chloroform record with a unit of measure listed as "NTU" would be flagged. All records in non-standard units were excluded from the occurrence dataset unless there was strong evidence of the correct standard unit to use (e.g., state response indicating the correct unit of measure, obvious data entry error, concentration is within the range of standard units and all other records from the state are reported in the standard units).

5.4 Potential Outliers

To identify potential high outliers, EPA flagged all detected concentrations for the DBP rule contaminants that were greater than four times the contaminant's MCL and all detected concentrations that were greater than ten times the contaminant's MCL. Any concentration identified in the greater than 10 times the MCL would be captured in the greater than 4 times MCL and then followed up with the state about them. To identify potential low outliers, EPA flagged all detected concentrations that were less than one-tenth the minimum MDL. Exhibit 8 provides a list of all relevant MCL values. Note that for total organic carbon (TOC) (not listed in Exhibit 8) all results greater than 100 mg/L were excluded from TOC data file.

EPA included questions to the state on each of these potential high and low outliers in their "flagged record report." Any changes suggested by the states were implemented for these records. For example, some states wrote back to say there were "no errors" in their high detect concentrations or that they had "no reason or evidence to show these data to be invalid." Other states stated that "all of the high results were due to using mg/L when they should have been µg/L." For the states that did not respond, all detected DBP concentrations greater than 100 times the contaminant's MCL were excluded from the dataset. No low-end cut-off was applied for the DBP data. All other potential outliers less than or equal to 100 times the contaminant's MCL were included in the datasets. The value of 100 times the MCL was chosen as a conservative high-end cut-off. For example, a TTHM detected concentration of 10,000 ug/L was excluded as it was assumed a data entry error.

Exhibit 8: List of DBP MCL Values

Contaminant	Maximum Contaminant Level (MCL)	
	Value	Unit of Measure
Chloroform	80[1]	µg/L
Bromoform	80[1]	µg/L
Bromodichloromethane	80[1]	µg/L
Dibromochloromethane	80[1]	µg/L
Total Trihalomethanes (TTHM) [1]	80	µg/L
Monochloroacetic Acid	60[2]	µg/L
Dichloroacetic Acid	60[2]	µg/L
Trichloroacetic Acid	60[2]	µg/L
Bromoacetic Acid	60[2]	µg/L
Dibromoacetic Acid	60[2]	µg/L
Haloacetic acids 5 (HAA5)	60	µg/L
Bromate	10	µg/L
Chlorite	1,000	µg/L

[1] The MCL for total trihalomethanes is 80 µg/L but the individual trihalomethane results were also compared against that MCL to identify potential outliers.
[2] The MCL for the sum of five haloacetic acids is 60 µg/L but the individual haloacetic acid results were also compared against that MCL to identify potential outliers.

5.5 Locational Flag

While the occurrence of DBPs could theoretically occur anywhere in a given water system, EPA is primarily focused on the occurrence in the distribution system. As such, EPA excluded any DBP records with a location sampling point type that was not obviously a part of the distribution system or entry point to the distribution system, such as sampling results from raw or source waters. Specifically, the following location sampling point types were not flagged for exclusion: "DS" (distribution system), "EP" (entry point), "FC" (first customer), "FN" (finished), "LD" (lowest disinfectant residual), "MD" (midpoint of distribution system), or "MR" (maximum residence time). For records whose sampling point location type was either null or labeled as a generic "Water System Facility Point," an additional filter was added to make sure any records with a water system facility type that was likely associated with the distribution system were not excluded. Specifically, the following facility type codes were not flagged for exclusion when the sampling point type code was listed as "WS" (water system facility point) or null: "CC" (consecutive connection), "DS" (distribution system), "TM" (transmission main), or "TP" (treatment plant).

Chapter 6 Quality Assurance Measures Applied to Microbial Contaminants

In addition to the QA measures described in Chapter 4, there were a handful of additional QA measures applied to only the microbial contaminants. Those QA measures are described in this chapter. Exhibit 9 is a flow chart of the additional QA measures applied to the microbial contaminants.

Exhibit 9: Summary of the Count of Records Removed via the QA Measures Applied to Microbial Rule Contaminants

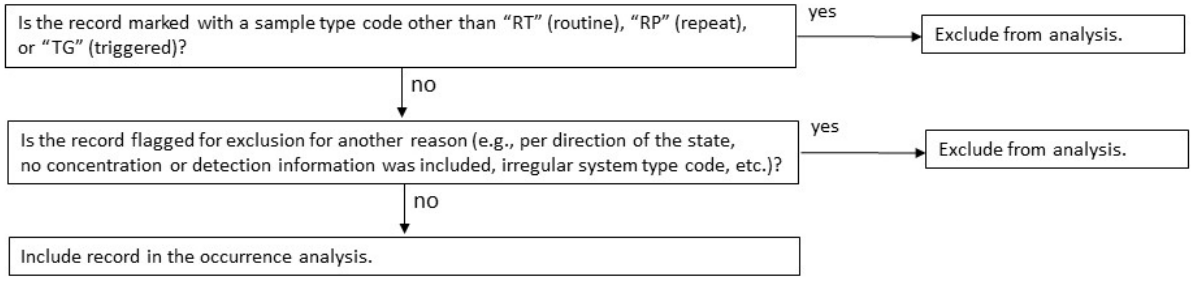


Exhibit 10 documents the specific counts of microbial records included and excluded in each QA step. After applying the various QA measures to more than 28 million SYR4 ICR microbial records, 99 percent of the records from 57 states and primacy agencies remained in the final dataset for use of analyses.

Exhibit 10: Summary of the Count of Analytical Samples Results Removed via the QA Measures Applied to Microbial Rule Contaminants[1]

QA Step	Count of Records	
	Included	Excluded
Original number of analytical samples results	28,329,039	
Step1: Removal of analytical sample results from non-public water systems.	28,315,533	13,506
Step 2: Removal of data from systems with missing source water type and/or population served information.	28,236,298	79,235
Step 3: Removal of data with a sample collection date outside of the Six-Year 4 date range of 2012 - 2019.	28,114,841	121,457
Step 4: Removal of data marked as being "not for compliance."	27,985,027	129,814
Step 5: Removal of microbial data with sample type code other than "RT" (routine), "RP" (repeat), or "TG" (triggered).	27,981,035	3,992
Step 6: Removal of records with no data/results	27,964,042	16,993
Step 7: Removal of records with irregular system type codes (specific to State of PA where unknown system type codes were included)	27,962,474	1,568

QA Step	Count of Records	
	Included	Excluded
Final number of records	27,962,474	
Percent Included	99%	

[1] The following analytes are included in the counts above: Total coliform, Fecal coliform, *E. coli, Cryptosporidium, Giardia lamblia, Enterococci,* and coliphage.

6.1 Non-Routine Samples

Some primacy agencies have regulations that are more stringent than the NPDWRs and require water systems to submit more sample results than federally required. Primacy agencies also may require laboratories to report all sample results from water systems including results from contaminants that are not regulated. Usually, non-routine sample results that are specifically listed as "special request" in the database are also identified as being "non-compliance" samples. Most other types of non-routine sample results, such as confirmation, repeat or maximum residence time sample results are "for compliance." While the extraction tool excluded sample results that were "not for compliance," some "special" sample results that were marked as being "for compliance" were included in the data extracted from SDWIS states. In addition, "non-routine / not for compliance" results were present in data from the non-SDWIS states. These data were flagged and inquired to the states. All results that were marked as routine ("RT"), repeat ("RP"), or triggered ("TG") were included in the occurrence datasets for the microbial contaminants.

6.2 Pairing Disinfectant Residual and Coliform Results for non-SDWIS states

Per requirements under the SWTR, surface water systems need to monitoring disinfectant residuals at the same locations and time as for routine TC under TCR/RTCR. Thus, the TC/EC datasets generally also contain paired disinfectant residual monitoring records. However, two non-SDWIS states, Wisconsin and Pennsylvania, submitted disinfectant residual concentration data as independent records not paired with total coliform (TC) samples. To enable evaluation of disinfectant residual concentrations versus TC positivity rates, EPA paired the residual chlorine data with the associated TC result. EPA paired the two sets of results based on the sample collection date, sample point ID, and lab assigned ID. Using a combination of two approaches, roughly 31 percent of Wisconsin and Pennsylvania's TC records were paired with free chlorine residuals, while around 5 percent were paired with total chlorine residuals. This method enabled more than 410,000 TC records to be paired with free chlorine residuals. In addition, more than 54,000 TC records were paired with total chlorine residuals. In an effort to pair more results, EPA applied a secondary approach to the remaining unpaired records which omitted the lab assigned ID as a necessary "join" field. This pairing effort enabled an additional 97,000 TC records to be paired with free chlorine residuals. Additionally, nearly 33,000 TC were paired with total chlorine residuals.

6.3 Updates to Absence and Presence Codes

Under the SYR4 ICR, some microbial records (total coliform, *E. coli*, and fecal coliform) were submitted without a presence indicator code (i.e., indicating whether the result was absent ("A") or present ("P")) but with a value in the measured concentration field (specifically, the CONCENTRATION_MSR field). EPA updated nearly 4 million microbial records with a null presence absence code and a concentration of zero to set the presence absence code equal to "A". In addition, EPA updated nearly 60,000 microbial records with a PRESENCE_IND_CODE of null to "P" when the concentration was greater than zero, indicating the presence of the microbe.

Chapter 7 References

United States Environmental Protection Agency (USEPA). 2016. Six-Year Review 3 Technical Support Document for Disinfectants/Disinfection Byproducts Rules.

USEPA. 2019. Information Collection Request Submitted to OMB for Review and Approval; Comment Request; Contaminant Occurrence Data in Support of the EPA's Fourth Six-Year Review of National Primary Drinking Water Regulations: October 31, 2019, Volume 84, Number 211, Page 58381-58382.

The Data Management and Quality Assurance/Quality Control Process for EPA's Fourth Six-Year Review's Microbial and Disinfection Byproduct Datasets: Appendices

Appendix A: Data request letter EPA sent June 3, 2020 contacting each Primacy Agency to request voluntary submission of its compliance monitoring data and treatment technique information for regulated chemical, radiological, and microbiological contaminants

UNITED STATES
ENVIRONMENTAL
PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF WATER

State Drinking Water Administrators
Association of State Drinking Water Administrators
1401 Wilson Blvd # 1225
Arlington, VA 22209

Dear State Drinking Water Administrator,

The 1996 Safe Drinking Water Act Amendments require the U.S. Environmental Protection Agency (EPA) to review and revise, if appropriate, existing National Primary Drinking Water Regulations (NPDWRs) at least every six years (i.e., the Six-Year Review). The Agency is currently preparing for the fourth round of the Six-Year Review (Six-Year Review 4).

As was done for the third Six-Year Review, the EPA is contacting each primacy agency (hereinafter referred to as "state") and requesting voluntary submission of its compliance monitoring data and treatment technique information for regulated chemical, radiological, and microbiological contaminants. We are requesting compliance monitoring data collected between January 2012 and December 2019. The Office of Management and Budget (OMB) has approved the information collection request for the EPA's fourth Six-Year Review under the provisions of the Paperwork Reduction Act, 44 U.S.C. 3501 et seq., and has assigned OMB control number 2040-0298.

These data are an important component in supporting the EPA's Six-Year Review of NPDWRs. We are encouraging each state to submit its contaminant monitoring and treatment technique information because these data will contribute directly to the EPA's understanding of national contaminant occurrence, treatment technique information, the population exposed to regulated contaminants, and exposure reductions associated with the current regulations. The EPA is requesting your voluntary submission by September 30, 2020.

The EPA is requesting only data that are currently stored electronically (no paper records), including both detection and non-detection results for compliance monitoring and treatment technique information. Exhibit 1 of the attachment provides a list of the regulated contaminants for which the EPA is requesting data. Exhibit 2 presents critical data elements needed for each sample result. To make your voluntary reporting as easy as possible, your state can transmit its compliance monitoring data set to the EPA using the same process your state currently uses to submit your SDWIS data quarterly. The attachment also answers questions about how the data will be transferred, managed, and used and provides some background information about why we are requesting these data.

In our previous Six-Year Review data collections, we have worked closely with state data managers to answer questions and facilitate data transfer. Soon after June 30, 2020 we will begin contacting data managers and coordinating directly with them by phone and/or email.

Thank you for your consideration of this request. Many of you voluntarily submitted your data for the Six-Year Review 3. We appreciated your participation and hope you will do so again. If you have any questions about this request or the intended uses of the data, please contact Lili Wang, Associate Chief, Standards and Risk Reduction Branch, at wang.lili@epa.gov or Nicole Tucker, Six-Year Review 4 Team Lead, at tucker.nicole@epa.gov.

Sincerely,

Jennifer L. McLain, Director
Office of Ground Water and Drinking Water

Enclosure: Attachment
cc: Regional Water Division Directors
Regional Drinking Water Branch Chiefs
Tribal Direct Implementation Contacts

ATTACHMENT

I. Details Regarding EPA's Request for Contaminant Monitoring Data

A. What regulated contaminants are included in this request?

EPA is requesting compliance monitoring information for chemical, radiological, and microbiological contaminants, as was requested under past Six-Year Reviews. Exhibit 1, below, lists the specific contaminants for which EPA is requesting monitoring data. EPA will work with you to make the data transfer as easy as possible. Voluntary submission of your regulated drinking water contaminant monitoring and treatment technique data is the most critical step in this national occurrence assessment for the Six-Year Review 4**.**

B. What specific data are being requested and what timeframe should the data cover?

EPA is requesting the voluntary submission of compliance monitoring data for regulated chemical, radiological, and microbiological contaminants (Exhibit 1) collected between January 2012 and December 2019. This request only includes those data that you have stored in ***electronic format***. The requested data include routine compliance monitoring samples (including repeat and confirmation samples) and treatment technique data. Please include all results for ***both analytical detections and non-detections.***

Exhibit 2 lists the data elements that are likely to be captured as part of your facility and treatment data, and likely to be in your compliance monitoring database. We encourage you to send us your data even if you feel that your data set is incomplete.

Exhibit 1: Occurrence Data Requested		
Chemical Contaminants (Phase I, II, IIB, and V Rules; Arsenic Rule; Lead and Copper Rule)		
Acrylamide	1,1-Dichloroethylene	Methoxychlor
Alachlor	cis-1,2-Dichloroethylene	Monochlorobenzene (Chlorobenzene)
Antimony	trans-1,2-Dichloroethylene	Nitrate (as N)
Arsenic	Dichloromethane (Methylene chloride)	Nitrite (as N)
Asbestos	1,2-Dichloropropane	Oxamyl (Vydate)
Atrazine	Di(2-ethylhexyl) adipate (DEHA)	Pentachlorophenol
Barium	Di(2-ethylhexyl) phthalate (DEHP)	Picloram
Benzene	Dinoseb	Polychlorinated biphenyls (PCBs)
Benzo[a]pyrene	Diquat	Selenium
Beryllium	Endothall	Simazine
Cadmium	Endrin	Styrene
Carbofuran	Epichlorohydrin	2,3,7,8-TCDD (Dioxin)
Carbon tetrachloride	Ethylbenzene	Tetrachloroethylene
Chlordane	Ethylene dibromide (EDB)	Thallium
Chromium (total)	Fluoride	Toluene
Copper	Glyphosate	Toxaphene

Exhibit 1: Occurrence Data Requested		
Cyanide	Heptachlor	2,4,5-TP (Silvex)
2,4-D	Heptachlor epoxide	1,2,4-Trichlorobenzene
Dalapon	Hexachlorobenzene	1,1,1-Trichloroethane
1,2-Dibromo-3-chloropropane (DBCP)	Hexachlorocyclopentadiene	1,1,2-Trichloroethane
1,2-Dichlorobenzene (*o*-Dichlorobenzene)	Lead	Trichloroethylene
1,4-Dichlorobenzene (*p*-Dichlorobenzene)	Lindane	Vinyl chloride
1,2-Dichloroethane (Ethylene dichloride)	Mercury (inorganic)	Xylenes (total)
Radiological Contaminants		
Combined Radium-226/228; and Radium-226 & Radium-228 *(if available)*	Gross beta	Tritium
	Iodine-131	Uranium
Gross alpha	Strontium-90	
Total Coliform Rule (TCR) and Revised Total Coliform Rule (RTCR)		
Total coliforms	Fecal coliforms	*Escherichia coli (E. coli)*
Disinfectants and Disinfection Byproducts Rules (DBPRs)		
Total Trihalomethanes (TTHMs): Chloroform Bromodichloromethane Dibromochloromethane Bromoform	Haloacetic Acids (HAA5): Monochloroacetic acid Dichloroacetic acid Trichloroacetic acid Bromoacetic acid Dibromoacetic acid	Bromate
		Chlorite
		Chlorine
		Chloramines
		Chlorine dioxide
Ground Water Rule (GWR)		
Escherichia coli (E. coli)	Enterococci	Coliphage
Surface Water Treatment Rules (SWTRs)		
Chlorine	*Cryptosporidium*	Heterotrophic Plate Count (HPC)
Chloramines	*Giardia lamblia*	
Filter Backwash Recycling Rule (FBRR)		
No specific occurrence data collected.		

Exhibit 2: Requested Data Categories	
Data Category	**Description**
System-Specific Information	
Public Water System Identification Number (PWSID)	The code used to identify each PWS. The code begins with the standard 2-character postal state abbreviation or Region code; the remaining 7 numbers are unique to each PWS in the state.
System Name	Name of the PWS.
Federal Public Water System Type Code	A code to identify whether a system is: • Community Water System; • Non-transient Non-community Water System; or Transient Non-community Water System.
Population Served	Highest average daily number of people served by a PWS, when in operation.

Exhibit 2: Requested Data Categories	
Federal Source Water Type	Type of water at the source. Source water type can be: • Ground water; or • Surface water; or • Ground water under the direct influence of surface water (GWUDI) (**Note**: Some States may not distinguish GWUDI from surface water sources. In those States, a GWUDI source should be reported as a surface water source type.)
Treatment Information	
Water System Facility	System facility data, including: treatment plant identification number, treatment plant information, treatment unit process/objectives, facility flow, treatment train (train or flow of water through treatment units within the treatment plant).
Filtration Type	Information relating to system filtration, including: filtration status, types of filtration (e.g., unfiltered, conventional filtration, and other permitted values).
Treatment Technique Information	Information pertaining to treatment processes. Types of treatment technique information including: disinfectants used and their doses for primary and secondary disinfection, coagulant/coagulant aid type and dose, disinfectant concentration, disinfection profile/bench mark data, log of viral inactivation/removal, contact time, contact value, pH, temperature.
Filter Backwash Information	Information about filter backwash that is returned to the treatment plant influent (e.g., information on: recycle/schematic status, alternative return location, corrective action requirements, and recycle flows and frequency).
Sample-Specific Information	
Sampling Point Identification Code	A sampling point identifier established by the state, unique within each applicable facility, for each applicable sampling location (e.g., entry point to the distribution system). This information enables occurrence assessments that address intra-system variability.
Sample Identification Number	Identifier assigned by state or the laboratory that uniquely identifies a sample.
Sample Collection Date	Date the sample is collected, including month, day, and year.
Sample Type	Indicates why the sample is being collected (e.g., compliance, routine, repeat, confirmation, additional routine samples, duplicate, special, special duplicate, etc.).
Sample Analysis Type Code	Code for type of water sample collected. • Raw (Untreated) water sample • Finished (Treated) water sample For lead and copper only: • Source • Tap *For TCR Repeats only; indicator of sampling location relative to sample point where positive sample was originally collected:* • Upstream • Downstream • Original
Contaminant	Contaminant name, 4-digit SDWIS contaminant identification number, or Chemical Abstracts Service (CAS) Registry Number for which the sample is being analyzed.
Sample Analytical Result - Sign	The sign indicates whether the sample analytical result was: • (<) "less than" means the contaminant was not detected or was detected at a level "less than" the minimum reporting level (MRL). • (=) "equal to" means the contaminant was detected at a level "equal to" the value reported in "Sample Analytical Result - Value." • (+) "positive result" *(For RTCR data, only positive E. coli result sign to be included.)*
Sample Analytical Result - Value	Actual numeric (decimal) value of the analysis for the chemical results, or the MRL if the analytical result is less than the contaminant's MRL. *(For the TCR and RTCR, TC and E. coli will indicate presence/absence, and positive E. coli will have numeric results.)*

Exhibit 2: Requested Data Categories	
Sample Analytical Result - Unit of Measure	Unit of measurement for the analytical results reported (usually expressed in either µg/L or mg/L for chemicals; or pCi/l or mrem/yr for radiological contaminants). *(Not required for TCR and RTCR data)*
Sample Analytical Method Number	EPA identification number of the analytical method used to analyze the sample for a given contaminant.
Minimum Reporting Level (MRL) - Value	MRL refers to the lowest concentration of an analyte that may be reported. *(Not required for TCR and RTCR data)*
MRL - Unit of Measure	Unit of measure to express the concentration value of a contaminant's MRL. *(Not required for TCR and RTCR data)*
Source Water Monitoring Information	Total organic carbon (TOC), including percent TOC removal, TOC removal summary, pH, alkalinity, monitoring data entered as individual results or included in DBP (or monthly operating report) summary records, alternative compliance criteria, results from round 2 monitoring under LT2 ESWTR (including *Cryptosporidium, E. coli*, turbidity, or state-approved alternate indicators).
Sample Summary Reports	Sample summaries for DBPRs, SWTRs, GWR corrective actions, and the Lead and Copper Rule (LCR) associated with analytical result records. Values used for compliance determination [e.g., turbidity (combined effluent/individual effluent), disinfectant residual levels in treatment plant and distribution system, treatment technique information, HPC, etc.]

1. For systems that are no longer required to individually monitor for nitrite, results should be reported for total nitrate plus nitrite (expressed as N) as SDWIS Analyte Code 1038 in lieu of individual results for nitrite and nitrate.

C. How do I prepare my data for submission to EPA?

We want to make this process as easy as possible for states that are volunteering to submit monitoring and treatment technique data. EPA developed and refined a SDWIS/State extraction tool, which runs a customized query to pull data for those using SDWIS/State. We believe this would be the most efficient (i.e., easiest) method of data extraction for those states using some or all of SDWIS/State. Currently, some states store and manage their data in more than one database. If it is easier for you to provide the electronic data for all contaminants that are stored in your data system, EPA can help you with a global extraction of the data. Please send inquiries to SixYearData@cadmusgroup.com. All data will be transmitted to EPA using the same process your state currently uses to submit your SDWIS data (see section D, below, for details).

Extracting data that are stored in SDWIS/State:

SDWIS/State Extract Tool: EPA has developed the SDWIS/State Extract Tool to extract the relevant data (specified in Exhibit 2) from a SDWIS/State database. The tool consists of three parts: PWS Inventory and Treatment, Analytical Results and Calculated Compliance Values. The first two parts were used in the Six-Year Review 3. States that use SDWIS/State for data storage and management and are interested in using the SDWIS/State extract tool can email SixYearData@cadmusgroup.com for instructions to download the extraction tool. EPA believes the extract tool would be the easiest mode of extraction for data that are stored in SDWIS/State. For the data transfer step, please see section D, below.

Note: If you have not migrated all drinking water monitoring data for the applicable period (January 2012 through December 2019) to SDWIS/State, a separate data submission to include all data back to January 2012 is requested, so that the data included in the Agency's Six-Year Review analysis is as complete and comparable as possible.

Automated Data Quality Assurance (QA) with SDWIS/State Extraction Tool*:* EPA has built in several automated data QA checks with this extraction tool. For example, the extraction tool will check for duplicate data, and analytical results that are >10 times the MCL. Before the data are extracted from SDWIS/State, the extraction tool runs these queries and returns a "flagged item report" for any data that meet these and other criteria that may indicate anomalies in your data (e.g., incorrect units of measurement, or data entry error). If there are entries in your "flagged item report," we strongly encourage you to review and resolve as many of these flags as possible before re-running and submitting your data. Doing this will help ensure your submitted data are of the highest quality possible. In addition, we will run these and other QA checks once we receive your data; so, by addressing flags before submitting your data, you will reduce the number of questions that need to be resolved once your data are submitted.

Format for Non-SDWIS/State data:

Virtually any electronic file format is acceptable. It would be ideal for states to submit their data sets in one of the following file formats: dBase™ (.dbf); Microsoft Access (.accdb); comma or tab delimited files (such as .csv or .txt), or; Microsoft Excel (.xls). However, you can submit the requested data "as is," by simply sending the compliance monitoring and treatment technique records in whatever structure or condition in which they are currently stored and submitting that copy of the electronic data to EPA. If it is easier for you to provide your entire electronic data set, EPA will extract the needed data. If you have further questions about this data submission, you can contact SixYearData@cadmusgroup.com.

Documentation:

EPA requests that your submission also include, at a minimum, a brief description of the basic format and structure of each data set, and definitions of all data elements, column/row headings, codes, acronyms, etc., used in each data set. (Note: EPA does not need this information if you are using SDWIS/State. EPA already has this information.) This “data dictionary” information will reduce the amount of time needed for questions and clarification later. EPA's primary goal is to obtain the most complete national occurrence and treatment technique data possible, and the Agency will work with the states to reconcile data questions where needed. If your data set is incomplete, or there are known anomalies, such as those that may have been identified by the SDWIS/State extract tool, it would be helpful if an explanation of these issues were included with your transmittal.

D. How do I send my data to EPA?

Regardless of whether data is stored in SDWIS/State, states can submit data using the same process your state currently uses to submit your SDWIS data. (Note some states using SDWIS/State may store some of the requested data outside of SDWIS/State and they should also follow these instructions.) Zip your files extracted from SDWIS/State or from some other location and name them SIXYEAR_REVIEW_XX.ZIP where XX is the Primacy Agency identifier. For example, Maryland would submit a file SIXYEAR_REVIEW_MD.ZIP. The files extracted from SDWIS/State by the extraction tool get zipped up and saved together with this naming convention. For more information on how to submit the data please see instructions file accompanying the extraction tool.

E. When do these data need to be submitted?

To help EPA meet its Six-Year Review 4 statutory timeframe and to allow ample time for data compilation, analysis and documentation of results, EPA requests that the data be submitted by September 30, 2020.

II. Background Information Regarding EPA's Occurrence Data Request

A. Why is EPA requesting this data?

The 1996 Safe Drinking Water Act (SDWA) Amendments require EPA to review and revise, if appropriate, existing National Primary Drinking Water Regulations (NPDWRs) at least every six years (i.e., the Six-Year Review). EPA is requesting monitoring and treatment technique data for NPDWRs to support the fourth Six-Year Review. Without an understanding of where and at what levels regulated drinking water contaminants are occurring in public drinking water, EPA cannot assess any potential need to revise the regulations.

In addition, the 1996 SDWA Amendments require the Agency to maintain a national drinking water contaminant occurrence database (i.e., the National Contaminant Occurrence Database or NCOD) using occurrence data for both regulated and unregulated contaminants. Through this data collection, EPA will be fulfilling various requirements set forth by Congress in the 1996 SDWA Amendments.

B. How will these data be used?

EPA's OGWDW will use the data to estimate the occurrence of regulated contaminants in public drinking water systems and to evaluate the number of people exposed and exposure reductions. Combined with results of other technical analyses (such as assessments of contaminant health effects), the results of the occurrence and exposure analyses will be used to help determine whether potential revisions to the current drinking water regulations are likely to maintain or provide for greater protection of public health for people served by public water systems. This data will help EPA to make well-informed regulatory decisions.

Once the Agency publishes the review results for the Six-Year Review 4, these data will be made publicly available. The procedures used to analyze these data will reflect those established and refined in prior Six-Year Reviews. Copies of EPA's Six-Year Review occurrence findings and methodology reports can be obtained at: http://water.epa.gov/lawsregs/rulesregs/regulatingcontaminants/sixyearreview/index.cfm. These documents contain the first, second, and third Six-Year Review occurrence findings and provide direct examples of the types of occurrence analyses that will be conducted using the compliance monitoring data you submit.

C. Why is it important to submit these data?

Regulatory decisions and the public health protection resulting from these decisions are improved by both the quality and quantity of the data. Each state that submits data can be

directly represented in any national occurrence estimates we develop. The Six-Year Review 4 data will be used in the review of existing regulations to determine whether current NPDWRs remain appropriate or if revisions should be considered. All data will undergo a comprehensive quality assurance/quality control (QA/QC) process required for the Six-Year Review 4 occurrence analyses. A copy of the resulting final, QA/QC reviewed contaminant data sets will be posted on the EPA Six-Year Review website.

D. What will happen once the data are submitted?

EPA will conduct uniform QA/QC assessments on each data set. Contaminant-specific analytical values will be assessed as part of the QA/QC review. For example, assessment of all analytical values for a specific contaminant will help identify possible unit errors or the presence of outliers. The data will also be checked for duplicate data entries (as defined by multiple rows of identical data elements) with duplicates excluded from the analysis, as needed. Identified errors that do not have straight-forward solutions will be addressed through consultations with the appropriate data management staff.

Based on EPA's experience with monitoring information provided by states for the prior Six-Year Reviews, the Agency will likely need to contact some states to address questions regarding the data format and content (e.g., outlier values, or missing or undefined data elements). EPA will document the QA/QC process and all edits or changes made to the submitted monitoring data.

After the data have undergone QA/QC editing and formatting, the data sets will be aggregated into national contaminant occurrence data sets for each contaminant. The national aggregate data sets will be used to generate statistical estimations of national occurrence. When the analyses are completed and reported, the data will be placed in the NCOD and in the docket to support any Six-Year Review 4 decisions.

Treatment information will also be compiled and assessed to support the Six-Year Review 4 decisions. However, the format of this information may not lend itself to analogous quantitative analysis and national summaries. Assessment of this information will be conducted and may be summarized in a more qualitative manner. Water system facility characteristics, filtration type, treatment technique information, and filter backwash information may be used to further inform the results of the occurrence data assessment.

Appendix B: User Guide to Downloading and Using Six-Year Review 4's Microbial and Disinfection Byproducts Information Collection Request data files from EPA's Website

This appendix includes a user guide for downloading and using the SYR4 MDBP data from EPA's website: https://www.epa.gov/dwsixyearreview/microbial-and-disinfection-byproduct-data-files-2012-2019-epas-fourth-six-year. In addition, instructions on importing the SYR4 MDBP datasets and data dictionary for the MDBP datasets are also included in this Appendix (see section 5 and 6, respectively).

Some datasets are described as "full" or reduced datasets. Full datasets are defined as all the QA-ed data for that contaminant. A "reduced" dataset is a subset of the QA-ed data that has be created by combining data from two or more contaminants to fit a particular purpose, e.g. pairing microbial contaminant data with its associated disinfectant residual and eliminating non-paired records is called a reduced dataset.

The data files are posted online in several zip files. Each zip file includes text files for multiple contaminants/parameters. The number of records and contaminants/parameters included in each file vary. The user may want to compare their counts of records downloaded for each contaminant of interest to the table of records provided in this user guide's exhibits to ensure that all of the records were correctly downloaded and imported. Note that these record counts reflect the data after the QA/QC process. For a list of data elements included in the data posted online, refer to Section 6 of this Appendix – Data Dictionary for Six-Year 4 ICR MDBP Database.

The remainder of this document is organized as follows:

Section 1. Background Information on Six-Year Review 4 Data

To support the national contaminant occurrence and exposure assessments performed under the fourth Six-Year Review process (SYR4), EPA collected compliance monitoring data and treatment technique information from public water systems (PWSs) for regulated drinking water contaminants. EPA conducted a voluntary data request from state and other primacy agencies to obtain compliance monitoring data and treatment technique information necessary to analyze national contaminant occurrence in support of SYR4. This data request was conducted through the Information Collection Request (ICR) process. EPA requested primacy agencies submit their Safe Drinking Water Act (SDWA) compliance monitoring data and treatment technique information collected between January 2012 and December 2019. For the MDBP data particularly, EPA collected the data recorded in the individual states databases related to these National Primary Drinking Water Regulations: Stage 1 and Stage 2 Disinfectants and Disinfection Byproducts Rules, Surface Water Treatment Rules, Interim Enhanced Surface Water Treatment Rule, and Long-Term 1 Enhanced Surface Water Treatment Rule. For more information on the process undertaken to request the voluntary submission of compliance monitoring data and treatment technique information by the states, see the fourth Six-Year Review ICR (84 FR 58381, USEPA, 2019).

EPA received compliance monitoring data and treatment technique information from both SDWIS state and non-SDWIS state users. For states that use SDWIS/state, EPA developed a tool, available upon request from primacy agencies, to extract the requested data identified in the SYR4 ICR from a SDWIS/State database. In all, 46 states and 13 other primacy agencies provided compliance monitoring data that included parametric records. Thirty-five states, Washington D.C, and six regional tribal entities used the extraction tool to extract all or some of their data. The 17 states/entities not using SDWIS/State submitted their compliance monitoring data and treatment technique "as is," resulting in a variety of formats, including dBase, MS Excel, XML, MS Access, and comma-delimited. With the exception of two states whose data were downloaded from their publicly available website (California and Florida), all states submitted their data over the Internet via EPA's Central Data Exchange. All data was conformed to a similar format with consistent units of measurement for consistency. For more details about the collection and formatting of SYR4 MDBP data see the main chapters of this document.

EPA conducted a quality assurance and control evaluation of these data submitted by primacy agencies, and assembled these data into a database. As noted in the main chapters, that only the data that passed the QA/QC process are posted online.

Section 2: Disinfection Byproducts

2A. Description of the Data Files for Disinfection Byproducts

The SYR4 disinfection byproducts (DBPs) datasets include data text files of regulated disinfection byproducts such as total trihalomethanes (TTHM) and sum of five haloacetic acids (HAA) along with the individual speciated DBPs within these groups, respectively.

2B. Data Files Posted for Disinfection Byproducts

The following SYR4 ICR data text files are located in their designated zip file at https://www.epa.gov/dwsixyearreview/microbial-and-disinfection-byproduct-data-files-2012-2019-epas-fourth-six-year under **Disinfection Byproducts**:

SYR4_THMs.zip *file contains individual files for:*

- Total Trihalomethanes (TTHM)
- Bromodichloromethane
- Bromoform
- Chloroform
- Dibromochloromethane

SYR4_HAAs.zip *file contains individual files for:*

- Haloacetic Acids (HAA5)
- Bromoacetic acid
- Dibromoacetic acid
- Dichloroacetic acid
- Monochloroacetic acid
- Trichloroacetic acid

SYR4_Bromate_Chlorite.zip *contains individual files for:*

- Bromate
- Chlorite

2C. Disinfection Byproducts Data Records

Exhibit 1 provides a count of states, total number of sample records and systems for each disinfection byproduct whose data is posted online.

Note the speciation data is higher for TTHM than HAA5. There were two more states that provided speciated THM results as compared to speciated HAA results. About 11,000 systems provided speciated THM data but not speciated HAA data and there are about 200 systems with speciated HAA data but no speciated THM data. In addition, the number of PWSs that provided

speciated TTHM data was higher than number of PWSs providing TTHM. There are approximately 8,000 systems that have data for the speciated THMs but not TTHM whereas there are only about 7,000 systems with data for TTHM but not the speciated THMs.

Exhibit 1: Number of Disinfection Byproduct Data Records and Zip filename(s)

Contaminant	Analyte ID	Number of States/Entities with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Disinfection Byproducts-Full Datasets					
TOTAL TRIHALOMETHANES (TTHM)	2950	57	1,089,557	46,297	SYR4_THMs.zip
DIBROMOCHLOROMETHANE	2944	46	981,059	47,172	SYR4_THMs.zip
BROMOFORM	2942	46	976,412	47,129	SYR4_THMs.zip
CHLOROFORM	2941	46	981,289	47,403	SYR4_THMs.zip
BROMODICHLOROMETHANE	2943	46	977,561	47,196	SYR4_THMs.zip
HALOACETIC ACIDS (HAA5)	2456	57	1,005,235	43,577	SYR4_HAAs.zip
DIBROMOACETIC ACID	2454	44	720,986	36,121	SYR4_HAAs.zip
DICHLOROACETIC ACID	2451	44	721,017	36,134	SYR4_HAAs.zip
MONOCHLOROACETIC ACID	2450	44	720,474	36,113	SYR4_HAAs.zip
TRICHLOROACETIC ACID	2452	44	720,706	36,125	SYR4_HAAs.zip
BROMOACETIC ACID	2453	44	720,595	36,095	SYR4_HAAs.zip
BROMATE	1011	38	23,298	444	SYR4_Bromate_Chlorite.zip
CHLORITE	1009	33	87,995	514	SYR4_Bromate_Chlorite.zip

Section 3: Disinfection Byproduct Related Parameters

3A. Description of Data Files Posted for Disinfection Byproduct Related Parameters

This DBP related parameters data posted includes data files for: total organic carbon (TOC), total alkalinity, Paired TOC-Alkalinity, pH, DOC, SUVA, and UV-absorbance.
Full datasets are provided for TOC, Alkalinity, pH, DOC, SUVA, and UV-absorbance.
A reduced dataset, Paired TOC-alkalinity, was created that included, for each treatment plant (listed as a water system facility in Exhibit 2), the average monthly concentrations of TOC and alkalinity in source (raw) water paired with the corresponding average finished water concentration of TOC. The "paired" TOC-alkalinity dataset was created to evaluate the percent removal of TOC using the SYR4 data and joined the average monthly TOC concentration with the average monthly alkalinity concentration for individual water system facilities when possible. This paired dataset is directly related to the treatment technique requirements for TOC removal under the Stage 1 DBPR. Historical efforts to evaluate the paired TOC-alkalinity data were described in Six-Year Review 3 Technical Support Document for Disinfectants/Disinfection Byproducts Rules" (USEPA, 2016).

Exhibit 3 contains the list of data elements, column names, and a brief description of the data for each data element included in the "paired" TOC-alkalinity dataset. For a list of data elements

included in the "full" TOC, alkalinity and pH datasets, refer to Section 6 Data Dictionary for the SYR4 ICR Database.

3B. Data Files Posted for Disinfection Byproduct Related Parameters

The following SYR4 ICR data text files are located in their designated zip file at https://www.epa.gov/dwsixyearreview/microbial-and-disinfection-byproduct-data-files-2012-2019-epas-fourth-six-year under **Disinfection Byproducts Related Parameters**:

SYR4_DBP_Related Parameters.zip *contains individual files for:*

- DOC
- pH
- SUVA
- Total Alkalinity
- Total Organic Carbon (TOC) (raw and finished TOC)
- Paired TOC and Alkalinity
- UV_absorbance

Exhibit 2: "Paired TOC-Alkalinity" Dataset Field Names and Definitions

Data Element	Column Name	Description
Public Water System Identification Number (PWSID)	NUMBER0	The code used to identify each PWS. The code begins with the standard 2- character postal state abbreviation or region code; the remaining 7 numbers are unique to each PWS in the state.
Sample Collection Date (Month)	Month	Month (1 through 12).
Sample Collection Date (Year)	Year	Year (2012 through 2019).
Retail Population-served	Population Served	Retail population served by the water system.
Federal Public Water System Type Code	System Type	Water system type according to federal requirements. C = Community water system NTNC = Non-transient non-community water system
Source Water Type	Source Water Type	Primary water source for the water system. GU = Ground water Under Direct Influence of Surface Water GW = Ground Water GWP = Purchased Ground Water SW = Surface Water SWP = Purchased Surface Water
Facility Identification Code	Water Facility ID	Unique identifier for each water system facility.
State Facility Identification Code	State Facility ID	Identifier for each water system facility that is unique within a particular state.
State Assigned Identification Code	State Assigned ID	A state-assigned value which identifies the water system facility.
Raw water TOC average concentration	Avg Of Raw TOC (mg/L)	Monthly average (in mg/L) total organic carbon (TOC) concentration in raw water.

Data Element	Column Name	Description
Raw water alkalinity average concentration	Avg Of Raw Alkalinity (mg/L)	Monthly average (in mg/L) alkalinity concentration in raw water.
Finished water TOC average concentration	Avg Of Finished TOC (mg/L)	Monthly average (in mg/L) total organic carbon (TOC) concentration in finished water.

3C. Disinfection Byproduct Related Parameters Data Records

Exhibit 3 provides a count of states, total number of sample records and systems for Total Organic Carbon (TOC)(raw and finished), Alkalinity, Paired TOC-Alkalinity, pH, DOC, SUVA, UV-absorbance.

The count of systems for raw and finished TOC samples are counted separately, so systems with samples in both categories are counted twice. Raw samples are identified as samples taken at source water sampling points. Records were marked as raw if [SOURCE_TYPE_CODE] = 'RW' OR [SOURCE_TYPE_CODE] was NULL but water system facility type code = ('IG' or 'IN' or 'RS' or 'SP' or 'WL' or 'CC'). Records were marked as finished if SOURCE_TYPE_CODE = 'FN' or SOURCE_TYPE_CODE was NULL but water facility type code = ('CW' or 'DS' or 'PF' or 'ST' or 'TM' or 'TP').

Note that within the "Full" TOC text file, raw/finished designations are not assigned. However, with the Paired TOC-alkalinity record reduced dataset, raw and finished designations are assigned.

Exhibit 3: Number of TOC, Alkalinity, pH, DOC, SUVA, and UV-absorbance Data Records and Zip Filename(s)

Contaminant	Analyte ID	Number of States/Entities with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Disinfection Byproduct Related Parameters - Full Datasets					
TOTAL ORGANIC CARBON (TOC)	2920	49	440,197	3,156	SYR4_DBP_Related Parameters.zip
RAW TOC	2920	42	188,358	2,494	SYR4_DBP_Related Parameters.zip
FINISHED TOC	2920	38	155,558	1,999	SYR4_DBP_Related Parameters.zip
ALKALINITY	1927	51	429,397	18,140	SYR4_DBP_Related Parameters.zip
PH	1925	52	632,821	28,660	SYR4_DBP_Related Parameters.zip
SUVA	2923	2	8,026	59	SYR4_DBP_Related Parameters.zip
UV-absorbance	2922	3	6,061	60	SYR4_DBP_Related Parameters.zip
DOC	2919	3	5,908	76	SYR4_DBP_Related Parameters.zip
Disinfection Byproduct Related Parameters – Reduced Dataset					
Paired TOC-alkalinity record[1]	N/A	33	92,666	1,192	SYR4_DBP_Related Parameters.zip

[1] The "paired" TOC-alkalinity dataset includes average monthly concentrations of TOC and alkalinity in source (raw) water paired with the corresponding average finished water concentrations of TOC.

Section 4: Microbial Contaminants, Microbial Related Parameters, and Associated Disinfectant Residuals

4A. Description of Data Files for Microbial Contaminants, Microbial Related Parameters, and Associated Disinfectant Residuals Data

Data for three microbial contaminants (total coliforms (TC), *Escherichia coli* (EC), and fecal coliform (FC)) were collected from 2012 to 2019 for SYR4. The total coliform datasets are separated into individual files by each year of data collected due the large volume of data collected on TC.

Reduced datasets were created to pair microbial data (TC, EC, FC) with associated disinfectant residual for disinfecting systems. Disinfectant residual results are shown as free residual chlorine and total chlorine in theses reduced datasets. These disinfectant residual data were collected on the same date and location as the microbial parameters. Additional data for disinfectant residual include datasets for chlorine and chloramine; those data were not reported as being collected on the same date and location as the microbial parameters.

Note that the TC/EC/FC data files contain the monitoring records under Total Coliform Rule/Revised Total Coliform Rule for systems with all source water types. The HPC disinfectants, disinfectant residuals, paired microbes disinfectant residuals files contain the monitoring records under SWTRs for surface water systems.

4B. Data Files Posted for Microbial Contaminants, Microbial Related Parameters, and Associated Disinfectant Residuals

The following SYR4 ICR data text files are located in their designated zip file at https://www.epa.gov/dwsixyearreview/microbial-and-disinfection-byproduct-data-files-2012-2019-epas-fourth-six-year under **Microbial Contaminants, Microbial Related Parameters, Associated Disinfectant Residuals**:

SYR4_TC.zip *contains individual files for:*

- Total Coliform_2012
- Total Coliform_2013
- Total Coliform_2014
- Total Coliform_2015
- Total Coliform_2016
- Total Coliform_2017
- Total Coliform_2018
- Total Coliform_2019

SYR4_EC_FC_HPC_Giardia.zip *contains individual files for:*

- *Escherichia coli* (EC)

- Fecal coliform (FC)
- *Giardia Lamblia*
- Heterotrophic Plate Count (HPC)

SYR4_Disinfectant Residuals.zip *contains individual files for:*

- Chloramines
- Chlorine
- Chlorine dioxide
- Free Residual Chlorine
- Residual Chlorine
- Total Chlorine

SYR4_Paired Microbes_DR (Disinfectant Residuals).zip *contains individual files for:*

- Paired EC_DR
- Paired FC_DR
- Paired TC_DR_2012
- Paired TC_DR_2013
- Paired TC_DR_2014
- Paired TC_DR_2015
- Paired TC_DR_2016
- Paired TC_DR_2017
- Paired TC_DR_2018
- Paired TC_DR_2019

4C. Microbial Contaminants, Microbial Related Parameters, and Associated Disinfectant Residuals Data Records

Exhibit 4 is a list of data elements included in the TC, EC, FC and Reduced Dataset for Analysis of Disinfecting Systems with Disinfectant Residual records.

Exhibit 4: Field Names and Descriptions for Paired Microbial Contaminants and Associated Disinfectant Residuals Datasets

Data Element	Column Name	Description
Presence Indicator Code	PRESENCE_ INDICATOR_CODE	Indication of whether results of an analysis were positive or negative for TC, EC and FC. • P = Presence • A = Absence.
Residual Field Free Chlorine	RESIDUAL_FIELD_ FREE_CHLORINE_MG_L	Amount of free chlorine residual (in mg/L) found in the water after disinfectant has been applied. These concentrations were measured in the field at the same time and location as coliform samples (TC-EC-FC samples).
Residual Field Total	RESIDUAL_FIELD_	Amount of total chlorine residual (in mg/L) found in the

Data Element	Column Name	Description
Chlorine	TOTAL_CHLORINE_ MG_L	water after disinfectant has been applied. These concentrations were measured in the field at the same time and location as coliform samples (TC-EC-FC samples).

Exhibit 5 provides a count of states, total number of sample records and systems for TC, EC, FC, and their associated free and total chlorine residual concentrations for both the full and reduced datasets.

Exhibit 5: Number of Microbial Contaminants, Microbial Related Parameters, and Associated Disinfectant Residuals Data Records and Zip Filename(s)

Contaminant	Analyte ID	Number of States/ Entities with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Microbes and Disinfectants – Full Datasets					
TOTAL COLIFORM (2012)	3100	54	2,349,687	102,423	SYR4_TC.zip
TOTAL COLIFORM (2013)	3100	54	2,398,740	102,713	SYR4_TC.zip
TOTAL COLIFORM (2014)	3100	56	2,521,212	105,515	SYR4_TC.zip
TOTAL COLIFORM (2015)	3100	56	2,513,937	104,532	SYR4_TC.zip
TOTAL COLIFORM (2016)	3100	57	2,656,932	113,099	SYR4_TC.zip
TOTAL COLIFORM (2017)	3100	57	2,780,743	114,328	SYR4_TC.zip
TOTAL COLIFORM (2018)	3100	57	2,849,385	114,954	SYR4_TC.zip
TOTAL COLIFORM (2019)	3100	57	2,675,476	111,385	SYR4_TC.zip
E. COLI (EC)	3014	57	7,175,363	93,728	SYR4_EC_FC_HPC_Giardia.zip
FECAL COLIFORM (FC)	3013	40	16,818	1,835	SYR4_EC_FC_HPC_Giardia.zip
HETEROTROPHIC BACTERIA (HPC)	3001	16	135,081	595	SYR4_EC_FC_HPC_Giardia.zip
GIARDIA LAMBLIA	3008	15	4628	229	SYR4_EC_FC_HPC_Giardia.zip
LEGIONELLA		0	0	0	N/A
CHLORINE[1]	0999	19	6,100,133	4,438	SYR4_Disinfectant Residuals.zip
TOTAL CHLORINE	1000	1	125,788	741	SYR4_Disinfectant Residuals.zip
CHLORAMINE[1]	1006	9	78,664	198	SYR4_Disinfectant Residuals.zip
RESIDUAL CHLORINE	1012	4	179,599	572	SYR4_Disinfectant Residuals.zip
FREE RESIDUAL CHLORINE[1]	1013	3	2,000,997	4,044	SYR4_Disinfectant Residuals.zip
CHLORINE DIOXIDE	1008	9	12,752	28	SYR4_Disinfectant Residuals.zip
Microbes and Associated Disinfectant Residuals - Reduced Dataset					
E. coli (EC) with Associated Disinfectant Residuals	3014	49	3,079,032	28,091	SYR4_Paired Microbes_DR.zip
Fecal Coliform (FC) with Associated Disinfectant Residuals	3013	24	5,966	534	SYR4_Paired Microbes_DR.zip
Total Coliform (TC) paired with Associated Disinfectant Residuals (2012)	3100	43	1,165,209	30,950	SYR4_Paired Microbes_DR.zip
Total Coliform (TC) paired with Associated Disinfectant Residuals (2013)	3100	44	1,173,926	31,132	SYR4_Paired Microbes_DR.zip

Contaminant	Analyte ID	Number of States/ Entities with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Total Coliform (TC) paired with Associated Disinfectant Residuals (2014)	3100	46	1,218,722	31,865	SYR4_Paired Microbes_DR.zip
Total Coliform (TC) paired with Associated Disinfectant Residuals (2015)	3100	47	1,241,995	31,880	SYR4_Paired Microbes_DR.zip
Total Coliform (TC) paired with Associated Disinfectant Residuals (2016)	3100	48	1,274,211	34,654	SYR4_Paired Microbes_DR.zip
Total Coliform (TC) paired with Associated Disinfectant Residuals (2017)	3100	50	1,331,868	37,217	SYR4_Paired Microbes_DR.zip
Total Coliform (TC) paired with Associated Disinfectant Residuals (2018)	3100	50	1,480,354	41,053	SYR4_Paired Microbes_DR.zip
Total Coliform (TC) paired with Associated Disinfectant Residuals (2019)	3100	50	1,498,050	38,029	SYR4_Paired Microbes_DR.zip

[1] Reported independently of the coliform sample results.

Section 5: Instructions on Importing Microbial and Disinfection Byproduct Datasets

These text files are tab delimited and have no text qualifier. Field names are included in the first row of each file. The data are available for download for each parameter and should be imported into a data management system that supports large datasets for analysis.

5A: Downloading Data Files (Note that instructions may vary depending on the version and software used to import data.)

1. Begin by reviewing the Data Field Names and Definitions (Section 6- Data Dictionary for the SYR4 ICR Database).

2. Access the SYR4 MDBP data files by going to https://www.epa.gov/dwsixyearreview/six-year-review-4-microbial-and-disinfection-byproduct-data-2012-2019.

3. Click on the desired zip file and select "Save As" to save the file to your computer.

4. Navigate to the location on your computer where you saved the zip file and extract the zip file contents by clicking "Open with" and using WinZip or a similar file compression software

5B: Importing Data into Microsoft Excel

Using Microsoft Excel 2013 or a newer version is recommended due to the size of the dataset(s). Note the following MDBP data files are too large to import into Microsoft excel: TTHM, HAA, Free Residual Chlorine, Total Chlorine, all TC files, EC, and all Paired microbes and Disinfectant Residual files.

5. Open a blank workbook in Microsoft Excel.

6. In the workbook, select Data among the tabs at the top of the page.

7. On the far left, top of the screen, go to the Get External Data section and select From Text.

8. You will be prompted to select a text file. Locate the text files you extracted in Step 4, and click "Import" on the text file of interest.

9. A preview of the file text converted to a table will appear. At the top, verify that File Origin (depending on your computer's operating system) displays "10000: Western European (Mac)" or "1252: Western European (Windows)." Select "Tab" as the Delimiter and "Based on first 200 rows" as the Data Type Detection. Click Load To...

10. In the next window, choose "Table" under Select how you want to view the data in your workbook. Select "Existing worksheet" for where to put the data and verify the table's origin cell origin displays as "=A1." Click OK.

11. A "Queries & Connections" window will appear on the right of the screen as Excel generates the new table. This step may take several minutes.

12. Save the Excel spreadsheet file once the table generation is complete.

5C: Importing Data into R

1. Open a blank R script.

2. Using the function read.delim(), import the text file using the following format:

    a. [analyte name] <- read.delim(file = [filepath], header = TRUE)

    Example: bromoform <- read.delim(file = "C:/Users/[username]/Desktop/SYR4-Microbes /SUMMARY_MDBPS_BROMOFORM.txt", header = TRUE)

3. Check the data frame that is generated to ensure correct formatting.

4. NOTE: data columns that should be in date format will be imported as character type. To

fix, include the line "df$DATE <- as.Date.character(df$DATE, format = "%d-%b-%y")" in the R code, replacing df with the name of the dataframe, and DATE with the name of the column containing date information.

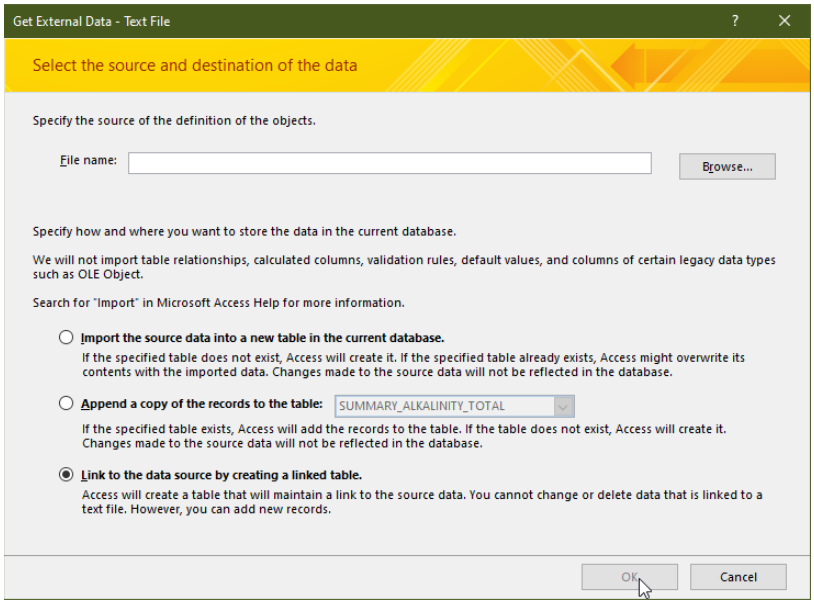
5D: Importing Data into Microsoft Access

1. Open a blank database in Microsoft Access.

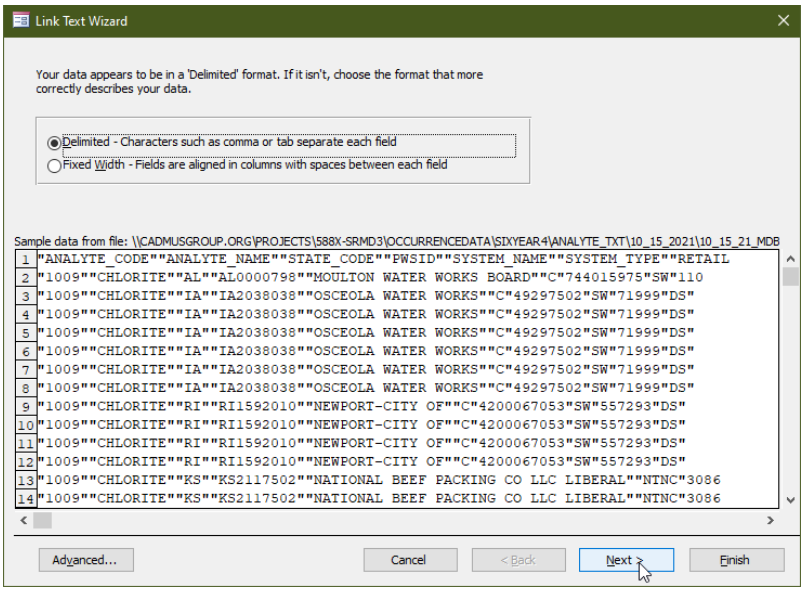
2. In the database, select External Data among the tabs at the top of the page.

3. On the far left, top of the screen, go to the New Data Source dropdown and select From File > Text File."

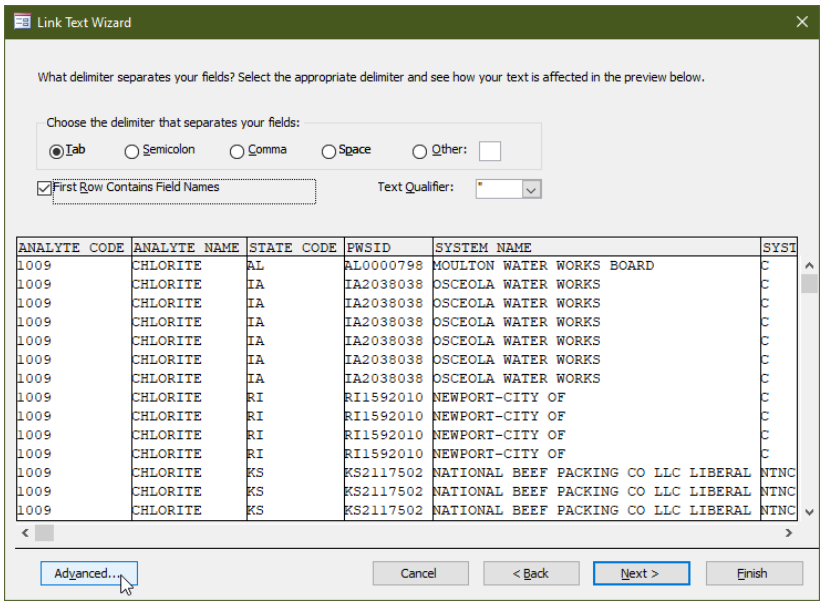
4. You will be prompted to select a text file. Locate the text files you extracted in Step 4, and with the following options: "import the source data into a new table in the current database", or "Link to the data source by creating a linked table". You can choose either method, but note that linking the file will maintain a smaller database size. Click OK.



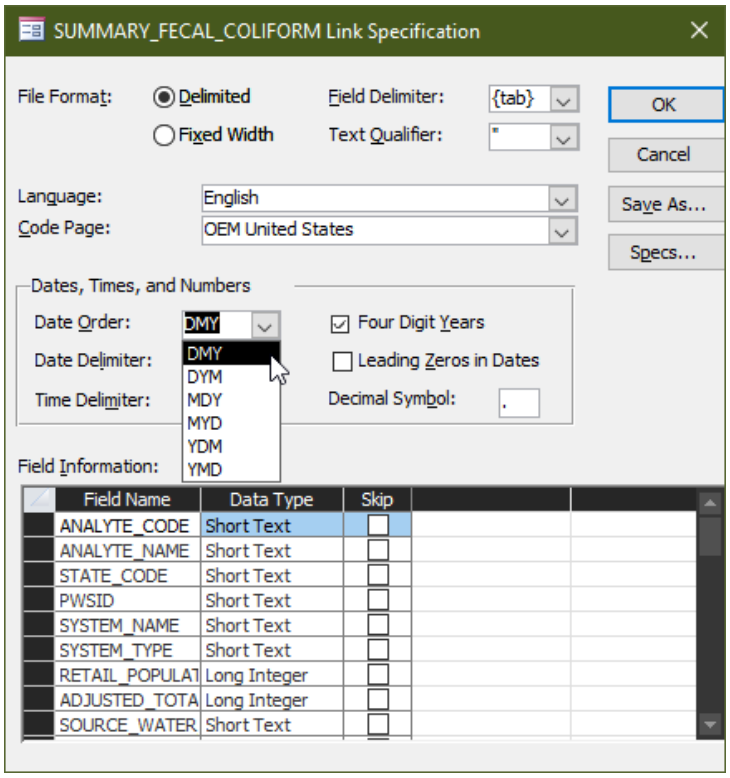
5. The Link (or Import) Text Wizard will appear. The default settings will be displayed and should have Delimited selected as the data format. Select Next>.



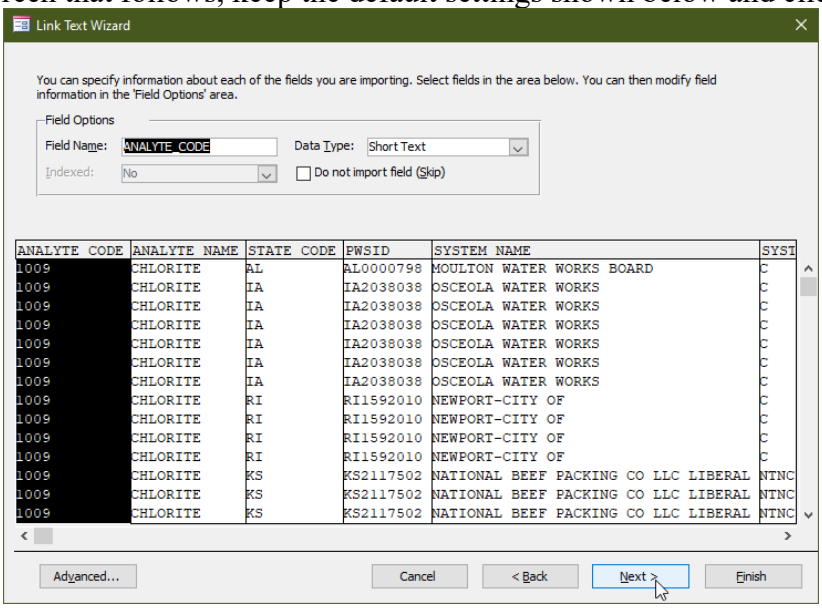
6. Default settings will display next and should have "Tab" selected as the delimiter. Select the checkmark box next to "First Row Contains Field Names." Next, click "Advanced…".



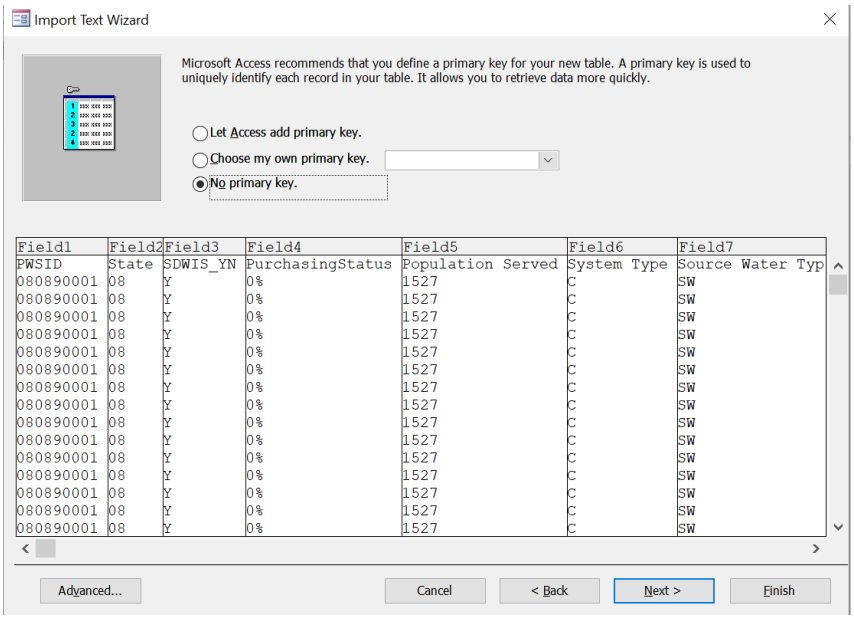
7. The Link (or Import) Specification window will appear. In the Dates, Times, and Numbers section, set the Date Order value to "DMY."



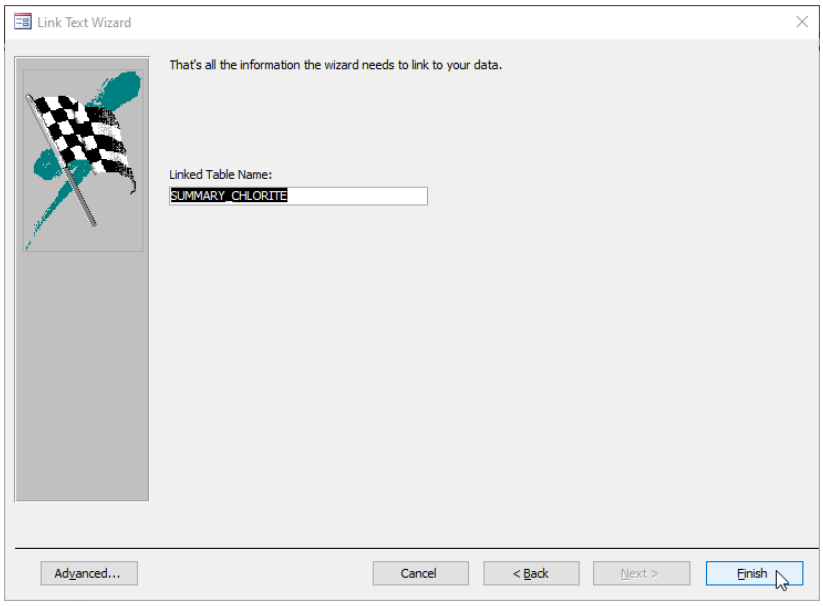
8. On the screen that follows, keep the default settings shown below and click Next>.



If you are importing instead of linking, a window will pop up related to setting a primary key. The default is set to “Let Access add a primary key”. Check “No primary key” and click Next >.



9. A final screen will appear. Enter a meaningful name for the linked/imported table. This field will be auto-populated with the name of the linked file. Click Finish.



Part Two: Filtering and Formatting Data in Excel

10. To efficiently search, have cell A1 selected, choose “Data” among the tabs on the top of the page and click on “Filter.” Each header title for each column now will have a small dropdown arrow displayed.

11. Filtering the data: a. If you want to look for a specific public water system, click the dropdown arrow for “PWSID” or “System Name.” Within the search field, type the name

and select from the displayed list. b. If you want to search for a different public water system, click the dropdown arrow and "Clear Filter from PWSID" or "Clear Filter from System Name." c. If you want to filter the data by contaminant, select "Analyte Name."

12. Multiple filters can be applied for example, allowing you to look for an individual water system's data for a specific contaminant of interest.

13. De-select Filter in the top menu bar and the entire database will again be displayed.

14. Note, all column formats are imported as the default General formatting. Column formats must be individually, manually changed in Excel after the download is complete to aid in data analysis. Use the Home screen in excel, highlight the column and select the format from the drop down menu. Suggested formats are:

Text fields	Analyte Name
	State Code
	PWSID
	System Name
	System Type
	Source Water Type
	Water Facility Type
	Sampling Point Type
	Source Type Code
	Sample Type Code
	Laboratory Assigned ID
	Sample Collection Date
	Detection Limit Unit
	Detection Limit Code
	Value Unit
	Presence Indicator Code
Numeric fields	Analyte ID
	Retail Population Served
	Adjusted Total Population Served
	Water Facility ID
	Sampling Point ID
	Six Year ID
	Sample ID
	Detection Limit Value
	Detect
	Value
	Residual Field Free Chlorine mg/L
	Residual Field Total Chlorine mg/L

Section 6: Data Dictionary for the SYR4 ICR Database

Exhibit 6 below contains a list of the data elements, column names and a brief description of the data for each data element included in each of the SYR4 ICR data text files.

Exhibit 6: Six-Year 4 Data Field Names and Definitions

Column Name	Data Element	Description
ANALYTE_CODE	Contaminant Identification Code	4-digit Safe Drinking Water Information System (SDWIS) contaminant identification number for which the sample is being analyzed.
ANALYTE_NAME	Contaminant Name	Common name of contaminant for which the sample is being analyzed.
STATE_CODE	State Code	2- digit state code. Note that the state code "IM" refers to non-community water system data from the State of Illinois.
PWSID	Public Water System Identification Number (PWSID)	The code used to identify each PWS. The code begins with the standard 2- character postal state abbreviation or region code; the remaining 7 numbers are unique to each PWS in the state.
SYSTEM_NAME	System Name	Name of the PWS.
SYSTEM_TYPE	Federal Public Water System Type Code	A code to identify whether a system is: • Community Water System (C); • Non-Transient Non-Community Water System (NTNC); or • Transient Non-Community Water System (NC).
RETAIL_POPULATION_SERVED	Retail Population served	Retail population served by a system.
ADJUSTED_TOTAL_POPULATION_SERVED[1]	Adjusted Total Population-served	Total population served by a system, adjusted to reduce double-counting of population served by purchasing water systems.
SOURCE_WATER_TYPE	Source Water Type	Type of water at the source. Source water type can be: • Ground water (GW); • Surface water (SW); • Purchased Surface Water (SWP); • Purchased Ground Water (GWP); • Ground Water Under Direct Influence of Surface Water (GU); or • Purchased Ground Water Under Direct Influence of Surface Water (GUP).
WATER_FACILITY_ID	Facility Identification Code	A unique identifier for each water system facility.
WATER_FACILITY_TYPE	Water Facility Type	Type of water system facility: • CC = Consecutive Connection; • CH = Common Headers; • CW = Clear Well; • DS = Distribution System; • IG = Infiltration Gallery; • IN = Intake; • OT = Other; • PC = Pressure Control; • PF = Pumping Facility;

Column Name	Data Element	Description
		• RS = Reservoir; • SI = Surface Impoundment; • SP = Spring; • SS = Sampling Station; • ST = Storage; • TM = Transmission Main (Manifold); • TP = Treatment Plant; • WH = Well Head; • WL = Well; or • XX = unknown.
SAMPLING_POINT_ID	Sampling Point Identification Code	A unique identifier for each sampling point location.
SAMPLING_POINT_ TYPE	Sampling Point Type	Location type of a sampling point: • DS = Distribution System; • EP = Entry point; • FC = First Customer; • FN = Finished Water Source; • LD = Lowest Disinfectant Residual; • MD = Midpoint in the Distribution System; • MR = Point of Maximum Residence; • PC = Process Control; • RW = Raw Water Source; • SR = Source Water Point; • UP = Unit Process; or • WS = Water System Facility Point
SOURCE_TYPE_CODE	Source Type Code	Type of water source, based on whether treatment has taken place. Source type can be: • Finished (FN); • Raw (RW); or • Unknown (null or X).
SAMPLE_TYPE_CODE	Sample Type Code	Type of sample: • CO = Confirmation; • MR = Maximum Residence Time; • RP = Repeat; • RT = Routine; • ST = Split; • MS = Matrix spike; • TG = Triggered; or • FB = Field Blank.
LABORATORY_ ASSIGNED_ID	Laboratory Assigned Identification Number	Unique lab identification, used to link up the total coliform positive (TC+) and *E. coli* / fecal coliform samples.
SIX_YEAR_ID	Six Year ID	Unique identifier for each analytical result.
SAMPLE_ID	Sample Identification Number	Identifier assigned by state or the laboratory that uniquely identifies a sample.
SAMPLE_ COLLECTION_DATE	Sample Collection Date	Date the sample was collected, including month, day, and year.
DETECTION_LIMIT	Detection Limit	Limit below which the specific lab indicated they could not

Column Name	Data Element	Description
_ VALUE	Value	reliably measure results for a contaminant with the methods and procedures used by the lab.
DETECTION_LIMIT _ UNIT	Detection Limit Unit	Units of the detection limit value.
DETECTION_LIMIT _ CODE	Detection Limit Code	Indicates the type of Detection Limit reported in the Detection Limit Value column (e.g., the Minimum Reporting Level, Laboratory Reporting Level, etc.)
DETECT	Sample Analytical Result - Sign	The sign indicates whether the sample analytical result was: • (0) "less than" means the contaminant was not detected or was detected at a level "less than" the MRL. • (1) "equal to" means the contaminant was detected at a level "equal to" the value reported in "Sample Analytical Result - Value."
VALUE	Sample Analytical Result - Value	For detections, this field is equal to the actual numeric (decimal) value of the analysis for the chemical result; for non-detections, this field is blank.
UNIT	Sample Analytical Result - Unit of Measure	Unit of measurement for the analytical results reported (usually expressed in either µg/L or mg/L for chemicals; or pCi/L for radionuclides).
PRESENCE_ INDICATOR_CODE	Presence Indicator Code	Indication of whether results of an analysis were positive or negative for TC, EC and FC. • P = Presence • A = Absence.
RESIDUAL_FIELD_ FREE_CHLORINE_ MG_L	Residual Field Free Chlorine	Amount of free chlorine residual (in mg/L) found in the water after disinfectant has been applied. These concentrations were measured in the field at the same time and location as coliform samples (TC-EC-FC samples).
RESIDUAL_FIELD_ TOTAL_CHLORINE_ MG_L	Residual Field Total Chlorine	Amount of total chlorine residual (in mg/L) found in the water after disinfectant has been applied. These concentrations were measured in the field at the same time and location as coliform samples (TC-EC-FC samples).

[1] Information for total population was not received. This value was generated for wholesale systems using buyer-seller relationships and calculating the adjusted total population served.

Appendix C: Six-Year Review 4 Microbial and Disinfection Byproduct Data Records by State

Appendix C contains exhibits with the number of Six-Year 4 Microbial and Disinfection Byproducts (MDBP) data records by category by state. The following is a list of the exhibits:

Exhibit C-1: Number of Microbial Contaminants (Total Coliform, *E.coli*, Fecal Coliform, *Giardia Lamblia*) Data Records by State
Exhibit C-2: Number of Total Trihalomethanes (TTHM) Data Records by State
Exhibit C-3: Number of Haloacetic acids (HAAs) Data Records by State
Exhibit C-4: Number of Chlorite and Bromate Data Records by State
Exhibit C-5: Number of Disinfection Byproduct Related Parameters Data Records by State

Exhibit C-1: Number of Microbial Contaminants (Total Coliform, *E.coli*, Fecal Coliform, *Giardia Lamblia*) Data Records by State

State	Total Coliform	*E. Coli*	Fecal Coliform	*Giardia lamblia*
Alaska	103,898	65,414	2,823	0
Alabama	284,580	90,650	6	60
Arkansas	394,314	6,089	0	0
American Samoa	13,186	13,184	0	0
Arizona	219,468	42,862	26	0
California	0	0	0	0
Colorado	352,349	204,889	24	0
Connecticut	382,725	219,854	14	23
Washington, D.C.	13,693	9,648	0	0
Delaware	70,366	13,042	3	0
Florida	2,342,672	350	21	0
Hawaii	16,035	13,593	13	0
Iowa	425,813	207,287	3	0
Idaho	193,935	14,451	3	7
Illinois	1,526,019	651,044	235	0
Indiana	398,481	13,702	0	0
Kansas	279,741	208,962	11	926
Kentucky	427,911	1,949	0	0
Louisiana	179,619	147,417	14	0
Massachusetts	0	0	0	0
Maryland	60,832	34,081	1,092	0
Maine	145,575	77,758	2	56
Minnesota	225,927	15,141	12	398
Missouri	601,095	282,873	1	0
Northern Mariana Islands	13,364	12,020	0	0
Montana	260,675	216,652	4,942	0

State	Total Coliform	*E. Coli*	Fecal Coliform	*Giardia lamblia*
North Carolina	926,048	628,350	4	0
North Dakota	95,674	936	1	0
Nebraska	218,891	153,908	0	0
New Hampshire	155,791	156,191	0	0
New Jersey	935,126	22,684	64	785
Navajo Nation	7,447	6,789	0	0
Nevada	81,129	13,499	0	256
New York	541,960	88,232	438	153
Ohio	1,022,164	112,768	112	0
Oklahoma	398,661	236,786	0	0
Oregon	477,951	16,078	1	0
Pennsylvania	854,438	246,817	730	0
Rhode Island	61,041	44,878	1,792	1
South Carolina	9,563	7,510	2	0
South Dakota	117,852	66,507	0	0
Tennessee	91,984	84	1,449	0
Texas	2,637,545	1,359,122	1,132	0
Utah	297,343	92,252	10	4
Virginia	703,226	343,357	150	8
Vermont	126,345	106,484	1	192
Washington	949,429	224,822	191	0
Wisconsin	693,211	545,150	0	0
West Virginia	187,869	4,082	11	1,689
Wyoming	108,011	87,686	1,409	0
Region 1 - Tribes	2,722	2,708	0	0
Region 2 - Tribes	912	84	0	0
Region 4 - Tribes	3,591	57	3	70
Region 5 - Tribes	19,648	145	1	0
Region 6 - Tribes	21,655	10,140	47	0
Region 7 - Tribes	2,468	2,237	0	0
Region 8 - Tribes	21,291	13,740	24	0
Region 9 - Tribes	21,764	17,844	0	0
Region 10 - Tribes	21,089	524	1	0

Exhibit C-2: Number of Total Trihalomethanes (TTHM) Data Records by State

State	TTHM	Chloroform	Bromoform	Bromodichloromethane	Dibromochloromethane
Alaska	4,546	4,557	4,548	4,559	4,558
Alabama	41,159	5,361	5,392	5,371	5,377
Arkansas	21,380	25,444	25,446	25,446	28,031
American Samoa	161	0	0	0	0

State	TTHM	Chloroform	Bromoform	Bromodichloromethane	Dibromochloromethane
Arizona	15,050	545	548	549	552
California	143,888	149,737	150,126	150,620	150,591
Colorado	24,986	11,810	11,812	11,816	11,815
Connecticut	11,128	20,566	20,567	20,567	20,565
Washington, D.C.	240	7	7	7	7
Delaware	3,031	5,389	5,386	5,389	5,389
Florida	48,865	0	0	0	0
Hawaii	2,732	2,649	2,674	2,671	2,670
Iowa	14,736	14,708	14,708	14,708	14,708
Idaho	4,822	366	357	358	364
Illinois	43,203	42,414	42,430	42,444	42,439
Indiana	19,042	2,867	2,871	2,870	2,871
Kansas	15,283	13,449	13,439	13,449	13,449
Kentucky	26,111	0	0	0	0
Louisiana	35,015	35,257	35,267	35,257	35,261
Massachusetts	22,494	15,614	15,540	15,621	15,586
Maryland	12,715	9,838	9,679	9,782	9,699
Maine	4,588	4,031	4,020	4,024	4,020
Minnesota	0	17,244	16,988	17,159	17,098
Missouri	20,303	26,742	26,743	26,743	26,743
Northern Mariana Islands	245	0	0	0	0
Montana	7,503	7,764	7,765	7,763	7,763
North Carolina	44,268	35,821	35,862	35,797	35,784
North Dakota	4,170	4,163	4,164	4,164	4,164
Nebraska	7,256	7,260	7,260	7,260	7,260
New Hampshire	6,394	9,776	9,766	9,774	9,774
New Jersey	32,013	46,887	47,020	46,925	47,230
Navajo Nation	1,369	0	0	0	0
Nevada	6,176	6,853	6,852	6,850	6,853
New York	48,574	44,873	44,696	44,777	44,732
Ohio	42,844	46,461	46,298	46,219	46,333
Oklahoma	30,421	30,611	30,614	30,615	30,616
Oregon	13,218	0	0	0	0
Pennsylvania	49,995	48,859	46,398	47,023	47,054
Rhode Island	3,175	1,477	1,477	1,475	1,477
South Carolina	19,816	19,818	19,818	19,817	19,815
South Dakota	4,095	0	0	0	0
Tennessee	22,006	0	0	0	0
Texas	113,625	154,480	154,480	154,479	154,480
Utah	9,277	7,852	7,840	7,826	7,796
Virginia	27,661	28,375	27,769	28,335	28,175

State	TTHM	Chloroform	Bromoform	Bromodichloromethane	Dibromochloromethane
Vermont	4,173	6,810	6,811	6,811	6,812
Washington	21,349	29,032	28,368	26,916	27,996
Wisconsin	9,976	16,404	15,506	16,223	16,046
West Virginia	13,049	13,028	13,026	13,022	13,022
Wyoming	4,803	3,293	3,287	3,293	3,293
Region 1 - Tribes	259	0	0	0	0
Region 2 - Tribes	62	0	0	0	0
Region 4 - Tribes	0	0	0	0	0
Region 5 - Tribes	543	0	0	0	0
Region 6 - Tribes	828	828	828	827	828
Region 7 - Tribes	137	71	72	72	73
Region 8 - Tribes	1,573	661	660	661	662
Region 9 - Tribes	2,243	0	0	0	0
Region 10 - Tribes	983	1,237	1,227	1,227	1,228

Exhibit C-3: Number of for Haloacetic acids (HAAs) Data Records by State

State	HAA5	Monochloro-acetic Acid	Dichloroacetic Acid	Trichloroacetic Acid	Monobromo-acetic Acid	Dibromoacetic Acid
Alaska	4,222	4,205	4,207	4,202	4,197	4,205
Alabama	41,186	0	0	0	0	0
Arkansas	21,435	21,445	21,442	21,439	21,442	21,442
American Samoa	158	0	0	0	0	0
Arizona	14,956	518	517	517	517	526
California	86,262	83,511	84,239	84,067	83,471	84,002
Colorado	23,814	9,290	9,290	9,413	9,290	9,290
Connecticut	10,777	8,925	8,925	8,925	8,924	8,905
Washington, D.C.	241	4	4	3	4	4
Delaware	2,981	3,014	3,016	3,016	3,013	3,014
Florida	48,591	0	0	0	0	0
Hawaii	2,223	2,144	2,161	2,160	2,162	2,164
Iowa	14,730	14,704	14,703	14,703	14,704	14,704
Idaho	4,039	164	165	164	164	167
Illinois	43,147	42,393	42,393	42,360	42,393	42,392
Indiana	19,024	0	0	0	0	0
Kansas	15,225	13,410	13,416	13,413	13,416	13,413
Kentucky	26,113	0	0	0	0	0
Louisiana	34,991	35,004	34,999	34,993	35,011	35,001
Massachusetts	21,448	15,525	15,558	15,545	15,495	15,485
Maryland	12,645	6,196	6,138	6,163	6,149	6,166
Maine	4,097	2,497	2,499	2,499	2,497	2,496

State	HAA5	Monochloro-acetic Acid	Dichloroacetic Acid	Trichloroacetic Acid	Monobromo-acetic Acid	Dibromoacetic Acid
Minnesota	0	11,390	11,501	11,469	11,385	11,415
Missouri	20,221	19,896	19,896	19,896	19,896	19,896
Northern Mariana Islands	209	0	0	0	0	0
Montana	3,824	3,809	3,807	3,805	3,809	3,810
North Carolina	44,217	35,794	35,720	35,721	35,802	35,783
North Dakota	4,161	4,155	4,155	4,155	4,155	4,155
Nebraska	2,903	2,903	2,903	2,903	2,903	2,903
New Hampshire	3,576	3,501	3,498	3,497	3,501	3,501
New Jersey	31,995	32,005	32,004	32,003	32,003	32,005
Navajo Nation	1,360	0	0	0	0	0
Nevada	5,265	5,238	5,238	5,232	5,235	5,237
New York	42,009	37,146	37,179	37,168	37,151	37,173
Ohio	42,508	42,510	42,510	42,483	42,529	42,462
Oklahoma	30,320	27,331	27,331	27,327	27,332	27,334
Oregon	13,221	0	0	0	0	0
Pennsylvania	50,166	15,471	15,487	15,483	15,481	15,479
Rhode Island	3,117	1,442	1,442	1,442	1,442	1,442
South Carolina	19,820	19,819	19,819	19,819	19,819	19,816
South Dakota	4,087	0	0	0	0	0
Tennessee	21,996	0	0	0	0	0
Texas	113,097	113,098	113,098	113,098	113,098	113,098
Utah	9,290	7,120	7,118	7,118	7,124	7,126
Virginia	27,387	21,656	21,724	21,732	21,677	21,646
Vermont	4,055	4,055	4,055	4,055	4,055	4,055
Washington	21,330	21,879	21,549	21,410	22,039	21,964
Wisconsin	9,848	9,850	9,848	9,847	9,849	9,849
West Virginia	13,021	12,990	12,992	12,990	12,995	12,989
Wyoming	3,755	2,246	2,248	2,247	2,246	2,249
Region 1 - Tribes	260	0	0	0	0	0
Region 2 - Tribes	55	0	0	0	0	0
Region 4 - Tribes	0	0	0	0	0	0
Region 5 - Tribes	476	0	0	0	0	0
Region 6 - Tribes	827	783	783	784	783	783
Region 7 - Tribes	127	47	49	49	47	49
Region 8 - Tribes	1,307	397	397	397	397	397
Region 9 - Tribes	2,146	0	0	0	0	0
Region 10 - Tribes	974	994	994	994	993	994

Exhibit C-4: Number of Chlorite and Bromate Data Records by State

State	Chlorite	Bromate
Alaska	0	203
Alabama	5,396	0
Arkansas	1,862	192
American Samoa	0	0
Arizona	2,418	601
California	1,520	6,065
Colorado	2,823	739
Connecticut	393	152
Washington, D.C.	0	0
Delaware	0	73
Florida	0	0
Hawaii	0	0
Iowa	2,128	94
Idaho	13	49
Illinois	1,897	222
Indiana	0	267
Kansas	4,933	651
Kentucky	1,786	0
Louisiana	0	0
Massachusetts	2,414	1,050
Maryland	31	0
Maine	350	214
Minnesota	66	189
Missouri	5,034	225
Northern Mariana Islands	0	0
Montana	5	779
North Carolina	920	540
North Dakota	0	201
Nebraska	195	30
New Hampshire	0	0
New Jersey	1,233	721
Navajo Nation	0	0
Nevada	2,031	886
New York	348	88
Ohio	1,391	364
Oklahoma	3,864	672
Oregon	3	235
Pennsylvania	15,344	306
Rhode Island	867	0
South Carolina	0	1
South Dakota	0	0
Tennessee	0	0

State	Chlorite	Bromate
Texas	26,960	4,289
Utah	2	314
Virginia	1,406	1,430
Vermont	0	0
Washington	0	2
Wisconsin	0	1,079
West Virginia	84	93
Wyoming	0	133
Region 1 - Tribes	0	0
Region 2 - Tribes	0	0
Region 4 - Tribes	0	0
Region 5 - Tribes	0	0
Region 6 - Tribes	0	96
Region 7 - Tribes	0	0
Region 8 - Tribes	47	0
Region 9 - Tribes	231	24
Region 10 - Tribes	0	29

Exhibit C-5: Number of Disinfection Byproduct Related Parameters Data Records by State

State	Alkalinity	pH	All Total Organic Carbon (TOC)	Raw Water TOC	Finished Water TOC	Free Chlorine Data[1]	Total Chlorine Data[1]	Free Chlorine Data[2]	Total Chlorine Data[2]
Alaska	1,533	191	2,915	524	169	55,417	498	176	0
Alabama	18,574	3,279	17,239	8,489	8,596	182,333	2,687	3,274	1,179
Arkansas	0	0	0	0	0	2	371,859	0	0
American Samoa	2	2	0	0	0	7,491	23	0	0
Arizona	3,540	776	4,221	2,114	2,107	7	5	0	0
California	0	125,308	32,893	18,884	13,637	0	0	0	0
Colorado	8,960	1,000	16,549	0	0	321,103	28,287	1,437	22
Connecticut	16,188	147,504	8,074	4,033	3,784	338,697	52,495	277	10
Washington, D.C.	1	41	17	6	0	8,688	13,666	64	159
Delaware	6,363	6,684	336	127	209	51,299	11,139	4,207	807
Florida	73	6,919	1	0	0	1,036,993	0	0	0
Hawaii	60	2	3	1	2	14,454	213	0	0
Iowa	2,587	216	6,555	2,838	0	315,592	366,036	11	11
Idaho	1,436	476	2,001	181	180	86,182	333	0	0
Illinois	12,766	3,111	17,715	8,857	8,858	828,654	521,300	0	0
Indiana	4,416	937	6,945	3,579	3,366	133,771	124,094	0	0
Kansas	10,654	3,114	15,085	7,479	7,510	143,389	129,953	0	0
Kentucky	15,521	3,331	27,990	13,997	13,993	351,946	133,281	0	0

State	Alkalinity	pH	All Total Organic Carbon (TOC)	Raw Water TOC	Finished Water TOC	Free Chlorine Data[1]	Total Chlorine Data[1]	Free Chlorine Data[2]	Total Chlorine Data[2]
Louisiana	5,594	5,854	1	1	0	126,762	67,051	51,477	28,978
Massachusetts	10	4,785	0	0	0	0	0	0	0
Maryland	3,479	1,646	4,908	2,412	0	1,569	1,112	0	0
Maine	6,958	4,533	1,744	977	767	17,767	41,444	0	0
Minnesota	4,948	22,506	3,320	0	0	0	0	0	0
Missouri	11,408	11,710	17,671	8,749	8,922	293,054	420,150	0	0
Northern Mariana Islands	0	261	0	0	0	6,196	1	0	0
Montana	4,109	724	6,456	2,561	2,774	59,138	21,696	5	47
North Carolina	30,586	39,246	26,460	12,990	13,470	631,245	315,687	0	0
North Dakota	1,554	472	2,176	1,083	1,093	0	0	0	0
Nebraska	0	7	989	0	289	27,968	46,685	733	141
New Hampshire	1,774	3,454	1,133	0	0	0	0	0	0
New Jersey	44,603	94,278	12,340	6,185	5,814	172,293	48,946	0	0
Navajo Nation	221	239	23	0	0	6,348	105	0	0
Nevada	4,995	7,692	1,027	539	488	41,407	226	0	0
New York	2,674	5,523	8,890	4,716	2,789	125,909	2,035	1,675	0
Ohio	1,769	2,402	156	1	14	761,430	802,905	0	0
Oklahoma	22,713	2,282	33,140	454	105	169,182	184,063	27,762	23,977
Oregon	3,145	0	7,699	4,600	3,097	346,652	5	0	0
Pennsylvania	60,459	86,188	33,174	17,903	0	180,486	87,219	0	0
Rhode Island	1,492	588	1,513	755	712	25,221	25,629	0	5
South Carolina	5,477	2,116	10,714	5,287	5,354	0	64	0	0
South Dakota	0	0	0	0	0	0	0	0	0
Tennessee	259	1,438	0	0	0	89,236	0	0	0
Texas	63,232	11,262	55,684	27,835	27,849	0	0	588,072	820,698
Utah	3,123	662	4,961	2,415	2,508	120,687	1,858	12,594	0
Virginia	20,257	10,176	20,652	10,312	10,340	596,794	6,035	8,278	0
Vermont	308	281	184	88	96	60,180	25,395	3,058	2,352
Washington	681	203	29	3	26	0	0	0	0
Wisconsin	6,786	6,187	3,442	0	0	327,565	0	0	0
West Virginia	10,692	2,507	17,588	4,807	4,246	13,231	176,456	0	132
Wyoming	1,634	111	3,071	1,584	1,452	70,446	12,085	14	4
Region 1 - Tribes	21	18	0	0	0	2,571	43	0	0
Region 2 - Tribes	0	0	0	0	0	575	24	0	0
Region 4 - Tribes	0	0	0	0	0	3,176	0	0	0
Region 5 - Tribes	0	0	0	0	0	16,392	0	0	0
Region 6 - Tribes	110	0	224	114	99	16,865	3,831	409	195
Region 7 - Tribes	83	4	176	0	0	1,480	906	0	0
Region 8 - Tribes	865	55	1,483	738	739	13,841	3,639	43	49

State	Alkalinity	pH	All Total Organic Carbon (TOC)	Raw Water TOC	Finished Water TOC	Free Chlorine Data[1]	Total Chlorine Data[1]	Free Chlorine Data[2]	Total Chlorine Data[2]
Region 9 - Tribes	400	361	379	0	0	16,145	41	0	0
Region 10 - Tribes	304	159	251	140	104	9,787	10,664	0	0

[1] Free and Total Chlorine data associated with Total Coliform

[2] Free and Total Chlorine data associated with DBPs