Data Management and Quality Assurance/Quality Control Process for the Fourth Six-Year Review Information Collection Request Dataset

Office of Water (4607M)
EPA-815-R-24-017
February 2024

Disclaimer

This document is not a regulation. It is not legally enforceable and does not confer legal rights or impose legal obligations on any party, including EPA, States, or the regulated community.

While EPA has made every effort to ensure the accuracy of any references to statutory or regulatory requirements, the obligations of the interested stakeholders are determined by statutes, regulations, or other legally binding requirements, not this document. In the event of a conflict between the information in this document and any statute or regulation, this document would not be controlling.

Executive Summary

The 1996 Amendments to the Safe Drinking Water Act (SDWA) require that the Environmental Protection Agency (EPA) "shall, not less often than every 6 years, review and revise, as appropriate, each national primary drinking water regulation." The National Primary Drinking Water Regulations (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 the SYR process, EPA generally issues an Information Collection Request (ICR) to the states and other primacy agencies to collect the recent data information that public water systems (PWSs) have submitted per requirements of NPDWRs. The data are 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 SYR 4 ICR see EPA's website: https://www.epa.gov/dwsixyearreview/six-year-review-4-drinking-water-standards-information-collection-request.

This report describes how the compliance monitoring data and treatment technique information for EPA's fourth Six-Year Review (SYR 4) of NPDWRs were obtained, evaluated, and formatted, where necessary, to enable national contaminant occurrence estimates. In addition, this document describes the data requested and received, data quality issues, and data management efforts to make it consistent and usable for subsequent analyses.

EPA conducted data management and quality assurance (QA) evaluations on the data received for contaminants evaluated for the SYR 4 to establish a national compliance monitoring and treatment technique dataset consisting of data from 59 states/primacy agencies (46 states plus territories, Washington, D.C., and tribes). The compliance monitoring data and treatment technique information for these 59 states/primacy agencies comprise more than 71 million analytical records from approximately 140,000 PWSs, which serve more than 301 million people nationally.[1] The ICR dataset for the fourth Six-Year Review (SYR 4 ICR dataset) is the largest and most comprehensive compliance monitoring data and treatment technique information dataset ever compiled and analyzed under EPA's drinking water program.

Information regarding the acquisition, storage, and management of the SYR 4 ICR data is presented in Sections 2 through 4 of this report. Detailed descriptions of the QA evaluations and data preparation for analyses are presented in Section 5 and Section 6, respectively. Additional technical information related to the SYR 4 ICR dataset is presented in the appendices to this report.

[1] These statistics reflect the portion of the overall dataset representing compliance monitoring samples collected for requested regulated contaminants. The initial dataset, including data not specifically requested by EPA but submitted voluntarily by some states, was comprised of over 83 million records from approximately 142,000 PWSs.

For the national contaminant occurrence assessments for the Chemical Phase Rules and Radionuclides Rule conducted in support of EPA's fourth Six-Year Review of NPDWRs, refer to the USEPA (2024a) report entitled *Analysis of Regulated Contaminant Occurrence Data from Public Water Systems in Support of the Fourth Six-Year Review of National Primary Drinking Water Regulations: Chemical Phase Rules and Radionuclides Rules*. For more detailed information on the microbial contaminants' occurrence analysis, refer to USEPA (2024b) report entitled *Six-Year Review 4 Technical Support Document for Microbial Contaminant Regulations.* The final SYR 4 ICR datasets are posted online at: https://www.epa.gov/dwsixyearreview.

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

Appendices

Exhibits

Abbreviations and Acronyms

ADWR	Airline Drinking Water Rule
CAS	Chemical Abstracts Service
CHEMID	Four Digit SDWIS Code
CO	Confirmation
CWS	Community Water System
DBCP	1,2-Dibromo-3-chloropropane
DBP	Disinfection Byproduct
DBPR	Disinfection Byproduct Rule
D/DBPR	Disinfectants and Disinfection Byproducts Rule
DEHA	Di(2-ethylhexyl) adipate
DEHP	Di(2-ethylhexyl) phthalate
EC	*Escherichia coli (E. coli)*
EDB	Ethylene dibromide
eDWR	Electronic Drinking Water Report
EPA	Environmental Protection Agency (United States)
FBRR	Filter Backwash Recycling Rule
FC	Fecal Coliform
GAC	Granular Activated Carbon
GW	Ground Water
GWP	Ground Water Purchased
GWR	Ground Water Rule
GWUDI (or GU)	Ground Water Under Direct Influence (of Surface Water)
GUP	Purchased 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
IOC	Inorganic Contaminant
LCR	Lead and Copper Rule
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
MFL	Million Fibers per Liter
mg/L	Milligrams per Liter
mrem/yr	Millirem per year
MR	Maximum Residence
MRDL	Maximum Disinfectant Residual Level
MRL	Minimum Reporting Level
NPDWR	National Primary Drinking Water Regulation
NTNCWS	Non-Transient Non-Community Water System

Abbreviations and Acronyms (cont.)

PCBs	Polychlorinated Biphenyls
pCi/L	Picocuries per Liter
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
SOC	Synthetic Organic Contaminant
SW	Surface Water
SWP	Purchased Surface Water
SWTR	Surface Water Treatment Rule
SYR 4	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
VOC	Volatile Organic Contaminant

1 Introduction

This document describes how the compliance monitoring data and treatment technique information for the fourth Six-Year Review (SYR 4) were obtained, evaluated, and formatted, where necessary, to enable national contaminant occurrence estimates in support of the Environmental Protection Agency's (EPA) SYR 4 of National Primary Drinking Water Regulations (NPDWRs). In addition, this document describes the data requested and received, data quality issues, and modifications to the data to make it consistent and usable for subsequent analyses. The actual analyses performed are described in other reports, referenced further in this section.

The 1996 Amendments to the Safe Drinking Water Act (SDWA) require that the EPA "shall, not less often than every 6 years, review and revise, as appropriate, each national primary drinking water regulation," (Section 1412(b)(9)). The NPDWRs are often referred to as the national drinking water contaminant regulations or drinking water standards. The purpose of the Six-Year Review is to evaluate current information for regulated contaminants to determine if there is new information on health effects, treatment technologies, analytical methods, occurrence, 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.

National contaminant occurrence assessments were conducted in support of EPA's SYR 4, using data from National Compliance Monitoring Information Collection Request (ICR) dataset for the fourth Six-Year Review (SYR 4 ICR dataset). These compliance monitoring data and treatment technique information were provided to EPA by States[2] via the ICR process. The report *Analysis of Regulated Contaminant Occurrence Data from Public Water Systems in Support of the Fourth Six-Year Review of National Primary Drinking Water Regulations: Chemical Phase Rules and Radionuclides Rules* (USEPA, 2024a) provides complete details on the national contaminant occurrence assessments of the contaminants regulated by the Phase I, II, IIb, and V Rules, the Arsenic Rule, and the Radionuclides Rule conducted in support of EPA's SYR 4. Included in that report are detailed descriptions of the national contaminant compliance monitoring and treatment technique dataset compiled and the statistical analytical methods employed to generate national estimates of regulated contaminant occurrence in public drinking water systems.

Compliance monitoring data for rules concerning microbial contaminants, disinfectants, and disinfection byproducts were also collected under SYR 4. For more detailed information on the microbial contaminants' occurrence analysis, refer to *Six-Year Review 4 Technical Support Document for Microbial Contaminant Regulations* (USEPA, 2024b). Occurrence analyses of disinfectants, disinfection byproducts, and certain microbial contaminants were not included in SYR 4 because these NPDWRs were identified as candidates for revision under Six-Year Review 3. However, the occurrence information collected under SYR 4 will be used to inform potential revisions to MDBP rules.

[2] In the remainder of this document, the terms "State" or "States" refers to primacy agencies in states of the United States, the District of Columbia, the Commonwealth of Puerto Rico, the Virgin Islands, Guam, American Samoa, the Commonwealth of the Northern Mariana Islands, the Trust Territory of the Pacific Islands, or an eligible Indian tribe.

The SYR 4 ICR data were received from the States in a variety of formats and data structures. The submitted data required restructuring to a uniform format to conduct the national contaminant occurrence analyses. EPA conducted a rigorous quality control evaluation of the data submitted by States, then assembled these data into a database. This document provides a description of the processes EPA used to assure overall data quality while developing the occurrence dataset for SYR 4 contaminant occurrence evaluations.

Specifically, this document describes the compliance monitoring data and treatment technique information requested and received and provides an overview of the data management and quality assurance/quality control (QA/QC) efforts used to prepare the data to analyze contaminant occurrence. Additional QA/QC processes specific to the microbial analyses are described in USEPA (2024b).

2 Data Acquisition

Compliance monitoring data and treatment technique information provide information critical to the Six-Year Review occurrence assessments. Without an understanding of where and at what levels these contaminants are occurring in public drinking water, EPA cannot assess the risk to public health and whether potential revisions are likely to maintain or improve public health protection. In addition, other compliance data can help in evaluating the effectiveness of current regulations.

The Federal Safe Drinking Water Information System database (SDWIS/Fed) contains information about public water systems (PWSs) and their violations of EPA's drinking water regulations. However, SDWIS/Fed does not receive nor store compliance monitoring data, which include non-detections as well as detections. To estimate national occurrence of regulated contaminants in PWSs, it was necessary to compile results from all compliance monitoring samples, including samples which showed analytical detections *and* non-detections. These data are collected by States but are not required to be submitted to SDWIS/Fed. Therefore, to obtain the compliance monitoring data and treatment technique information used in support of national occurrence assessments for SYR 4, EPA conducted a voluntary data call-in from the States, through the ICR process. For more information on the process undertaken to request the voluntary submission of compliance monitoring data and treatment technique information from States, see the SYR 4 ICR (84 FR 58381, USEPA, 2019).

Similar to prior rounds of the Six-Year Review, EPA contacted each State via letter requesting the voluntary submission of their compliance monitoring data for regulated chemical, radiological, microbial, and disinfection byproduct (DBP) contaminants and treatment technique information for all NPDWRs and related parameters that were collected between January 2012 and December 2019. See Appendix A for the compliance monitoring data and treatment technique information request letter.

EPA requested only information stored electronically (i.e., no paper records) that represented routine compliance monitoring data and treatment technique information. Exhibit 1 shows the regulated contaminants for which EPA requested data, and Exhibit 2 shows the requested data elements (e.g., columns, fields) for each sample result. See Appendix B: Crosswalk of Data Elements Requested for SYR 4 ICR and the SDWIS Data Element Names for a crosswalk table between the data elements requested and the actual data element names as they appear in SDWIS. In some cases, EPA did not receive any data for the elements and/or analytes requested.

Exhibit 1: List of Contaminants/Parameters Identified in SYR 4 ICR for which Data Were Requested from States

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)

Chemical Contaminants (Phase I, II, IIB, and V Rules; Arsenic Rule; Lead and Copper Rule)		
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
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 (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	
Ground Water Rule (GWR)		
Escherichia coli (E. coli)	Enterococci	Coliphage
Surface Water Treatment Rules (SWTRs)		
Chlorine**	*Cryptosporidium****	Heterotrophic Plate Count (HPC)
Chloramines**		
Filter Backwash Recycling Rule (FBRR)		
No specific occurrence data collected.		

Source: Attachment A to the letter EPA sent to each State to request voluntary submission of its compliance monitoring data and treatment technique information for regulated chemical, radiological, and microbiological contaminants. See Appendix A for the data request letter.

* 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 (LT2ESWTR), is being reviewed and will be available along with the SYR 4 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.

Data Category	Description
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).
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.)*
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).

Data Category	Description
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.]

Source: Attachment A to the letter EPA sent to each State to request voluntary submission of compliance monitoring data and treatment technique information for regulated chemical, radiological, and microbiological contaminants. See Appendix A for the data request letter.

[1] These are the data elements requested in the SYR 4 ICR. The "Data Category" and "Description" columns were intentionally descriptive rather than prescriptive. This allowed the States that do not use SDWIS/State the flexibility to provide as much information as possible. EPA accepted all data "as is" without prescribing structure or format.

About 78 percent of the 50 U.S. 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 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 has the flexibility to support the most complex primacy program implementation while maintaining a common core of data elements required for reporting to SDWIS/Fed. In an attempt to make the SYR 4 data submittal process as easy for States as possible, EPA developed a SDWIS/State Extraction Tool (also referred to as "extraction tool" throughout this document), which enabled States to run a customized query to pull the requested data from a SDWIS/State database maintained by those States. All of the States using SDWIS/State that submitted data to EPA for SYR 4 used the extraction tool to extract and compile the EPA-requested compliance monitoring and treatment technique data.

SDWIS/State supports the Electronic Drinking Water Report (eDWR) XML Schema used by laboratories throughout the nation to electronically report sample analytical results as structured data to SDWIS/State (for more information, see the full eDWR description and schema details https://exchangenetwork.net/data-exchange/electronic-drinking-water-reports/). As a result, States receive tabular data from laboratories that is batch-processed into SDWIS/State rather than manually entered. Consequently, States have a substantial amount of structured data available in SDWIS/State. In all, for SYR 4, 46 states and 13 other jurisdictions provided compliance monitoring data and treatment technique information that included parametric records. The seven States that did not provide data were Georgia, Michigan, Mississippi, New Mexico, Guam, Puerto Rico, and U.S. Virgin Islands.

Exhibit 3 lists the States that submitted SYR 4 data and indicates whether they used the extraction tool. Thirty-five states, Washington D.C, and six regional tribal entities used the extraction tool to transmit all or some of their chemical and microbial data; therefore, those datasets were all submitted in a similar format. The 17 States not using SDWIS/State submitted their compliance monitoring data and treatment technique information "as is," resulting in a variety of formats, including dBase, Excel, XML, Access, and comma-delimited. Apart from California, Colorado, and Florida, whose data were downloaded from their publicly available websites, all States submitted their data online via EPA's Central Data Exchange.

Exhibit 3: Summary of States that Provided Compliance Monitoring Data and Treatment Technique Information for SYR 4

	State/Entity Name		
States/Tribes that DID use the SDWIS/State Extraction 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 Extraction Tool	American Samoa California[1] Colorado[1] 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 SYR 4 data	Georgia Guam Michigan	Mississippi New Mexico	Puerto Rico U.S. Virgin Islands

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

3 Data Storage

EPA designed the SYR 4 ICR database similarly to SDWIS/State to house the data that States sent in response to the SYR 4 ICR data request. The SYR 4 ICR database is an Oracle relational database which consists of tables, relationships, import scripts, and other objects that support populating the database tables. Because of the likelihood of duplicate record identifiers in the source tables (e.g., same IDs from different States), most tables in the SYR 4 database contain a unique record identifier (i.e., a primary key). The unique record identifiers ensured that all relevant records were imported and that duplicate record identifiers present in the source data did not cause relevant records to be excluded. The relational database structure is an appropriate method of storing large volumes of data because it allows each table to store unique information. The SYR 4 database was designed to ensure information was not duplicated between tables and to maintain the logical relationships inherent to the data.

Exhibit 4 presents a description of the tables included in the SYR 4 ICR database. The database includes 17 primary tables and 2 transaction tables. The primary tables include SDWIS data elements, codes, and the compliance monitoring data and treatment technique information. The two additional transaction tables that relate to the QA/QC review were created by EPA to manage the QA/QC review effort. The QA/QC review documentation codes are called transactions in the database and are listed in Exhibit 4 with the word "transaction" in the title. For a list of all of the data elements included in each table, as well as available codes for each data element, refer to Appendix C: Data Dictionary for the SYR 4 Database.

Exhibit 4: Description of Tables Included in SYR 4 ICR Database

Table Name	Brief Description	Description of Contents of Table
T6YWS	Water system (Ws) table	Inventory information: PWSID, source water type, system type, population, etc.
T6YWSF	Water system facility (Wsf) table	Facility identification information: facility ID, facility type, etc.
T6YSPT	Sample point (Spt) table	Sample point identification information: sample point type, source type, etc.
T6YANALYTE	Analyte table	Analyte identification information: contaminant name, 4-digit chemical IDs, etc.
T6YSAR	Sample analytical result (Sar) table	Monitoring records: sample date, sample type code, analyte, concentration, reporting level, method, etc.
T6YDBPSUM	Disinfectant Byproduct summaries table	Summary used to enter sampling requirements and collection information in support of the SWTR/IESWTR and DBP rules.
T6YFANL	Facility analyte levels table	Includes information from primacy agencies where they specify and maintain M&R and level compliance values for an analyte at a water system facility.

Table Name	Brief Description	Description of Contents of Table
T6YSAMPSUM	Lead and Copper Rule and Total Coliform Rule sample summaries table	Quantity of each different type of sample (e.g., total samples collected, or number of repeat samples) and the result (e.g., total positive samples, total negative samples) of the sample analysis summaries for an analyte.
T6YCMCLV	Compliance Monitoring Compliance Level Violations	Includes information on calculated compliance values.
T6YCORACT	Corrective actions table	Includes information on corrective actions.
T6YMCL_MDL	Maximum contaminant level and minimum detection level table	Includes information on the values and units of the maximum contaminant level, four times the maximum contaminant level, minimum detection level, and one tenth the minimum detection level.
T6YWSFPLT	Treatment plant water system facilities table	Includes information on treatment plant facilities.
T6YTREATPROCESS	Treatments associated to treatment plants table	Includes information pertaining to the treatment processes and objectives.
T6YWSFFLOWS	Water system facility flows table	Includes information on the relationship or connection between the different water system facilities of a water system.
T6YWSFIND	Water system facility indicators table	Includes information on the recording of an indicator for a Water System Facility.
T6YWSIND	Water system indicators table	Includes information on the recording of an indicator for a Water System.
T6YWSPURCH	Water system buyers and sellers	Includes information on the purchase of water between water systems.
T6YSAR_TRANSACTION	Transaction table for sample analytical results	Flagged monitoring records: reason record was flagged, action taken on flagged record, response from the State (when available), and any other relevant notes/remarks. Some records have multiple entries in the transaction table if the record was flagged for more than one reason.
T6YWS_TRANSACTION	Transaction table for water systems	Flagged water systems: reason record was flagged, action taken on flagged record, response from the State (when available), and any other relevant notes/remarks. Some records have multiple entries in the transaction table if the record was flagged for more than one reason.

4 Data Management

This section provides descriptions of the data management tasks that were implemented to prepare the SYR 4 datasets for QA/QC review. The SDWIS/State Extraction Tool transferred the SDWIS/State data to Microsoft Access. Data from States that did not use the extraction tool were restructured into a similar format. The two subdatasets (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 in order to arrange them into the same format. After reformatting and transforming data from the non-SDWIS States, all data were combined into the final SYR 4 ICR dataset.

A status documentation file was maintained that included information for each State. Specifically, the status documentation described 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. Original datasets were saved and maintained exactly as received and stored in an EPA database. Any subsequent changes to a State's dataset were made to a copy of the original dataset, and all changes were documented.

4.1 Review of SYR 4 Dataset Content

Similar to prior rounds of the Six-Year Review, the first assessment of the submitted SYR 4 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 SYR 4 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 when 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, etc.

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 contacted the State via email for clarification. EPA created a flagged record report for each State to summarize questions regarding potential data quality concerns, data completeness, statewide waivers, and any other unique aspects of their dataset. In addition, many of the non-SDWIS States submitted datasets with more data elements than requested. In those cases, EPA determined which data elements corresponded to the SYR 4 ICR.

EPA also confirmed that all of the requested contaminants from the SYR 4 ICR were included in each State's 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 provided). 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 use an incorrect four-digit SDWIS code for a particular contaminant. EPA included issues regarding missing contaminants or undetermined CHEMIDs in the flagged record reports that were sent to each State to ask for clarification.

Sample collection dates were reviewed for consistency with the SYR 4 ICR timeframe (2012–2019). If sample collection dates were suspicious or incorrect, EPA tried to use other data elements to infer 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.

4.2 Restructuring Non-SDWIS State Data

Datasets received from the non-SDWIS States were restructured through a series of Microsoft Access queries into a format similar to the structure of the data from the SDWIS States to allow for the construction of a unified database for the SYR 4 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.

Several States submitted their data as a single flat file. However, the SYR 4 ICR database was designed as a relational database so the structure of that flat file had to be modified (i.e., mapped) into the structure of the relational database. The various data elements were mapped from the single flat file table into three separate inventory tables for water systems, facilities, and sample points (T6YWS, T6YWSF, and T6YSPT, respectively). As an example, a flat file from a State may have contained columns for PWSID, population served, and system type for every sample analytical result. However, in the final SYR 4 ICR database, the sample analytical result table (T6YSAR) stores the sample analysis results with a water system ID to link it to a single record in the water system table (T6YWS) with the corresponding inventory information. In this case, a unique list of water systems and their system-level information was created from the flat file and imported into T6YWS. The same procedure was followed with the sample point and facility information. In some cases, a State provided sample point information but not facility information. Within the SYR 4 ICR database, both the sample point and facility tables had to be fully populated. In these cases, facility IDs were set equal to sample point IDs.

For each non-SDWIS State, EPA compiled a list of all tables and data elements, including permitted values and a description of each element. An example of a permitted value is a recognized system type code such as “C” (community) or “NTNC” (non-transient non-community). From this framework, the submitted values were matched to the corresponding values within SDWIS/Fed for the federally reportable data elements. The remaining data elements and permitted values were mapped to the corresponding SDWIS/State values where possible. For example, the source water type column in a non-SDWIS State’s dataset could be called “PSource”; in this instance, EPA created a crosswalk table[3] indicating that “PSource” should be mapped to the SDWIS/Fed field “D_FED_PRIM_SRC_CD”. Generally, the States that did not use the extraction tool provided enough information in data dictionaries or other documentation for EPA to accurately organize the data in the SDWIS/Fed format.

[3] A “crosswalk table” shows equivalent data elements in more than one database schema (e.g., a non-SDWIS/State dataset format to the SDWIS/State dataset format). It maps the elements in one database to the equivalent elements in another database.

Prior to populating the SYR 4 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.

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 datasets in Oracle, a database storing all SYR 4 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 SYR 4 ICR database. All queries were created and saved in Access to document the transformation, ensuring that this process is reproducible.

4.3 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, "=" for detections), detection limits, and analytical results data in multiple fields. EPA added a "DETECT" field to the SYR 4 ICR dataset to identify the results sign and facilitate analysis. 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. Chemical monitoring data were received in both milligrams per liter (mg/L) and micrograms per liter (µg/L). For this analysis, EPA converted all data for inorganic contaminants (IOCs), synthetic organic contaminants (SOCs), volatile organic contaminants (VOCs), uranium, trihalomethanes (THMs), and haloacetic acids (HAAs) to µg/L. Data for alpha particles, beta particles,[4] and combined radium-226/228 were analyzed in picocuries per liter (pCi/L). Except for asbestos and radionuclides, all thresholds and concentrations in this report are expressed in µg/L. As described in Section 5.3.3, all records with missing or unusual units in the SYR 4 ICR dataset were sent back to States for input as part of the flagged records reports mentioned earlier.

[4] Although the MCL for beta particles is in the unit of measure of millirem per year (i.e., 4 mrem/yr), the primary unit of analytical measure is picocuries per liter (pCi/L). This unit of measure relates to screening thresholds of 15 pCi/L and 50 pCi/L that are defined in the 2000 Radionuclides Rule. More than 99 percent of all compliance monitoring data for beta particles submitted by the States to EPA were in units of pCi/L.

5 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 an extensive QA/QC evaluation on the data to be included in the SYR 4 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 prior to analysis. Not all QA/QC measures described were conducted on all States, as noted in this chapter.

5.1 Completeness and Representativeness of the Six-Year Review ICR Dataset

The final SYR 4 ICR dataset consists of compliance monitoring data and treatment technique information received from 59 of 66 States. It represents a large sample of PWSs across the United States and the largest compliance monitoring dataset ever compiled and analyzed under EPA's drinking water program. The 59 States that provided data for the SYR 4 ICR dataset comprise 88 percent of all PWSs and 92 percent of the total population served by PWSs nationally. The SYR 4 ICR dataset is geographically representative of PWSs nationwide.

The absence of data from seven States in the final SYR 4 ICR dataset could potentially bias the dataset's representation of the national occurrence of contaminants. However, the seven States, representing 12 percent of PWSs and 8 percent of the population served by PWSs nationally, are expected to have a relatively small influence when compared to the PWSs and populations represented by the States that did submit data. The seven States that did not provide compliance monitoring data or treatment technique information are Georgia, Michigan, Mississippi, New Mexico, Puerto Rico, Guam, and the U.S. Virgin Islands. Although Georgia and Mississippi, two sizeable States in the southeastern United States did not provide data, all other southeastern States did provide data, allowing for substantial regional coverage, especially from a population-based perspective. All other regions of the conterminous United States had at most one State not included in the dataset. The SYR 4 ICR dataset, with 59 of the 66 States represented, is therefore considered reasonably complete and nationally representative as the basis of the contaminant occurrence estimates for this Six-Year Review. However, to further address the issue of potential bias, EPA assessed the contaminants regulated by the Chemical Phase and Radionuclides Rules by comparing occurrence in the States that contributed data to the SYR ICR dataset to those that did not.

Because a complete compliance monitoring dataset for every PWS was not available to EPA, it was not possible to monitor national occurrence with complete certainty or to confirm that the SYR 4 ICR dataset is representative of the States that did not voluntarily contribute data. Therefore, an indicator of occurrence was developed using data available from the SDWIS/Fed database, which does not have complete compliance monitoring data but does include violation data from all 66 States. EPA compiled SDWIS/Fed records of MCL violations for the Chemical Phase and Radionuclides Rules only, used here as an indicator of contaminant occurrence, by

State for the same years as the SYR 4 ICR dataset (2012–2019).[5] The MCL violation records were used to determine if the violation rate in the 7 missing States was significantly different than the violation rate in the 59 States in the dataset, or if the violation rate in the 59 States could be considered representative (i.e., drawn from the same statistical population). EPA conducted this assessment for select chemical and radiological analytes evaluated under SYR 4.

The MCL violation rate for each contaminant (i.e., the percentage of systems with at least one MCL violation) was calculated for the 59 States in the dataset and separately for the 7 States not in the SYR 4 ICR dataset. For each contaminant, a Mann-Whitney U test, also known as a Wilcoxon rank-sum test, was used to determine whether the population of MCL violation rates by State significantly differs between the two groups (59 States versus 7 States). The non-parametric Mann-Whitney test was chosen, as opposed to a parametric t-test, because the small sample sizes (Exhibit 5) do not support an assumption that the data fit a normal distribution. The resulting p-values from the Mann-Whitney U test were first compared to an alpha (α) level of 0.05, a common threshold of significance, then to 0.1, a less-stringent threshold considered to account for small sample sizes. If the p-value resulting from the Mann-Whitney U test was less than 0.1, EPA rejected the null hypothesis that the two populations of MCL violation rates were equal and accepted the alternative hypothesis that they were unequal. Exhibit 5 summarizes the results of the Mann-Whitney U test analysis.

Of the 69 chemical and radiological contaminants evaluated, only 10 contaminants had at least one MCL violation listed in the SDWIS/Fed database for the 2012–2019 period for both groups (i.e., 59 States that submitted data to the SYR 4 ICR dataset versus the 7 States that did not). As States are only required to submit MCL violations to SDWIS/Fed but are not otherwise required to submit compliance monitoring data, only States with at least one violation in SDWIS/Fed for the specified contaminant were used in this analysis. Therefore, Mann-Whitney U tests were conducted on only these 10 contaminants (Exhibit 5). The resulting p-values were greater than 0.1 for 9 of the 10 contaminants: arsenic, combined radium, uranium, fluoride, gross-alpha (excluding radon and uranium), nitrate, nitrite, selenium, and thallium. Thus, EPA failed to reject the null hypothesis that the two populations of MCL violation rates are equal. For one contaminant (chromium), only one State in each group had an MCL violation, and so the Mann-Whitney U test could not be applied effectively.

[5] While the SDWIS/Fed database does not store complete compliance monitoring parametric records, the database does maintain the most current and complete national and state records of contaminant MCL violations. Annual MCL compliance data were extracted from SDWIS/Fed by EPA in November 2021.

Exhibit 5: Mann-Whitney U Test for MCL Violation Rates in States Included in SYR 4 versus States Not Included

Contaminant Name	Number of States with MCL Violations		Median of State-Level Violation Rates (percent)		p-Value
	States in SYR 4 ICR	States NOT in SYR 4 ICR	States in SYR 4 ICR	States NOT in SYR 4 ICR	
Uranium	26	2	6.68	32.91	0.259
Thallium	7	2	0.30	0.11	0.333
Radium-226/228 (combined)	35	4	5.98	4.01	0.460
Selenium	7	1	2.21	6.79	0.500
Arsenic	43	4	8.00	4.61	0.519
Nitrite	10	1	0.22	0.08	0.545
Fluoride	23	3	0.82	0.23	0.648
Nitrate	35	2	4.74	12.11	0.721
Alpha/photon emitters	29	3	1.79	4.53	0.903
Chromium	1	1	0.68	0.08	n/a[1]

[1] The Mann-Whitney test is not appropriate for this small sample size.

To further evaluate the completeness of each State's dataset, EPA used the SDWIS/Fed database as a reference and compared the number of PWSs by State in the SYR 4 ICR dataset to the number of systems by State in the SDWIS/Fed database (frozen fourth quarter 2019). Only the SDWIS/Fed database records from the 59 States that are also in the SYR 4 ICR dataset were included. Although the system inventories represented in the two data sources are similar, they are not equivalent. The main difference is that the SYR 4 ICR dataset counts reflect the total number of active water systems with compliance monitoring data during any of the eight years represented in the dataset (2012–2019), while the SDWIS/Fed 2019 fourth quarter data freeze counts reflect the total number of active water systems in a single year (2019). Since systems open, close, and consolidate over time, the number of systems in each State will understandably be somewhat different between the two data sources. Population changes in system service areas over time could also contribute to differences in population served numbers for systems between the two data sources. Exhibit 6 presents this comparison between the SDWIS/Fed and SYR 4 ICR datasets. If a system had more than one specified population served value in the submitted data, the most frequently occurring population served value was included in the SYR 4 ICR dataset.

Exhibit 6 compares the number of systems and population served by these systems in the December 2019 SDWIS/Fed freeze and the SYR 4 ICR dataset by State. The counts of systems and population served presented in for the SYR 4 ICR dataset only include systems that provided data for the requested regulated contaminants, including chemicals, radionuclides, microbes, and DBPs, prior to QA/QC review. The comparison between the counts of systems in the two data sources indicates a 9 percent difference between the number of systems listed in the December

2019 SDWIS/Fed freeze compared to the number of systems in the SYR 4 ICR dataset. In Exhibit 6, positive values for percent difference indicate that more systems are reported in the SYR 4 ICR dataset, while negative values indicate that more systems are reported in the 2019 SDWIS/Fed freeze. Comparing the number of systems for each State, the absolute percentage difference between SDWIS/Fed and the SYR 4 ICR dataset ranges from 0 percent (e.g., Region 1 tribes, Region 2 tribes, Region 4 tribes, Navajo Nation, Washington, D.C., Kentucky, and Hawaii) to 24 percent (e.g., Oklahoma) in the number of systems. Based on the population served by systems, the absolute percentage difference between the total population served by systems listed in SDWIS/Fed and that listed in the SYR 4 ICR dataset is less than 1 percent. Comparing population served values for individual States, the absolute percentage difference between SDWIS/Fed and the SYR 4 ICR dataset ranges from 0 percent (e.g., Region 2 tribes, Region 4 tribes, and Washington, D.C.,) to 30 percent (e.g., Utah).

Exhibit 6: Comparison of the Total Number of Systems and Population Served in SDWIS/Fed and the SYR 4 ICR Dataset, By State

State	Total Number of Systems[1,2]			Population Served		
	2019 SDWIS/Fed Freeze	SYR 4 ICR Dataset	Percent Difference[3]	2019 SDWIS/Fed Freeze	SYR 4 ICR Dataset	Percent Difference[3]
Alabama	579	592	2%	5,782,465	5,935,212	3%
Alaska	1,378	1,370	-1%	849,984	851,634	0.2%
American Samoa	111	100	-11%	59,379	58,476	-2%
Arizona	1,526	1,528	0.1%	6,739,728	6,777,613	1%
Arkansas	1,051	1,042	-1%	2,909,279	2,932,762	1%
California	7,498	8,394	11%	40,916,430	41,647,398	2%
Commonwealth of the Northern Mariana Islands	70	69	-1%	76,157	74,076	-3%
Connecticut	2,432	2,485	2%	2,877,830	2,882,881	0.2%
Colorado	2,048	2,500	18%	6,745,814	6,397,009	-5%
Delaware	482	521	7%	980,130	1,014,200	3%
Florida	5,241	5,962	12%	20,862,887	20,860,764	0.0%
Hawaii	136	136	0%	1,525,474	1,521,687	-0.2%
Idaho	2,007	1,976	-2%	1,495,882	1,516,508	1%
Illinois	5,353	6,181	13%	12,502,127	12,608,341	1%
Indiana	4,036	4,692	14%	5,512,342	5,658,801	3%
Iowa	1,817	1,982	8%	2,949,070	2,976,894	1%
Kansas	982	979	-0.3%	2,835,829	2,875,770	1%
Kentucky	433	433	0%	4,508,752	4,502,282	-0.1%
Louisiana	1,317	1,486	11%	5,074,387	5,320,364	5%
Maine	1,910	2,209	14%	931,352	968,213	4%

State	Total Number of Systems[1,2]			Population Served		
	2019 SDWIS/Fed Freeze	SYR 4 ICR Dataset	Percent Difference[3]	2019 SDWIS/Fed Freeze	SYR 4 ICR Dataset	Percent Difference[3]
Maryland	3,302	3,337	1%	5,867,239	5,861,767	-0.1%
Massachusetts	1,727	1,759	2%	9,811,383	9,788,373	-0.2%
Minnesota	6,703	6,628	-1%	5,037,593	5,027,228	-0.2%
Missouri	2,761	3,045	9%	5,622,969	5,660,127	1%
Montana	2,196	2,176	-1%	1,067,458	1,063,777	-0.3%
Navajo Nation	171	171	0%	176,792	176,750	0.0%
Nebraska	1,339	1,494	10%	1,660,734	1,681,763	1%
Nevada	601	594	-1%	2,891,787	2,899,400	0.3%
New Hampshire	2,513	2,747	9%	1,218,513	1,256,653	3%
New Jersey	3,625	4,180	13%	9,607,693	9,718,394	1%
New York	8,401	9,454	11%	21,265,451	18,006,468	-18%
North Carolina	5,366	5,946	10%	8,975,117	9,047,042	1%
North Dakota	400	502	20%	709,109	718,937	1%
Ohio	4,418	5,241	16%	10,916,586	11,149,543	2%
Oklahoma	1,386	1,822	24%	3,721,779	3,785,103	2%
Oregon	2,496	2,720	8%	3,748,090	3,784,217	1%
Pennsylvania	8,167	9,968	18%	12,670,902	12,931,009	2%
Region 1 tribes	5	5	0%	75,826	75,845	0.0%
Region 2 tribes	9	9	0%	12,565	12,565	0%
Region 4 tribes	30	30	0%	27,571	27,571	0%
Region 5 tribes	106	123	14%	136,541	149,532	9%
Region 6 tribes	87	92	5%	187,255	194,809	4%
Region 7 tribes	14	15	7%	15,926	15,506	-3%
Region 8 tribes	148	147	-1%	140,568	141,174	0.4%
Region 9 tribes	309	302	-2%	530,167	528,365	-0.3%
Region 10 tribes	134	139	4%	132,798	143,367	7%
Rhode Island	483	479	-1%	1,134,075	1,134,759	0.1%
South Carolina	1,410	1,169	-21%	4,081,703	4,078,161	-0.1%
South Dakota	651	749	13%	839,311	849,252	1%
Tennessee	783	921	15%	7,219,007	7,269,841	1%
Texas	7,040	6,955	-1%	28,945,548	29,290,499	1%
Utah	1,046	1,055	1%	3,327,756	4,721,824	30%
Vermont	1,403	1,539	9%	614,390	628,868	2%
Virginia	2,813	3,218	13%	7,510,864	7,835,414	4%

State	Total Number of Systems[1,2]			Population Served		
	2019 SDWIS/Fed Freeze	SYR 4 ICR Dataset	Percent Difference[3]	2019 SDWIS/Fed Freeze	SYR 4 ICR Dataset	Percent Difference[3]
Washington	4,457	4,386	-2%	8,029,486	8,184,593	2%
Washington, D.C.	6	6	0%	665,602	665,602	0%
West Virginia	857	831	-3%	1,597,832	1,599,584	0%
Wisconsin	11,325	12,835	12%	5,040,624	5,109,898	1%
Wyoming	778	764	-2%	589,509	588,998	-0.1%
Total	**129,873**	**142,190**	**9%**	**301,959,417**	**303,183,463**	**0.4%**

[1] The majority of the water systems with data in the SYR 4 ICR dataset are transient non-community water systems. Because only the nitrate/nitrite regulations require compliance monitoring by these transient systems (see Exhibit 7), data from the transient systems were included only for the nitrate and nitrite occurrence analyses and were excluded for all occurrence analyses for IOCs, SOCs, VOCs, and radiological contaminants.

[2] The data shown did not undergo QA procedures.

[3] The "percent difference" was calculated by subtracting the 2019 SDWIS/Fed Freeze total number of systems (or population served by systems) from the SYR 4 ICR dataset total number of systems (or population served by systems). That difference was then divided by the total number of systems (or population served by systems) from the SYR 4 ICR dataset. The percent difference is less than zero if the SYR 4 ICR dataset indicated a smaller number of systems (or population served by systems).

Exhibit 7 compares the number of systems and population served by these systems in the December 2019 SDWIS/Fed freeze and the SYR 4 ICR dataset stratified by source water type and system type. The total differences for all 59 States indicate 9 percent more systems and 0.4 percent greater population served is reported in the SYR 4 ICR dataset than in SDWIS/Fed. For community water systems (CWSs), the difference is 3 percent based on the number of systems and 1 percent based on the population served by systems. For non-transient non-community water systems (NTNCWSs), the difference is 8 percent based on the number of systems and 3 percent based on the population served by systems. For transient non-community water systems (TNCWSs), the difference is 10 percent based on the number of systems and 9 percent based on the population served by systems. Overall, these comparisons indicate that the SYR 4 ICR dataset is suitable for use as the basis of national contaminant occurrence estimates. As stated earlier in this report, the 59 States that provided data for the SYR 4 ICR dataset comprise 88 percent of all PWSs and 92 percent of the total population served by PWSs, representing a nationwide distribution of PWSs.

Exhibit 7: Comparison of the Total Number of Systems and Population Served in SDWIS/Fed and the SYR 4 ICR Dataset, By Source Water Type and System Type

Source Water Type	2019 SDWIS/Fed Freeze				SYR 4 ICR Dataset				
	CWS	NTNCWS	TNCWS	Total	CWS	NTNCWS	TNCWS	Unknown[1]	Total
Number of Systems									
Ground Water (GW)	33,613	14,905	67,564	116,082	35,528	16,181	75,027	745	127,481
Surface Water (SW)	10,807	755	2,172	13,734	10,145	701	2,240	135	13,221
Unknown	27	8	22	57	119	96	312	961	1,488
Total	44,447	15,668	69,758	129,873	45,792	16,978	77,579	1,841	142,190
Population Served									
Ground Water (GW)	81,806,757	4,631,058	8,663,270	95,101,085	107,516,099	4,954,238	9,600,777	49,520	122,120,634
Surface Water (SW)	202,988,465	1,363,942	2,486,544	206,838,951	179,187,202	1,211,353	533,646	4,474	180,936,675
Unknown	11,676	4,855	2,850	19,381	33,000	16,735	75,105	1,314	126,154
Total	284,806,898	5,999,855	11,152,664	301,959,417	286,736,301	6,182,326	10,209,528	55,308	303,183,463

[1] Systems with unknown system type (i.e., system type not reported by the State) were included in the fourth Six-Year Review analyses.

5.2 Quality Assurance Measures Applied to All Contaminants

Before analyzing contaminant occurrence, EPA performed a rigorous QA/QC evaluation of the data from each State. When necessary, EPA contacted States, sent detailed flagged records reports, and asked 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's 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 statewide waiver of the requirement to monitor for the contaminant(s). The information provided by each State was recorded.

Exhibit 8 lists the contaminant groups that each system type is required to monitor. All data that passed the QA/QC process from these systems were included in the SYR 4 occurrence analyses. Data from systems that were not required to sample for a given contaminant (e.g., SOC data from transient systems, radionuclide data from non-community systems) were excluded from the SYR 4 analyses.

Exhibit 8: 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)
Inorganic Contaminants (IOCs)	All non-purchased community water systems and non- transient non-community water systems are required to sample for IOCs.	All purchased systems and transient non-community water systems are not required to sample for IOCs.
Lead and Copper	All (non-purchased and purchased) community water systems and non-transient non-community water systems are required to sample for lead and copper.	Transient non-community water systems are not required to sample for lead and copper.
Nitrate and Nitrite	Non-purchased community water systems, non-transient non-community water systems, and transient non-community water systems are all required to sample for nitrate and nitrite.	All purchased systems are not required to sample for nitrate and nitrite.
Synthetic Organic Contaminants (SOCs)	All non-purchased community water systems and non- transient non-community water systems are required to sample for SOCs.	All purchased systems and transient non-community water systems are not required to sample for SOCs.
Volatile Organic Contaminants (VOCs)	All non-purchased community water systems and non- transient non-community water systems are required to sample for VOCs.	All purchased systems and transient non- community water systems are not required to sample for VOCs.
Radiological Contaminants	All non-purchased community water systems are required to sample for the radionuclides.	All purchased systems and non-purchased non- transient non-community and non-purchased transient non-community water systems are not required to sample for radionuclides.

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.

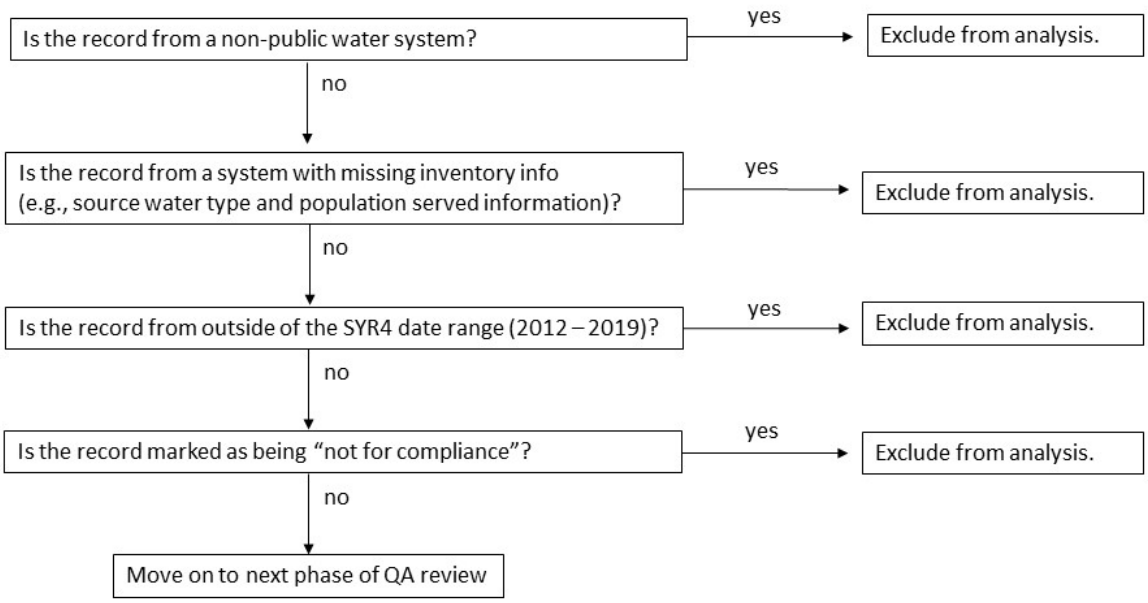
EPA created several automated data QA checks within the SYR 4 ICR dataset. These QA checks identified (i.e., flagged) records of potential data quality concerns. EPA sent out a detailed flagged record report to each State describing the identified records. These reports included the counts of flagged records by category, as well as specific questions for each category. In addition, an attachment identified the specific records that were flagged. EPA requested that each State provide the appropriate disposition (e.g., delete, make corrections) 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, 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.

Sections 5.2 through 5.5 provide a description of the various QA measures applied to the SYR 4 dataset 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. This code could be changed if EPA were to revise their decision about the exclusion of particular records from the occurrence analyses.

Section 5.2.1 through Section 5.2.5 describe the QA measures that were applied to the entire database (i.e., all regulated contaminant monitoring data). Exhibit 9 provides a visual representation of the overall flow of the QA/QC process for QA measures applied to all SYR 4 contaminants. Additional QA/QC measures applied to specified groups of contaminants are

included in Section 5.3 (chemicals and radionuclides), Section 5.4 (DBPs and related parameters), and Section 5.5 (microbes and residuals). Additional QA/QC measures were also taken to identify and exclude fluoride samples from fluoridated water systems prior to the occurrence analysis. See "Review of Fluoride Occurrence for the Fourth Six-Year Review" (USEPA, 2024c) for more information on additional QA/QC measures for fluoride data.

Exhibit 9: Flow Chart of QA Measures Applied to All SYR 4 Contaminants



5.2.1 Non-Public Water Systems

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

5.2.2 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, no population served). When inventory data were incomplete or missing, the missing data were populated 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. The inventory information for a given system may differ over time, so the SDWIS/Fed 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 occurrence analysis.

5.2.3 Sample Results Collected Outside of the Date Range

The SYR 4 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 occurrence analysis.

5.2.4 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 analyses, EPA assumed that all data from these three States were for compliance. All sample results flagged as "not for compliance" were excluded from the occurrence analysis.

5.2.5 Uniform System Inventory Information

For analysis, 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 as well as systems using ground water under direct influence of surface water were considered surface water systems to include in the occurrence analyses. This methodology to designate source may underestimate the number of groundwater 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.

5.3 Quality Assurance Measures Applied to Chemicals and Radionuclides

In addition to the QA measures described in Section 5.2, there were several other QA measures applied to only the chemical contaminants and radionuclides. Those QA measures are described in Sections 5.3.1 through 5.3.10. Additional QA measures are shown in Exhibit 10.

Exhibit 10. Flow Chart of Additional QA Measures Specific to Chemicals, Radionuclides, and Lead and Copper

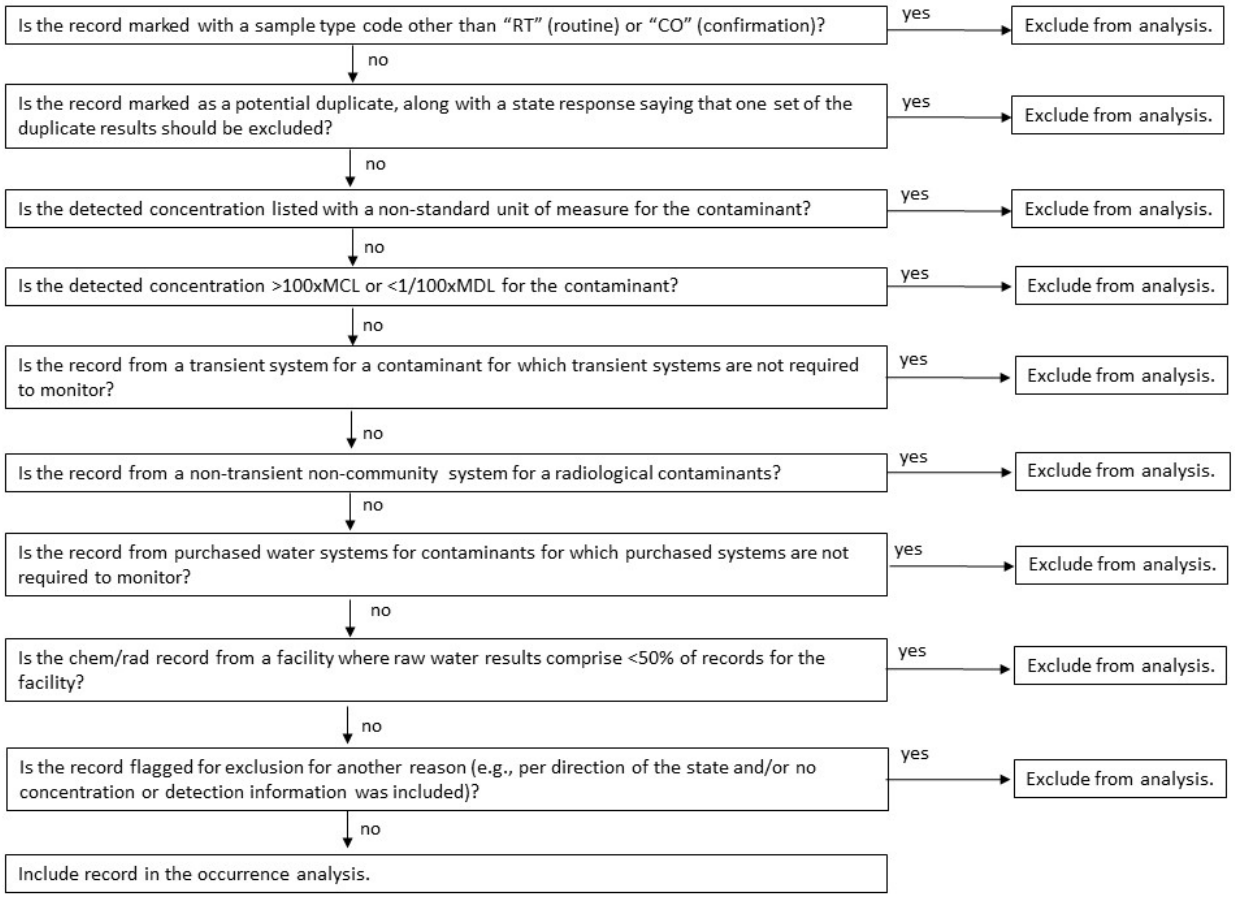


Exhibit 11 documents the specific counts of records included and excluded in each QA step. After applying the various QA measures to nearly 26 million SYR 4 ICR records for the Chemical Phase, Radionuclides, and Lead and Copper Rules' contaminants, 96 percent of the records remained in the final dataset. Most of the records were removed in either Step 9, removal of records from transient water systems for contaminants for which transient water systems are not required to sample or Step 11, removal of records from consecutive water systems, which are not required to sample for the Chemical Phase or Radionuclides Rules' contaminants.

Exhibit 11: Summary of the Count of Sample Analytical Results Removed via the QA Measures Applied to Chemical Phase, Radionuclides and Lead and Copper Rules' Contaminants

QA Step	Count of Records	
	Included	Excluded
Original number of analytical sample results[1]	25,756,988	
Step 1: Removal of analytical sample results from non-public water systems	25,752,276	4,712
Step 2: Removal of data from systems with missing source water type and/or population served information	25,712,838	39,438
Step 3: Removal of data with a sample collection date outside the SYR 4 date range of 2012 - 2019	25,637,677	75,161
Step 4: Removal of data marked as being "not for compliance"	25,567,220	70,457
Step 5: Removal of records marked with a sample type code other than routine or confirmation	25,455,914	111,306
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.	25,448,501	7,413
Step 7: Removal of data with detected concentrations with non-standard / blank unit of measure for the contaminant	25,448,171	330
Step 8: Removal of detected concentrations identified as potential high or low outliers	25,435,824	12,347
Step 9: Removal of records from transient water systems for contaminants for which transients are not required to sample	25,086,334	349,490
Step 10: Removal of records from non-transient water systems for radionuclides	25,070,331	16,003
Step 11: Removal of records from consecutive water systems	24,625,831	444,500
Step 12: Removal of raw water records where less than half the facility's records are raw	24,611,906	13,925
Step 13: Other flags (e.g., State responded that nitrate / nitrite records had been incorrectly entered, State included rows of data with no concentration value or detect / non-detect identifier)	24,596,843	15,063
Final number of records	24,596,843	
Percent Included	95%	

[1] The following 72 analytes are represented in the counts above: lead, copper, arsenic, barium, cadmium, chromium, cyanide, fluoride, mercury, nitrate-nitrite, nitrate, nitrite, selenium, antimony, total, beryllium, total, thallium, total, asbestos, endrin, bhc-gamma, methoxychlor, toxaphene, dalapon, diquat, endothall, glyphosate, di(2-ethylhexyl) adipate, oxamyl, simazine, di(2-ethylhexyl) phthalate, picloram, dinoseb, hexachlorocyclopentadiene, carbofuran, atrazine, alachlor lasso, 2,3,7,8-tcdd, heptachlor, heptachlor epoxide, 2,4-d, 2,4,5-tp, hexachlorobenzene, benzo(a)pyrene, pentachlorophenol, 1,2,4-trichlorobenzene, cis-1,2-dichloroethylene, total polychlorinated biphenyls (PCBs), 1,2-dibromo-3-chloropropane, ethylene dibromide, xylenes, total, chlordane, dichloromethane, o-dichlorobenzene, p-dichlorobenzene, vinyl chloride, 1,1-dichloroethylene, trans-1,2-dichloroethylene, 1,2-dichloroethane, 1,1,1-trichloroethane, carbon tetrachloride, 1,2-dichloropropane, trichloroethylene, 1,1,2-trichloroethane, tetrachloroethylene, chlorobenzene, benzene, toluene, ethylbenzene, styrene, gross alpha, excl. radon & uranium, combined uranium, combined radium (-226 & -228), and gross beta particle activity.

5.3.1 Non-Routine

Some States have regulations that are more stringent than the NPDWRs and require water systems to submit more sample results than federally required. States 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. All results that were marked as routine (RT) or confirmation (CO) were included in the occurrence analyses for the Chemical Phase Rules (i.e., contaminants evaluated in USEPA (2024a)); all other sample results for those contaminants were considered "non-routine" and were excluded from the occurrence analyses.

5.3.2 Duplicate Records

Potential duplicate sample analytical results for chemical contaminants and radionuclides 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 analysis unless the State responded to indicate that records were indeed duplicates and should be excluded.

5.3.3 Units of Measure

EPA identified all detection records for the Chemical Phase and Radionuclides Rules' contaminants where the units of measure reported were not one of the standard units used for the particular contaminant (i.e., not mg/L, µg/L, MFL (million fibers per liter), or pCi/L). For example, a benzene record with a unit of measure listed as NTU would be flagged since NTU is the unit of measure specifically for turbidity. EPA excluded all records in non-standard units from the occurrence analyses unless there was strong evidence of the correct standard unit (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).

5.3.4 Potential Outliers

To identify potential high outliers, EPA flagged all detected concentrations for the Chemical Phase and Radionuclides Rules' contaminants that were greater than 4 times the contaminant's MCL and all detected concentrations that were greater than 10 times the contaminant's MCL. All detected concentrations greater than 10 times the MCL were also included in the set of detected concentrations that were greater than 4 times the MCL. To identify potential low outliers, EPA flagged all detected concentrations that were less than one-tenth the minimum MDL. Exhibit 12 provides a list of all relevant MCL and MDL values for these contaminants.

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 explained 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 concentrations greater than 100 times the contaminant's MCL were excluded from the analysis, as were all detected concentrations less than one-hundredth the contaminant's minimum MDL. All other potential outliers less than or equal to 100 times the contaminant's MCL or greater than or equal to one-hundredth the contaminant's minimum MDL were included in the analysis. The values of 100 times the MCL and one-hundredth times the minimum MDL were chosen as conservative high-end and low-end cut-offs, respectively. For example, a benzene detected concentration of 1,600 ug/L was excluded as it was a likely data entry error. Likewise, a thallium record with a detected concentration of 0.00254 ug/L was excluded.

Exhibit 12: List of Contaminant MCL and MDL Values

Contaminant	Maximum Contaminant Level (MCL)		Method Detection Limit (MDL)	
	Value	Unit of Measure	Value	Unit of Measure
Inorganic Contaminants				
Antimony	6	µg/L	0.4	µg/L
Arsenic	10	µg/L	0.5	µg/L
Asbestos	7	MFL	--	MFL
Barium	2,000	µg/L	0.8	µg/L
Beryllium	4	µg/L	0.2	µg/L
Cadmium	5	µg/L	0.05	µg/L
Chromium (Total)	100	µg/L	0.08	µg/L
Copper	AL[1] = 1,300	µg/L	0.5	µg/L
Cyanide	200	µg/L	5	µg/L
Fluoride	4,000	µg/L	0.01	µg/L
Lead	AL[1] = 15	µg/L	0.6	µg/L
Mercury (Inorganic)	2	µg/L	0.2	µg/L
Nitrate (as N)	10,000	µg/L	0.002	µg/L
Nitrite (as N)	1,000	µg/L	0.004	µg/L
Selenium	50	µg/L	0.6	µg/L
Thallium	2	µg/L	0.3	µg/L
Synthetic Organic Contaminants				
Alachlor	2	µg/L	0.009	µg/L
Atrazine	3	µg/L	0.003	µg/L
Benzo(a)pyrene	0.2	µg/L	0.016	µg/L

Contaminant	Maximum Contaminant Level (MCL)		Method Detection Limit (MDL)	
	Value	Unit of Measure	Value	Unit of Measure
Carbofuran	40	µg/L	0.52	µg/L
Chlordane	2	µg/L	0.001	µg/L
Dalapon	200	µg/L	0.054	µg/L
Di(2-ethylhexyl)adipate (DEHA)	400	µg/L	0.09	µg/L
Di(2-ethylhexyl)phthalate (DEHP)	6	µg/L	0.46	µg/L
1,2-Dibromo-3-chloropropane (DBCP)	0.2	µg/L	0.009	µg/L
2,4-Dichlorophenoxyacetic acid	70	µg/L	0.055	µg/L
Dinoseb	7	µg/L	0.166	µg/L
Diquat	20	µg/L	0.72	µg/L
Endothall	100	µg/L	0.7	µg/L
Endrin	2	µg/L	0.002	µg/L
Ethylene Dibromide (EDB)	0.05	µg/L	0.008	µg/L
Glyphosate	700	µg/L	6	µg/L
Heptachlor	0.4	µg/L	0.0015	µg/L
Heptachlor Epoxide	0.2	µg/L	0.001	µg/L
Hexachlorobenzene	1	µg/L	0.001	µg/L
Hexachlorocyclopentadiene	50	µg/L	0.004	µg/L
Lindane (gamma-Hexachlorocyclohexane)	0.2	µg/L	0.003	µg/L
Methoxychlor	40	µg/L	0.003	µg/L
Oxamyl (Vydate)	200	µg/L	0.86	µg/L
Pentachlorophenol	1	µg/L	0.014	µg/L
Picloram	500	µg/L	0.05	µg/L
Polychlorinated biphenyls (PCBs)	0.5	µg/L	0.039	µg/L
Simazine	4	µg/L	0.008	µg/L
Toxaphene	3	µg/L	0.13	µg/L
2,3,7,8-TCDD (Dioxin)	0.00003	µg/L	0.0000044	µg/L
2,4,5-Trichlorophenoxypropionic Acid (Silvex)	50	µg/L	0.033	µg/L
Volatile Organic Contaminants				
Benzene	5	µg/L	0.1	µg/L
Carbon Tetrachloride	5	µg/L	0.002	µg/L
1,2-Dichlorobenzene	600	µg/L	0.02	µg/L
1,4-Dichlorobenzene	75	µg/L	0.01	µg/L
1,2-Dichloroethane	5	µg/L	0.02	µg/L

Contaminant	Maximum Contaminant Level (MCL)		Method Detection Limit (MDL)	
	Value	Unit of Measure	Value	Unit of Measure
1,1-Dichloroethylene	7	µg/L	0.05	µg/L
cis-1,2-Dichloroethylene	70	µg/L	0.02	µg/L
trans-1,2-Dichloroethylene	100	µg/L	0.03	µg/L
Dichloromethane	5	µg/L	0.02	µg/L
1,2-Dichloropropane	5	µg/L	0.01	µg/L
Ethylbenzene	700	µg/L	0.01	µg/L
Monochlorobenzene	100	µg/L	0.01	µg/L
Styrene	100	µg/L	0.01	µg/L
Tetrachloroethylene	5	µg/L	0.002	µg/L
Toluene	1,000	µg/L	0.01	µg/L
1,2,4-Trichlorobenzene	70	µg/L	0.02	µg/L
1,1,1-Trichloroethane	200	µg/L	0.005	µg/L
1,1,2-Trichloroethane	5	µg/L	0.01	µg/L
Trichloroethylene	5	µg/L	0.002	µg/L
Vinyl Chloride	2	µg/L	0.01	µg/L
Xylenes (Total)	10,000	µg/L	0.01	µg/L
Radiological Contaminants				
Alpha Particles	15	pCi/L	--	--
Beta Particles[2]	50	pCi/L	--	--
Combined Radium-226 & -228	5	pCi/L	--	--
Uranium	30	µg/L	--	--

[1] AL – Action Level

[2] The analyses presented here are based on compliance monitoring data represented in units of pCi/L and are conducted relative to the screening threshold of 50 pCi/L.

5.3.5 Transient Water Systems

Transient non-community water systems (TNCWS) operate for at least 60 days per year and serve at least 25 people per day. With regard to the Chemical Phase and Radionuclides Rules, transient water systems are only required to submit nitrate, nitrite, or total nitrate/nitrite sample results collected from entry points. Unless a State responded to say that the system in question used to be a CWS or NTNCWS at the time of sampling (and thus the records should be included), all data from transient water systems were excluded from the occurrence analyses presented in USEPA (2024a), except for nitrate, nitrite, or total nitrate/nitrite which TNCWS are required to monitor.

5.3.6 Non-Community Water Systems (Radionuclides Only)

Transient non-community water systems and non-transient non-community water systems are not required to submit radiological sample results. All data from non-community water systems were excluded from the occurrence analyses for the radionuclides.

5.3.7 Source Water Type Adjustment

As explained in Section 5.2.5, each system is defined with a single source water type and population-served category. For the Chemical Phase and Radionuclides Rules analyses, an adjustment to the source water type was necessary for a select group of systems whose water came from a mix of consecutive connections and their own sources. Specifically, these were systems that do not have their own surface intake or other SW facilities but do purchase some SW, in addition to using their own GW wells. In these cases, because the system does include some purchased surface water (SWP) sources, the federal source water type is listed as SWP in SDWIS/Fed and in the States' compliance monitoring data. This is the case even if the system only purchases a small portion of their water and the rest of the water comes from GW wells. To capture the legitimate (and required) compliance monitoring data from purchased systems (e.g., SWP, GWP) with their own GW wells, EPA reclassified the source water type of these systems prior to occurrence and preliminary exposure analyses. To identify purchased systems with their own GW wells, EPA reviewed all non-emergency, active facilities within a system. When active facilities with GW wells were identified, the system's source water type code was updated to "GW" in the SYR 4 ICR database. When all active, non-emergency facilities were classified as purchased sources according to SDWIS/Fed database (frozen fourth quarter 2019), the system was designated as a consecutive system (see Section 5.3.8).

5.3.8 Consecutive Water Systems

Consecutive water systems purchase 100 percent of their water from another water system(s). These systems do not have sources that require entry point monitoring for the Chemical Phase or Radionuclides Rules except for lead and copper. Analytical records from consecutive systems were excluded from the occurrence analyses for chemicals and radionuclides presented in USEPA (2024a) because this monitoring was not required for compliance. Population-served values and occurrence estimates in USEPA (2024a) were generated using the adjusted total populations served. Section 5.3.8 describes the process of identifying consecutive systems, and Section 6.2 discusses the adjustments of the population served to account for consecutive systems.

5.3.9 Samples from Source/Raw Water

EPA investigated source water samples (i.e., raw water samples) in some cases. In some States, systems are allowed to monitor raw water before treatment, rather than finished drinking water. If a contaminant is detected in a raw water sample at or above a level specified by the State, the system is required to collect a follow-up sample at the entry point to the distribution system, unless the water is not treated. EPA reviewed the raw (i.e., untreated, unfinished) samples related to the contaminants regulated under the Chemical Phase and Radionuclides Rules. EPA reviewed data at the facility-level (e.g., GW well, treatment plant) and excluded raw water records from the analysis if raw water records comprised less than 50 percent of the overall number of records

for the facility. EPA assumed that non-compliance source water samples had been incidentally included in the ICR reporting when they comprised less than half of the monitoring records for a given facility. When source water samples represented more than 50 percent of a facility's samples, EPA assumed that source water samples were intended for compliance.

5.3.10 Mismatched Nitrate and Nitrite Data

In some cases, data appeared to be mismatched for nitrate and nitrite. EPA reviewed data for instances where a nitrate and a nitrite result were reported as having an identical analytical result in the same water system on the same date and took corrective actions such as removing such data from the analysis or determining that the intent had been to report a single total nitrate plus nitrite result. EPA also evaluated cases where it was likely that nitrate and nitrite results were reversed and corrected them per State response when available.

5.4 Quality Assurance Measures Applied to DBPs and Related Parameters

In addition to the QA measures described in Section 5.2 that were applied to all contaminants, several additional contaminant-specific QA measures were applied to DBP data. For this reason, QA measures applied to DBP data will differ from those QA measures applied to chemical, radionuclide, and microbial contaminant data. The QA measures applied to DBPs and DBP-related parameters are described in this section. Exhibit 13 presents a flow chart of these additional QA measures for DBPs and DBP related parameters.

Exhibit 13. Flow Chart of Additional QA Measures Specific to DBPs and DBP-Related Parameters

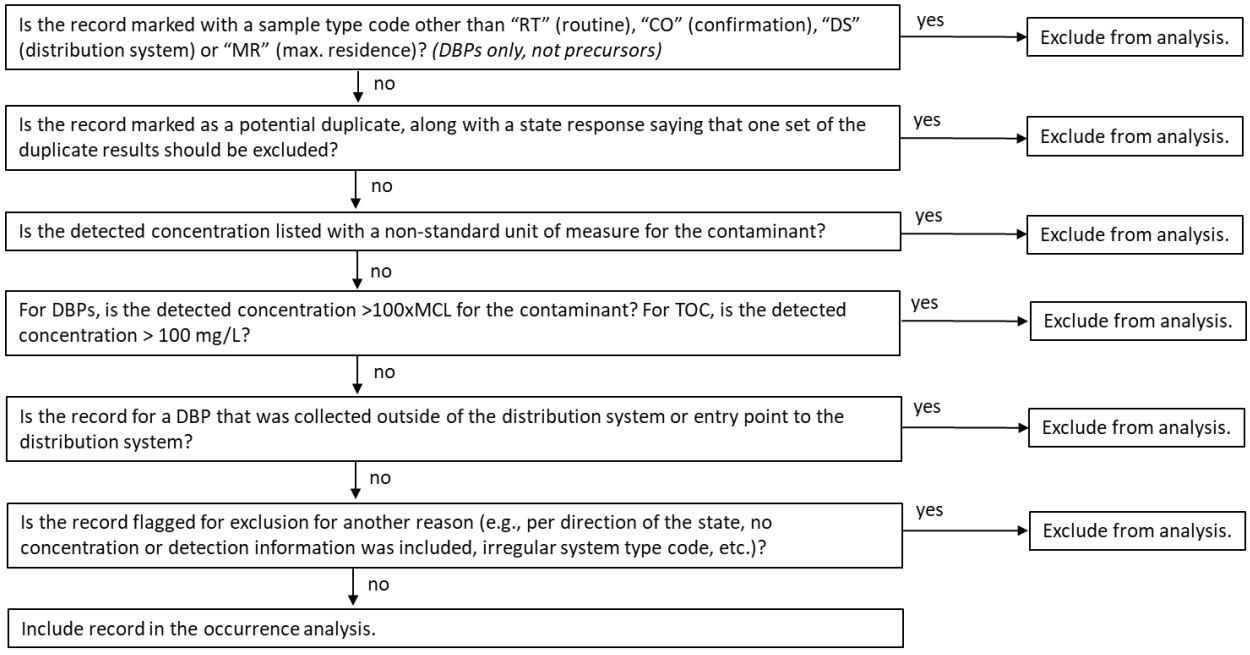


Exhibit 14 documents the specific counts of DBP records included and excluded in each QA

step. After applying the various QA measures to nearly 12 million SYR 4 ICR records for the DBPs and DBP related parameters, 96 percent of the records from 58 States remained in the final dataset. Exhibit 14 includes records for the following DBP contaminants: total trihalomethanes (TTHM), bromoform, chloroform, dibromochloromethane, bromodichloromethane, five haloacetic acids (HAA5), dibromoacetic acid, dichloroacetic acid, monobromoacetic acid, monochloroacetic acid, trichloroacetic acid, bromate, chlorite and DBP-related parameters: pH, alkalinity, and total organic carbon (TOC).

Exhibit 14: 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, monobromoacetic acid, monochloroacetic acid, trichloroacetic acid, bromate, chlorite, pH, alkalinity, and total organic carbon (TOC).

5.4.1 Non-Routine Samples

Some States have regulations that are more stringent than the NPDWRs and require water systems to submit more sample results than federally required. States 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 "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.

5.4.2 Duplicate Records

In the SYR 4 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 occurrence analyses.

5.4.3 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 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 (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).

5.4.4 Potential Outliers

To identify potential high outliers, EPA flagged all detected concentrations for the DBP-rule contaminants that were greater than 4 times the contaminant's MCL and all detected concentrations that were greater than 10 times the contaminant's MCL. All detected concentrations greater than 10 times the MCL were also included in the set of detected concentrations that were greater than 4 times the 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. Exhibit 15 provides a list of all relevant MCL values. For total organic carbon, which is not listed in Exhibit 15, all results greater than 100 mg/L were excluded from the 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 explained 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 analyses. 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 occurrence analysis. 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 15: List of DBP MCL Values

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

[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.

5.4.5 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).

5.5 Quality Assurance Measures Applied to Microbial Contaminants

In addition to the QA measures described above in Section 5.2, there were a handful of additional QA measures applied to only microbial contaminants. Those QA measures are described in this section. Exhibit 16 is a flow chart of the additional QA measures.

Exhibit 16. Flow Chart of Additional QA Measures Specific to Microbial Contaminants

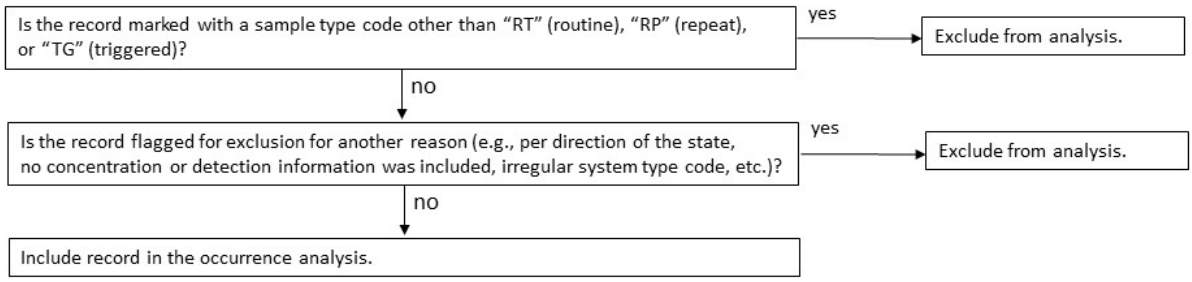


Exhibit 17 documents the specific counts of microbial records included and excluded in each QA step. After applying the various QA measures to more than 28 million SYR 4 ICR microbial records, 99 percent of the records from 57 States remained in the final dataset that was used for conducting occurrence analyses.

Exhibit 17: 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

QA Step	Count of Records	
	Included	Excluded
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
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.

5.5.1 Non-Routine Samples

Some States have regulations that are more stringent than the NPDWRs and require water systems to submit more sample results than federally required. States 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 analyses for the microbial contaminants.

5.5.2 Pairing Disinfectant Residual and Coliform Results for non-SDWIS States

Per the requirements under the Surface Water Treatment Rule (SWTR), surface water systems need to monitor disinfectant residuals at the same locations and time as for routine total coliform (TC) under the total coliform rule (TCR) and Revised TCR (RTCR). Thus, the TC data submitted by States generally also contain paired disinfectant residual monitoring records. However, some non-SDWIS States submit disinfectant residual concentration data as independent records not paired with TC samples. These data were submitted under different analyte codes: chlorine (0999), total chlorine (1000), chloramine (1006), chlorine dioxide (1008), residual chlorine (1012), and free residual chlorine (1013), depending on the State. To enable evaluation of disinfectant residual concentrations versus TC positivity rates, EPA paired the residual chlorine data with the associated TC result based on the sample collection date, sample point ID, and lab assigned ID. Specifically, EPA conducted this pairing for Wisconsin and Pennsylvania, two non-SDWIS States which submitted disinfected residual concentration data as independent records. Pennsylvania and Wisconsin were the only non-SDWIS States that had the necessary information needed to conduct this pairing. For Pennsylvania, 83,785 TC records (10 percent) were paired with free chlorine residuals (1013) and 54,395 TC (6 percent) were paired with total chlorine residuals (1000). For Wisconsin, 327,230 TC records (47 percent) were paired with free chlorine residuals (1013). 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 96,701 TC records in Pennsylvania and 335 TC records in Wisconsin to be paired records to be paired with free chlorine residuals (1013). An additional 32,824 TC records in Pennsylvania were paired with total chlorine residuals (1000). This resulted in a total of 267,705 TC records paired in Pennsylvania (31 percent) and 327,565 records paired in Wisconsin (47 percent). EPA did not have enough information to conduct pairing using the remaining analyte codes, including whether reported concentrations represent free or total chlorine. However, EPA is still making those unpaired disinfectant residual records available in the public release of the SYR 4 dataset (see Appendix E).

5.5.3 Updates to Absence and Presence Codes

Under the SYR 4 ICR, some microbial records (TC, EC, 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.

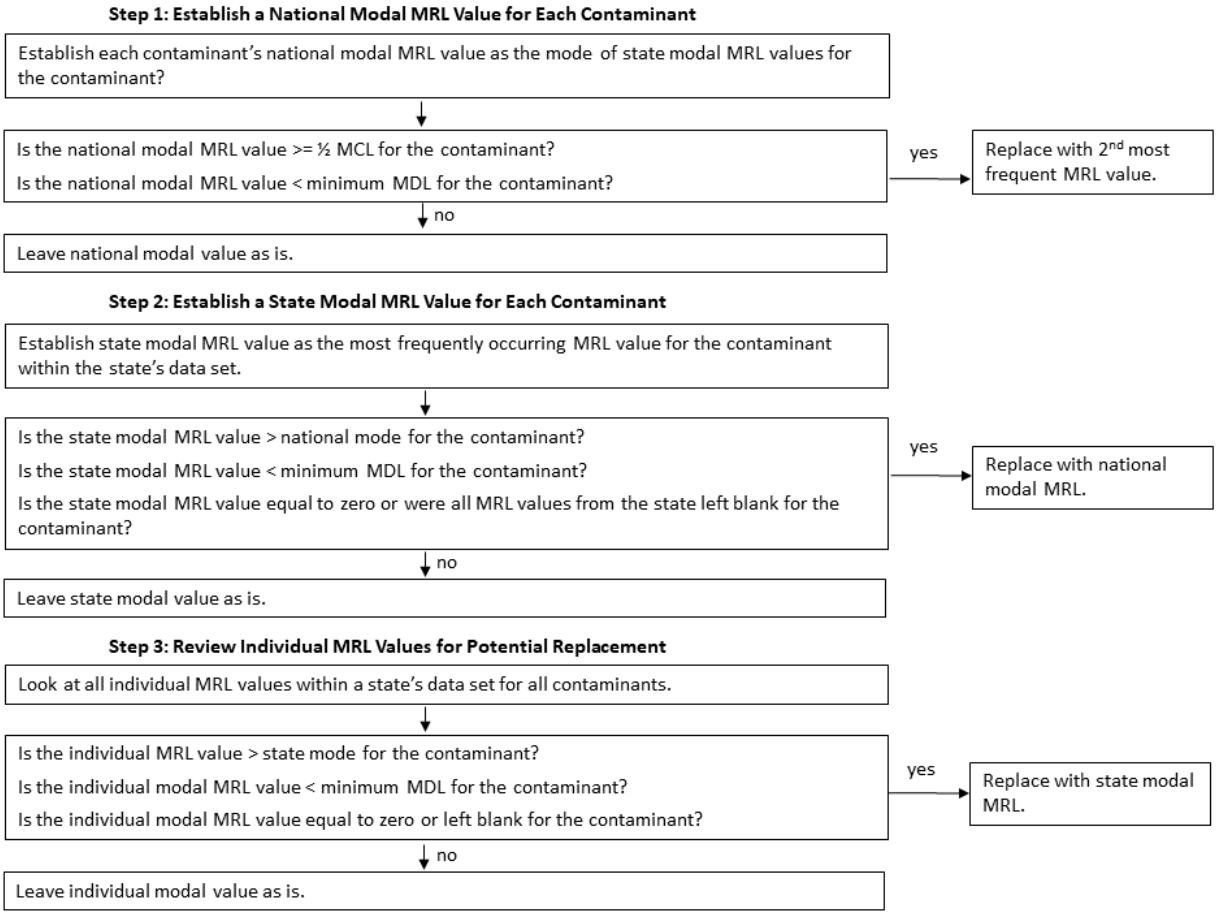
6 Data Preparation for Chemical Phase and Radionuclides Rules' Analyses

6.1 Non-Detection Record Replacement

Within the SYR 4 ICR dataset, each sample analytical result specifies a value and a sign to indicate whether that result is a detection (i.e., greater than or equal to the MRL) or a non-detection. Sample records reported as non-detections were less uniform and less complete than sample records for analytical detections. For some of the States that did report MRL data, this information was recorded in the analytical result field, along with a "<" sign in a corresponding field to identify the record as a non-detection. Other States simply included a zero or negative result in the analytical result field to signify a non-detection. For some of the occurrence analyses, EPA calculated system mean concentrations using a "simple substitution" approach that substitutes MRL values for reported analytical non-detections. Non-zero MRL numeric values were needed to replace all analytical results that were reported either as zero, "non-detection," "ND," etc.

A convention was established where EPA replaced any missing MRL data for non-detection results with the modal MRL value for the State in which the system was located. The State-specific modal MRLs were derived directly from the SYR 4 ICR dataset. In some cases, though, all MRL data for a specific contaminant's data from an entire State were missing. In these cases, the missing values were replaced with the national modal MRL derived as the mode of all the State-specific modal MRL values for that contaminant. If State-specific modal MRL values were greater than the national modal MRL or less than the minimum MDL for the contaminant, a process was developed to identify and replace such values with more reasonable MRL values. Exhibit 18 provides a description of the three-step process.

Exhibit 18. Process to Establish Contaminant National Modal MRLs



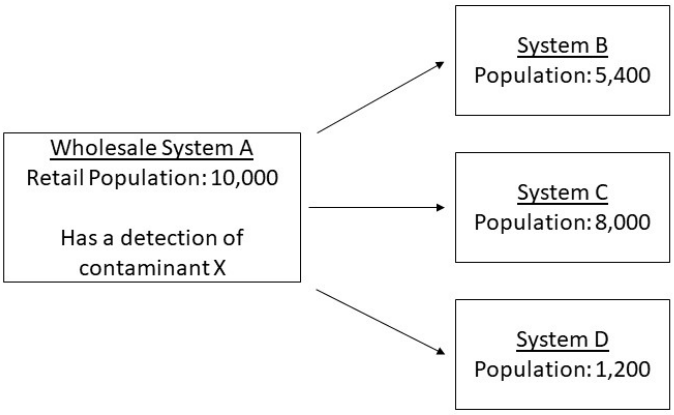
6.2 Adjustments of Population Served by Public Water Systems

Consecutive water systems purchase all of their water from other systems (i.e., seller or wholesale systems). Compliance monitoring requirements are different for consecutive water systems compared to other systems because their water has already been treated and monitored by the wholesale water system. For the occurrence analyses of the Chemical Phase and Radionuclides Rules' contaminants presented in USEPA (2024a), EPA excluded data from consecutive systems, as those systems are not required to sample for those contaminants.[6] However, EPA did adjust the population values of the wholesale systems to include the population of consecutive systems that buy their water. The population served directly by these wholesale systems is the retail population, and the population served indirectly through the purchased systems is the wholesale population. The sum of the retail and wholesale populations is the adjusted total population. Adjusting for the total population served ensured that the entire relevant population was included in the exposure estimates.

[6] Note that consecutive water systems do their own sampling for lead and copper, as well as the microbial contaminants and DBPs; thus, the data from these systems were not excluded from the lead, copper, microbial, or DBP occurrence datasets (see USEPA, 2024a and USEPA, 2024b).

Exhibit 19 illustrates a simple example of these adjustments. In the diagram, Systems B, C, and D (consecutive systems) buy 100 percent of their water from System A (wholesale system). System A is required to monitor for contaminant X; however, Systems B, C, and D are not required to monitor. If contaminant X was detected and population values were not adjusted, the exposure estimates would not account for the populations served by Systems B, C, and D, even though these populations could be exposed to contaminant X. To correct for this, EPA uses the adjusted total population served (i.e., retail plus wholesale populations) for System A for all population-served estimates, which is equal to 24,600 people.

Exhibit 19: Illustration of the Adjusted Total Population Served by Wholesale Systems

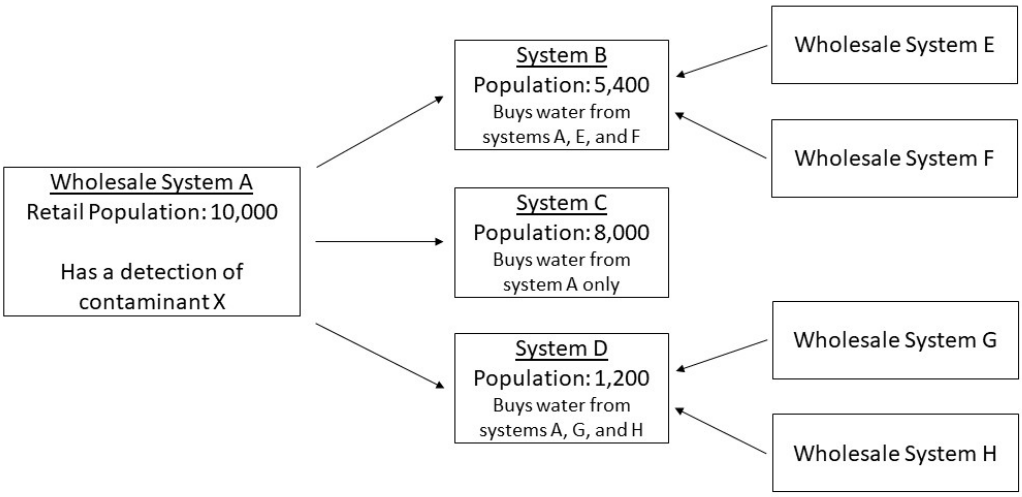


Total population served by wholesale system A exposed to detection of contaminant x
= retail population + wholesale population
= 10,000 + (5,400 + 8,000 + 1,200)
= 24,600

For some systems, a slightly more complicated adjustment to the wholesalers' total population served values was required. Many consecutive water systems buy water from more than one wholesale system. Because of this, their entire population should not be attributed to a single wholesale system, and EPA must instead distribute the population across the wholesale systems. The actual relative quantities of water purchased from the different wholesalers are not available; therefore, in the cases of multiple wholesalers, the population served by the consecutive system was assumed to be uniformly distributed across the wholesalers.

Exhibit 20 illustrates the complete population adjustment for System A, including the uniform distribution of the consecutive systems' population served. In the diagram, for example, System B, a system serving a population of 5,400 purchases its water from three different wholesale systems – Systems A, E, and F. To account for the population served by System B in the population exposure estimates, a third of System B's population (5,400 ÷ 3 =1,800) is uniformly distributed across Systems A, E, and F.

Exhibit 20: Illustration of the Allotment of Consecutive System Populations to Wholesale Systems



Adjusted population served by wholesale system A exposed to detection of contaminant x
= retail population + wholesale population
= 10,000 + (5,400/3 + 8,000 + 1,200/3)
= 20,200

To make adjustments across the SYR 4 ICR dataset, EPA compiled a list of all wholesale and consecutive systems. This list of buyer-wholesaler relationships was from SDWIS/Fed, fourth quarter of 2019. EPA then created a crosswalk linking the consecutive systems to the wholesale systems from which they purchased their water. Finally, EPA distributed the population served by each consecutive system evenly across the relevant wholesale system populations, according to the calculations described. As a result, the contaminant occurrence measures are associated with the adjusted total population (i.e., retail plus wholesale) served by these wholesale systems included in the Six-Year Review dataset.

7 Public Access to SYR 4 ICR Data

Through extensive data management efforts and QA evaluations, including consultations with state data management staffs, EPA established a compliance monitoring and treatment technique dataset (SYR 4 ICR dataset) that consists of data from 59 States (46 states of the United States, Washington, D.C., American Samoa, Navajo Nation, Commonwealth of the Northern Mariana Islands, and other tribes). The initial SYR 4 ICR dataset included more than 83 million analytical records from approximately 142,000 PWSs that serve approximately 303 million people nationally.[7] More than 73 million analytical contaminant records underwent QA/QC review to be included in the SYR 4 ICR dataset to support the SYR 4 analyses in USEPA (2024a-d). After the QA/QC review was completed on these analytical records and a small percentage of records that did not meet quality standards were omitted from analyses, the final SYR 4 ICR dataset comprise almost 71 million analytical records from approximately 140,000 PWSs that serve approximately 301 million people nationally.[8]

EPA maintains the final SYR 4 ICR compliance monitoring data and treatment technique information online at https://www.epa.gov/dwsixyearreview. The public can download the final SYR 4 ICR data (i.e., all records that passed the QA/QC review) that were used in support of the evaluation of regulated contaminant levels in drinking water. Appendix E includes a user guide to obtaining and using the SYR 4 ICR compliance monitoring, treatment technique, and related data from EPA's website.

[7] This count of 142,000 PWSs represents all water systems with any SYR 4 data, including data for information not specifically requested.
[8] This count of 140,000 PWSs serving 301 million people represents water systems that provided data for requested contaminants that passed QA/QC review.

8 References

United States Environmental Agency (USEPA). 2016. Six-Year Review 3 Technical Support Document for Disinfectants/Disinfection Byproducts Rules. EPA-810-R-16-012. December 2016.

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.

USEPA. 2024a. Analysis of Regulated Contaminant Occurrence Data from Public Water Systems in Support of the Fourth Six-Year Review of National Primary Drinking Water Regulations: Chemical Phase Rules and Radionuclides Rules. EPA-815-R-24-014. February 2024.

USEPA. 2024b. Six-Year Review 4 Technical Support Document for Microbial Contaminant Regulations. EPA-815-R-24-022. February 2024.

USEPA. 2024c. Review of Fluoride Occurrence for the Fourth Six-Year Review. EPA-815-R-24-021. February 2024.

USEPA. 2024d. Analytical Feasibility Support Document for the Fourth Six-Year Review of National Primary Drinking Water Regulation. EPA-815-R-24-015. February 2024.

Data Management and Quality Assurance/Quality Control Process for the Fourth Six-Year Review Information Collection Request Dataset: Appendices

9 List of Appendices

Appendix A: Data Request Letter that EPA Sent on June 3, 2020 to Each Primacy Agency to Request Voluntary Submission of 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

Exhibit 1: Occurrence Data Requested		
Copper	Glyphosate	Toxaphene
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.

Exhibit 2: Requested Data Categories	
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/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.)*

Exhibit 2: Requested Data Categories	
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.
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).

1. 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.

2. 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.

3. 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 datasets will be aggregated into national contaminant occurrence datasets for each contaminant. The national aggregate datasets 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: Crosswalk of Data Elements Requested for SYR 4 ICR and the SDWIS Data Element Names

Exhibit B.1 provides a crosswalk of the data elements requested in the SYR 4 ICR letter to the States compared with the actual data elements as they appear in the SDWIS/State databases. These were the data elements extracted via the SDWIS/State Extraction Tool.

Exhibit B.1: Crosswalk Table of Data Elements in SYR 4 ICR Request and SDWIS

Data Category	SDWIS Mapping ([Table Name].[Data Element])
System-Specific Information	
Public Water System Identification Number (PWSID)	TINWYS.NUMBER0
System Name	TINWSYS.NAME
Federal Public Water System Type Code	TINWSYS.D_PWS_FED_TYPE_CD
Population Served	TINWSYS.D_POPULATION_CNT
Federal Source Water Type	TINWSYS.D_FED_PRIM_SRC_CD
Treatment Information	
Water System Facility	T6YWSF; [TINWSF_IS_NUMBER] and [TINWSF_ST_CODE]
Filtration Type	TINWSYS.D_SWGUDI_INT_CD; TINTRPLT.FILTER_TYPE
Treatment Technique Information	TINTROBJ.NAME; TINTRPRO.NAME; TINTRPLT.DBM_VIR_INACT_LOG?; TINTRPLT.DBM_VIR_INACT_DT?; TINTRPLT.DBM_VIR_INACT_STAT?; TINTRPLT.DBM_VIR_INACT_PCT?; TSAOSAM.NAME; TSOSAM.VALUE_NUMBER; TSOSAM.UOM_CODE
Filter Backwash Information	TINTRPLT.FBR_SCHEMATIC_STAT; TINTRPLT.FBR_SCHEMA_RCV_DAT; TINTRPLT.FBR_SCHEMA_RVW_DAT; TINTRPLT.FBR_ALTR_RTN_RQS; TINTRPLT.FBR_ALTR_RTN_DT; TINTRPLT.FBR_CORCTV_ACT_RQS; TINTRPLT.FBR_CORCTV_ACT_DT
Sample-Specific Information	
Sampling Point Identification Code	TSASMPPT.IDENTIFICATION_CD
Sample Identification Number	TSASAMPL.ST_ASGN_IDENT_NUM
Sample Collection Date	TSASAMPL.COLLECTION_END_DATE
Sample Type	TSASAMPL.TYPE_CODE
Sample Analysis Type Code	TSASAMPL.REPEAT_LOC_TYP_CD
Contaminant	TSAANLYT.CAS_REGISTRY_NUM (TSAANLYT.CODE)
Sample Analytical Result- Sign	TSASAR.LESS_THAN_IND (TSAANLYT.LESS_THAN_CODE)
Sample Analytical Result- Value	TSASAR.CONCENTRATION_MSR
Sample Analytical Result- Unit of Measure	TSASAR.UOM_CODE
Sample Analytical Method Number	TSASMN.CODE
Minimum Reporting Level (MRL) - Value	TMNALRA.MEASURE (TSASAR.DETCTN_LIMIT_NUM, TSASAR.DETECTN_LIM_UOM_CD)
MRL - Unit of Measure	TMNALRA.UOM_CODE (TSASAR.UOM_CODE)
Source Water Monitoring Information	TMNFANL.* (TMNMPAVG.PRC_ACH_RMVL_RA_NO,TMNMPAVG.PRC_ACH_RMVL_RA_TX)
Sample Summary Reports	TSASMPSM.* (TSAMDBPS.)

Appendix C: Data Dictionary for the SYR 4 ICR Database

This appendix contains 19 tables presenting the various tables and their data elements in the SYR 4 relational database, along with all permitted values in those tables. The data dictionary for ADWR compliance data is in Appendix E, Section 6.

Exhibit C.1: Description of T6YWS (water system table)

Field Name	Data Type	Description
T6YWS_ID	Number	Unique identifier for each water system record.
TINWSYS_IS_NUMBER	Number	Identifier for each water system that is unique when combined with TINWSYS_ST_CODE.
TINWSYS_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the system.
NUMBER0	Text	Public water system identification number (PWSID)
WS_NAME	Text	Water system name
D_POPULATION_COUNT	Number	Retail population served by the water system.
D_FED_PRIM_SRC_CD	Text	Updated primary water source for the water system. (Updated for systems that were listed as purchased but are not truly 100% purchased.) GU = Ground water Under Direct Influence of Surface Water GUP = Purchased Ground Water Under Direct Influence of Surface Water GW = Ground Water GWP = Purchased Ground Water SW = Surface Water SWP = Purchased Surface Water
D_PWS_FED_TYPE_CD	Text	Water system type according to federal requirements. C = Community water system NC = Non-community water system NTNC = Non-transient non-community water system NP = Non-public water system (This field has been corrected as a part of the QA/QC process)
WS_ACTIVITY_STATUS_CD	Text	Activity status of the water system. A = Active (i.e., water system that is producing water on a regular basis (obtaining, treating, pumping, storing, or distributing)); I = Inactive
WS_ACTIVITY_DATE	Date	For SDWIS States, the ACTIVITY_DATE is the date of the ACTIVITY_STATUS_CD. For non-SDWIS States, it's the date that the water system was deactivated (if applicable).
STATE_CODE	Text	Two-letter code that identifies the U.S. state in which the system is located. This differs from TINWSYS_ST_CODE for tribal systems.
WHOLESALE_POPULATION	Number	Wholesale population served (for seller systems only)
TOTAL_POPULATION	Number	Total retail plus wholesale population served (for seller systems only)
ADJUSTED_TOTAL_POPULATION	Number	Adjusted total population served (retail plus adjusted wholesale population served as not to double-count buyer systems that purchase from multiple seller systems). For non-seller systems, this value is equal to D_POPULATION_COUNT.
ORIGINAL_ D_FED_PRIM_SRC_CD	Text	Original primary water source for the water system. GU = Ground water Under Direct Influence of Surface Water GUP = Purchased Ground Water Under Direct Influence of Surface Water GW = Ground Water GWP = Purchased Ground Water SW = Surface Water SWP = Purchased Surface Water

Exhibit C.2: Description of T6YWSF (water system facility table)

Field Name	Data Type	Description
T6YWSF_ID	Number	Unique identifier for each water system facility record.
T6YWS_ID	Number	Identifier matching each record to T6YWS
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that is unique when combined with TINWSF_ST_CODE.
TINWSF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.
TINWSYS_IS_NUMBER	Number	Identifier for each water system that is unique when combined with TINWSYS_ST_CODE.
TINWSYS_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the system.
WSF_ACTIVITY_STATUS_CD	Text	Activity status of the water system facility. A = Active; I = Inactive
WSF_ACTIVITY_DATE	Date/Time	For SDWIS States, the ACTIVITY_DATE is the date of the ACTIVITY_STATUS_CD. For non-SDWIS States, it's the date that the water system facility was deactivated (if applicable).
ST_ASGN_IDENT_CD	Text	A State-assigned value which identifies the water system facility.
WSF_NAME	Text	Name of the water system facility.
WSF_TYPE_CODE	Text	Type of the water system facility (permitted values). CC = Consecutive Connection; CH = Common Headers; CS = Cistern; CW = Clear Well; DS = Distribution System/Zone; IG = Infiltration Gallery; IN = Intake; NN = Non-piped, non-purchased; NP = Non-piped; OT = Other; PC = Pressure Control; PF = Pump Facility; RC = Roof Catchment; RS = Reservoir; SI = Surface Impoundment; SP = Spring; SS = Sampling Station; ST = Storage; TM = Transmission Main (Manifold); TP = Treatment Plant; WH = Well Head; WL = Well
FILTRATION_STATUS	Text	Indicates whether a non-emergency surface water source or a non-emergency ground water under the influence of surface water source is required to install filtration by a certain date or is successfully avoiding filtration.
FILTRATION_STAT_DT	Date/Time	Date the Filtration Status was determined.

Exhibit C.3: Description of T6YSPT (sample point table)

Field Name	Data Type	Description
T6YSPT_ID	Number	Unique identifier for each sample point record.
T6YWSF_ID	Number	Identifier that relates each record to the unique record in the T6YWSF table.
T6YWS_ID	Number	Identifier that relates each record to the unique record in the T6YWS table.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that is unique when combined with TINWSF_ST_CODE.
TINWSF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.
TSASMPPT_IS_NUMBER	Number	Identifier for each sample point that is unique when combined with TSASMPPT_ST_CODE.
TSASMPPT_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the sample point.
TSASMPPT_TYPE_CODE	Text	Location type of a sampling point (permitted values). DN = Within 5 service connections Downstream; DS = Distribution System; EP = Entry point; NF = Near the first service connection; OR = Original location; SR = Source sampling point; UP = Within 5 service connections Upstream
SOURCE_TYPE_CODE	Text	The type of water source, based on whether treatment has taken place. FN = Finished, treated; RW = Raw, untreated; x = unknown
IDENTIFICATION_CD	Text	Unique code for identifying a water system facility's sample point. This value must be unique within the Water System Facility.
DESCRIPTION_TEXT	Text	Description of the sample point location.
LD_CP_TIER_LEV_TXT	Text	Indicates if the sample point is a Lead and Copper Tier 1, 2, or 3 site.

Exhibit C.4: Description of T6YANALYTE (analyte table)

Field Name	Data Type	Description
T6YANALYTE_ID	Number	Unique identifier for each analyte record.
TSAANLYT_IS_NUMBER	Number	Identifier for each analyte that is unique when combined with TSAANLYT_ST_CODE.
TSAANLYT_ST_CODE	Text	This value is “HQ” for all SDWIS/Fed contaminants. If the value is not “HQ,” the analyte code is specific to the primacy agency.
ANALYTE_CODE	Text	4-digit EPA Analyte code
ANALYTE_NAME	Text	Analyte name
ALTERNATE_NAME	Text	Synonym for analyte name
FIRSTIMPORTSTATE	Text	First State from which the analyte was added (if a non-requested contaminant from a non-SDWIS State).

Exhibit C.5: Description of T6YSAR (sample analytical result table)

Field Name	Data Type	Description
T6YSAR_ID	Number	Unique identifier for each sample analytical result record.
T6YWS_ID	Number	Identifier that relates each record to the unique record in the T6YWS table.
T6YWSF_ID	Number	Identifier that relates each record to the unique record in the T6YWSF table.
T6YSPT_ID	Number	Identifier that relates each record to the unique record in the T6YSPT table.
T6YANALYTE_ID	Number	Identifier that relates each record to the unique record in the T6YANALYTE table.
TSASAR_IS_NUMBER	Number	Identifier for each sample analytical result that is unique when combined with TSASAR_ST_CODE.
TSASAR_ST_CODE	Text	Two-digit code that identifies the State that submitted data.
TSASAMPL_IS_NUMBER	Number	Identifier for each sample that must be combined with TSASAMPL_ ST_CODE when used. These values may not be unique.
TSASAMPL_ST_CODE	Text	Two-digit code that identifies the State that submitted data.
TSASMN_IS_NUMBER	Number	Identifier for each standard method number that must be combined with TSASMN_ST_CODE when used. These values may not be unique.
TSASMN_ST_CODE	Text	Two-digit code that identifies the State that submitted data.
TSASAMPL0IS_NUMBER	Number	Identifier for each sample that must be combined with TSASAMPL0ST_CODE when used. These values may not be unique. This relates a confirmation or repeat sample to the originating routine sample.
TSASAMPL0ST_CODE	Text	Two-digit code that identifies the State that submitted data.
LAB_ASGND_ID_NUM	Text	An identifier used for reconciliation with the State data system or sample identification number assigned by the laboratory.
COLLLECTION_END_DT	Date/Time	Sample Collection Date.
COMPL_PURP_IND_CD	Text	Indicates whether or not the sample result is used for compliance determination. Y = "yes" (use for compliance determination) N = "no" (taken for reasons other than compliance determination such as lab performance, etc.)
TSASAMPL_TYPE_CODE	Text	Sample Type Code (permitted values): BB = Batch Blank; CN = Continuous; CO = Confirmation; DU = Duplicate; FB = Field Blank; GR = Grab; MR = Maximum Residence Time; MS = Matrix spike; PE = Performance Evaluation; RI = Replacement for Invalid; RL = Replacement; RP = Repeat; RT = Routine; SB = Shipping Blank; SL (or ST) = Split; SP =Special; TE = Technical Evaluation; TG = Triggered
REPEAT_LOC_TYP_CD	Text	The location of the repeat/check/confirmation sample with respect to the location of the original routine sample.

Field Name	Data Type	Description
LESS_THAN_IND	Text	Indication of whether the result is "less than" the Lab Reporting Limit or "less than" the Regulatory Minimum Reporting Limit. "Y" = "yes" result is less than (i.e., a non-detection) "N" = "no" result is not less than (i.e., a detection)
LESS_THAN_CODE	Text	When valued, indicates that the analytical result (concentration) was below the Regulatory Minimum Reporting Level or below the Laboratory Reporting Level. DL = Detection Limit; MDL = The lab reported the analytical result was less than the Method Detection Limit; MRL = The lab reported the analytical result was less than the Minimum Reporting Level.
DETECTN_LIMIT_NUM	Number	Limit established by the laboratory below which scientifically reliable results cannot be achieved.
DETECTN_LIM_UOM_CD	Text	Unit of measure associated with the detection limit.
REPORTED_MSR	Text	Value (in text form) that represents the result obtained from a sample analysis. This field maintains the level of precision of the result (i.e., maintains the correct number of trailing zeroes in the analysis result).
CONCENTRATION_MSR	Number	A numeric value that represents the result obtained from a sample analysis.
UOM_CODE	Text	Unit of measure.
PRESENCE_IND_CODE	Text	Indicates whether results of an analysis were positive (P-Presence) or negative (A-Absence). Indication of presence or absence creates an analytical result for a microbial analyte.
COUNT_QTY	Number	The number of organisms counted or estimated in a microbiological sample. Usually expressed as "# of colonies per 100 milliliter sample."
COUNT_TYPE	Text	Type of microbiological unit that is being counted per specified count unit. Count type varies with the microbiological organism where count has been recorded.
COUNT_UOM_CODE	Text	The units of measure associated with the microbial analytical result count.
FF_CHLOR_RES_MSR	Number	Amount of free chlorine residual disinfectant found in the water after disinfection has been applied.
FLDTOT_CHL_RES_MSR	Number	Amount of total chlorine residual disinfectant found in the water after disinfection has been applied.
FIELD_TEMP_MSR	Number	Temperature of the water being sampled at the time and place of sample collection.
TEMP_MEAS_TYPE_CD	Text	Enables selection of "C" for centigrade or "F" for Fahrenheit degrees.
FIELD_TURBID_MSR	Number	Turbidity of the water being sampled at the time and place of sample collection in Nephelometric Turbidity Units (NTU).
FIELD_PH_MEASURE	Number	pH of the water being sampled at the time and place of sample collection (pH units).
FIELD_FLOW_RATE	Number	Flow of the water being sampled at the time and place of sample collection.
METHOD_CODE	Text	Method used to analyze the sample.
METHOD_NAME	Text	Name of method used to analyze the sample.
DETECT	Number	DETECT = 1 for all detections. Detections were identified as records with [CONCENTRATION_MSR] > 0 and [LESS_THAN_IND] was <> to "Y" or was null. DETECT = 0 for all non-detections. Non-detections were identified as records with [CONCENTRATION_MSR] = 0 and/or [LESS_THAN_IND] = "Y."
VALUE	Number	For all non-detections (i.e., [DETECT] = 0), [VALUE] was left blank. For all detections (i.e., [DETECT] = 1), [VALUE] = [CONCENTRATION_MSR].
UNITS	Text	Unit of measure associated with [VALUE]
TSASMPPT_IS_NUMBER	Number	Identifier for each sample point that is unique when combined with TSASMPPT_ST_CODE.
TSASMPPT_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the sample point.
ASSAY_UOM_CODE	Text	Unit of measure for microbiological analytical result

Exhibit C.6: Description of T6YDBPSUM (DBP summary table)

Field Name	Data Type	Description
T6YDBPSUM_ID	Number	Unique identifier for each DBP summary record.
T6YWS_ID	Number	Identifier that relates each record to the unique record in the T6YWS table.
T6YSPT_ID	Number	Identifier that relates each record to the unique record in the T6YSPT table.
T6YFANL_ID	Number	Identifier that relates each record to the unique record in the T6YFanls table.
TSAMDBPS_IS_NUMBER	Number	Identifier for each MDBP summary that must be combined with TSAMDBPS_ST_CODE when used.
TSAMDBPS_ST_CODE	Text	Two-digit code that identifies the State that submitted the MDBP summary.
SOURCE_TYPE_CODE	Text	The type of water source, based on whether treatment has taken place.
IDENTIFICATION_CD	Text	The unique code for identifying a water system facility sample point. This value must be unique within the Water System Facility.
DESCRIPTION_TEXT	Text	A description of the monitoring requirement.
LD_CP_TIER_LEV_TXT	Text	"Tiers" for sampling sites by water systems, established by the lead and copper rules: Tier 1: Single family residences that contain copper pipe and lead solder installed after 1982 and/or served by a lead service line Tier 2: Same as above but multi-family buildings Tier 3: Single family residence with copper pipe and lead solder installed before 1983
TYPE_CODE_CV	Text	Type of Microbial Disinfection Byproduct Summary.
REPORTED_DATE	Date/Time	Date that the MDBP Summary is reported to regulating agency.
SAMPLES_REQUIRED	Number	Number of samples required for specified analyte and water system facility.
SAMPLES_COLLECTED	Number	Number of samples collected for specified analyte and water system facility.
MR_COMPLIANCE_IND	Text	Indicates status of M&R compliance for specified analyte and water system facility.
LVL_COMPLIANCE_IND	Text	Indicates status of level compliance for the specified analyte and water system facility.
SMPLS_BYND_MEA_LVL	Number	The total number of outlier samples (i.e., samples that exceed the Max, Min, or 95P Measure Level), stored as a number.
PRCNT_BYND_MEA_LVL	Number	The percentage of outlier samples (i.e., samples that exceed the Max, Min, or 95P Measure Level), stored as a number.
PRCNT_BYND_MEA_TXT	Text	The percentage of outlier samples (i.e., samples that exceed the Max, Min, or 95P Measure Level), stored as text.
HIGHEST_MSR	Number	The highest measure during the specified monitoring period.
HIGHEST_MSR_TXT	Text	The highest measure during the specified monitoring period stored as text to preserve the trailing zeros (which indicate the precision of the measure).
CP_PRD_BEGIN_DT	Date/Time	Compliance Period Begin Date
CP_PRD_END_DT	Date/Time	Compliance Period End Date
TINWSYS_IS_NUMBER	Number	Identifier for each water system that is unique when combined with TINWSYS_ST_CODE.
TINWSYS_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the system.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that is unique when combined with TINWSF_ST_CODE.
TINWSF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.
T6YWSF_ID	Number	Unique identifier for each water system facility record.
TSASMPPT_TYPE_CODE	Text	Location type of a sampling point.
TSASMPPT_IS_NUMBER	Number	Identifier for each sample point that is unique when combined with TSASMPPT_ST_CODE.
TSASMPPT_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the sample point.

Exhibit C.7: Description of T6YFANL (facility analyte levels table)

Field Name	Data Type	Description
T6YFANL_ID	Number	Unique identifier for each facility analyte level record.
T6YANALYTE_ID	Number	Identifier that relates each record to the unique record in the T6YANALYTE table.
TMNFANL_IS_NUMBER	Number	Identifier for each facility analyte level that must be combined with TINWSYS_ST_CODE when used.
TINWSYS_IS_NUMBER	Number	Identifier for each water system that must be combined with TINWSYS_ST_CODE when used.
TINWSYS_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the system.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that must be combined with TINWSF_ST_CODE when used.
TINWSF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.
EFFECTIVE_BEG_DAT	Date/Time	The first date a facility analyte level was made effective.
EFFECTIVE_END_DAT	Date/Time	The last date a facility analyte level was effective.
REPORTED_MSR	Text	A numeric value that represents the result obtained from a single analysis, or the average result obtained from multiple analyses.
FANL_UOM_CODE	Text	A code or abbreviation for a unit of measure.
NUM_DAYS_PER_MONTH	Number	The number of days per month during the annual operation period for which this water system facility is normally in operation and/or must monitor for the analyte specified in this FANL. The number 31 is meant to signify each day within the month.
SAMPLE_RQT_PER_DAY	Number	The number of samples that must be collected during a 24-hour period from midnight to midnight for which this water system facility must monitor for the analyte specified. The number 24 is meant to signify continuous.
IND_FILT_MNTRG_FLG	Text	Individual Filter Monitoring Required Flag -- either Yes/No
SUM_TYPE_CODE_CV	Text	Type of Microbial Disinfection Byproduct Summary.
MDBP_SUM_CHK_FLG	Text	Indicates whether MDBP Summaries will be used in checking for compliance at the Facility Analyte Level.
CONTROL_LVL_MSR	Number	The measure of facility analyte control level captured as a number.
FANL_ANALYTE_CODE	Text	4-digit EPA Analyte code
FANL_ANALYTE_NAME	Text	Analyte name
T6YWS_ID	Number	Identifier that relates each record to the unique record in the T6YWS table.
T6YWSF_ID	Number	Unique identifier for each water system facility record in the T6YWSF table.

Exhibit C.8: Description of T6YSAMPSUM (sample summaries table)

Field Name	Data Type	Description
T6YSAMPSUM_ID	Number	Unique identifier for each sample summary record.
T6YANALYTE_ID	Number	Identifier that relates each record to the unique record in the T6YANALYTE table.
TSASSR_IS_NUMBER	Number	Identifier for each sample summary result that must be combined with TSASSR_ST_CODE when used.
TSASSR_ST_CODE	Text	Two-digit code that identifies the State that submitted the sample summary result.
TSASMPSM_IS_NUMBER	Number	Identifier for each sample summary that must be combined with TSASMPSM_ST_CODE when used.
TSASMPSM_ST_CODE	Text	Two-digit code that identifies the State that submitted the sample summary result.
TINWSYS_IS_NUMBER	Number	Identifier for each water system that must be combined with TINWSYS_ST_CODE when used.
TINWSYS_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the system.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that must be combined with TINWSF_ST_CODE when used.

Field Name	Data Type	Description
TINWSF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.
COLLECTION_STRT_DT	Date/Time	The earliest date the samples represented in the sample summary were collected.
COLLECTION_END_DT	Date/Time	The latest date the samples represented in the sample summary were collected.
COMPL_PURP_IND_CD	Text	Indicates whether or not the sample summary was used for compliance determination.
SAMP_SUM_TYPE_CODE	Text	Analyte Codes CU90 and PB90: 90 - 90th percentile value (lead and copper only) 95 - 95th Percentile value (lead and copper only) AL – Number of samples greater than the action level (lead and copper only) Analyte Code 3100: RT - routine samples with negative results from the distribution system.
COUNT_QTY	Number	Number of analytical results represented in the sample summary record
SAMP_SUM_MEASURE	Number	The calculated value of the results represented in the sample summary defined by the sample summary's TYPE_CODE.
SAMP_SUM_UOM_CODE	Text	The unit of measure (UOM) that is associated with the value reported for the sample summary measure.
TSAANLYT_IS_NUMBER	Number	Identifier for each analyte that is unique when combined with TSAANLYT_ST_CODE.
TSAANLYT_ST_CODE	Text	This value is "HQ" for all SDWIS/Fed contaminants. If the value is not "HQ," the analyte code is specific to the primacy agency.
ANALYTE_CODE	Text	4-digit EPA Analyte code
ANALYTE_NAME	Text	Analyte name
T6YWS_ID	Number	Identifier that relates each record to the unique record in the T6YWS table.
T6YWSF_ID	Number	Identifier that relates each record to the unique record in the T6YWSF table.

Exhibit C.9: Description of T6YCMCLV (Compliance monitoring and compliance level violations table)

Field Name	Data Type	Description
T6YANALYTE_ID	Number	Unique identifier for each treatment record.
T6YWS_ID	Number	Identifier that relates each record to the unique record in the T6YWSF table.
T6YWSF_ID	Text	Unique identifier for each water system facility record.
T6YSPT_ID	Text	Unique identifier for each sample point record.
CP_PRD_BEGIN_DT	Date	Compliance Period Begin Date.
CP_PRD_END_DT	Date	Compliance Period End Date.
AVG_TYPE_CODE	Text	The type of average represented by the MCL Value.
TSAANLYT_IS_NUMBER	Number	Identifier for each analyte that is unique when combined with TSAANLYT_ST_CODE.
TSAANLYT_ST_CODE	Text	This value is "HQ" for all SDWIS/Fed contaminants. If the value is not "HQ," the analyte code is specific to the primacy agency.
CALCULATED_VALUE	Number	The value for a given analyte, sampling location and period of time that is compared against an MCL to determined compliance.
UOM_CODE	Text	The measurement units used to express the measure or value.
NUMB_RESULTS_USED	Number	The number of results used in the calculation of a given Monitoring Period Average.
PRC_ACH_RMVL_RA_NO	Number	Precursor Achieved Removal Ratio Number Used by the Calculate MCL
AVG_DUR_TYPE_CD	Text	The type of monitoring period, i.e., monthly, quarterly, annually.
AVG_NBR_MON_PRD	Number	The number of monitoring periods covered by the average.
BIN_NUMBER	Text	The BIN assignment for the period of time covered by the average.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that is unique when combined with

Field Name	Data Type	Description
		TINWSF_ST_CODE.
TINWSF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.
TSASMPPT_IS_NUMBER	Number	Identifier for each sample point that is unique when combined with TSASMPPT_ST_CODE.
TSASMPPT_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the sample point.
MP_TYPE_CODE	Text	The code of monitoring period, i.e., monthly, quarterly, annually.
T6YCMCLV_ID	Number	Unique identifier for each calculated compliance value.

Exhibit C.10: Description of T6YCORACT (Corrective Actions)

Field Name	Data Type	Description
T6YCORACT_ID	Number	Unique identifier for each corrective action.
T6YWS_ID	Number	Identifier that relates each record to the unique record in the T6YWSF table.
TINWSYS_IS_NUMBER	Number	Identifier for each water system that is unique when combined with TINWSYS_ST_CODE.
TINWSYS_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the system.
DATE_ISSUE_IDENTIFIED	Text	Date the corrective action was identified.
SCHEDULE_TYPE	Text	Type of schedule for the corrective action.
SCHEDULE_DESCRIPTION	Text	Schedule for the corrective action.
CORACT_CATEGORY_CODE	Text	Category code for the corrective action.
CORACT_NAME	Text	Name of the corrective action.
DUE_DATE	Date	Due date for the required corrective action.
ACHIEVED_DATE	Date	The date that the water system achieved the corrective action required.
TENSCHD_IS_NUMBER	Number	Identifier for each corrective action compliance schedule that must be combined with TENSCHD_ST_CODE when used.
TENSCHD_ST_CODE	Text	Two-digit code that identifies the State of the corrective action compliance schedule.

Exhibit C.11: Description of T6YMCL_MDL (Maximum contaminant level and minimum detection level table)

Field Name	Data Type	Description
T6YMCL_MDL_ID	Number	Unique identifier for each MCL or MDL
ANALYTE_CODE	Text	4-digit EPA Analyte code
CHEMGRP	Text	Chemical Group
DB_MCL	Number	Maximum Contaminant Level
DB_MCL_UNIT	Text	Maximum Contaminant Level Unit of Measure
DB_4XMCL	Number	Four times the Maximum Contaminant Level
MDL	Number	Method Detection Limit
MDL_10TH	Number	One-tenth the Method Detection Limit

Exhibit C.12: Description of T6YWSFPLT (Treatment plant water system facilities table)

Field Name	Data Type	Description
T6YWSFPLT_ID	Number	Unique identifier for each treatment plant water system facility record.
T6YWSF_ID	Number	Identifier that relates each record to the unique record in the T6YWSF table.
ST_ASGN_IDENT_CD	Text	A State-assigned value which identifies the treatment plant water system facility.
WSF_TYPE_CODE	Text	The value extracted from SDWIS/State will be "TP" (treatment plant).
FILTER_TYPE	Text	(Unfiltered (UF), Conventional Filtration (CF), Direct Filtration (DF), Diatomaceous Earth (DE), Other (OT), and other permitted values that the System Administrator may add)
FILTER_DESCRIPTION	Text	A description of the filter.
DISINFECT_CONCENTN	Text	Disinfectant Concentration in mg/L
CONTACT_TIME_STAT	Text	Contact Time Status (Permitted values): RQD – Required; NRQD - Not Required; REQT – Requested; RECV – Received; URVW - Under Review; RVWD – Reviewed; APVD – Approved; DTMD – Determined; DENY – Denied; RESB - Resubmitted
CT_TIME_DETERM_DAT	Date/Time	Date the Contact Time was determined
CONTACT_TIME	Text	Contact Time in minutes--the number of minutes the water was in contact with the disinfectant to be properly disinfected. The range of values is 0001 to 2400
CT_VALUE	Text	Contact value in mg/min/liter
DBM_GIA_INACT_LOG	Number	The disinfection profile benchmark for *Giardia* inactivation in Logs.
DBM_GIA_INACT_STAT	Text	The status of the disinfection profile benchmark for *Giardia* inactivation. See CONTACT_TIME_STAT for permitted values and description
DBM_GIA_INACT_DT	Date/Time	The date the disinfection virus benchmark was determined.
DBM_GIA_INACT_PCT	Number	The disinfection profile benchmark for *Giardia* inactivation percent.
DBM_VIR_INACT_LOG	Number	The disinfection profile benchmark for virus inactivation in Logs.
DBM_VIR_INACT_STAT	Text	The status of the disinfection profile benchmark for Virus inactivation. See CONTACT_TIME_STAT for permitted values and description
DBM_VIRUS_INACT_DT	Date/Time	The date the disinfection virus benchmark was determined.
DBM_VIR_INACT_PCT	Number	The disinfection profile benchmark for virus inactivation percent.
BIN_STATUS	Text	The status of the BIN determination for the Long Term 2 Surface Water Treatment Rule. See CONTACT_TIME_STAT for permitted values and description.
BIN_LT2	Number	The BIN number for the Long Term 2 Surface Water Treatment Rule.
BIN_DETERM_DT	Date/Time	The date the BIN number was determined for the Long Term 2 Surface Water Treatment Rule.
FBR_SCHEMATIC_STAT	Text	Under the Filter Backwash Rule, a water system is required to submit a schematic of this treatment plant to the primacy agency for review to demonstrate the percentage of filter backwash that is returned to the treatment plant influent. See CONTACT_TIME_STAT for permitted values and description.
FBR_SCHEMA_RCV_DAT	Date/Time	Date primacy agency received treatment plant schematic to demonstrate the percentage of filter backwash that is returned to the treatment plant influent.
FBR_SCHEMA_RVW_DAT	Date/Time	Date primacy agency completes review of treatment plant schematic and determines the percentage of filter backwash that is returned to the treatment plant influent.
FBR_ALTR_RTN_RQS	Text	The status of a request from the water system to request an alternate location for return of the filter backwash.
FBR_ALTR_RTN_DT	Date/Time	The date that the water system requested an alternate location for return of the filter backwash.
FBR_CORCTV_ACT_RQS	Text	The status of corrective action by the water system as required by the primacy agency after review of the schematic of the filter backwash flow in the treatment plant.

Field Name	Data Type	Description
FBR_CORCTV_ACT_DT	Date/Time	The date that the water system achieved the corrective action required for the filter backwash.
WSF_NAME	Text	Name of the water system facility.
FBR_COMMENTS	Text	A memo field into which a user may enter comments about the Filter Backwash Recycling Rule.
DSNF_BMRK_REASON	Text	Text description associated with the Disinfection Benchmark Reason
CONTACT_TIM_REASON	Text	Text description associated with the Contact Time

Exhibit C.13: Description of T6YTREATPROCESS (Treatments associated to treatment plants table)

Field Name	Data Type	Description
T6YTREATPROCESS_ID	Number	Unique identifier for each treatment record.
T6YWSF_ID	Number	Identifier that relates each record to the unique record in the T6YWSF table.
TINTROBJ_CODE	Text	A coded value that categorizes the treatment objective.
TINTROBJ_NAME	Text	The name of the treatment objective.
TINTRPRO_CODE	Text	A coded value that categorizes the treatment process.
TINTRPRO_NAME	Text	The name of the treatment process.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that is unique when combined with TINWSF_ST_CODE.
TINWSF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.

Exhibit C.14: Description of T6YWSFFLOWS (Water system facility flows table)

Field Name	Data Type	Description
T6YWSFFLOWS_ID	Number	Unique identifier for each water system facility flow record.
T6YWSF_ID	Number	Identifier that relates each record to the unique record in the T6YWSF table.
TINWSFF_IS_NUMBER	Number	Identifier for each water system facility flow entry that is unique when combined with T6YWSF_ID.
TINWSFF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility flow entry.
TRAIN_ID	Text	This attribute identifies the water system facilities that are part of the same flow.
SEQUENCE_ID	Text	This attribute identifies the order of the water system facilities in a specific flow.
PROCESS_WATER_TYPE	Text	A system administrator-controlled code of the type of water flowing between the facilities.
WATER_QTY_MSR	Number	A value that represents the number of gallons of water purchased.
WATER_QTY_MSR_UNIT	Text	A coded value which specifies the unit of measurement for the quantity of water purchased.
CONNECTION_TYPE_CD	Text	Categorizes the type of connection between the water system facilities.
CONNECTION_DATE	Date/Time	The date of the connection of the water system facility to another water system facility.
DISCONNECTION_DATE	Date/Time	The date of the disconnection of the water system facility from another water system facility.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that is unique when combined with TINWSF_ST_CODE.
TINWSF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.
TINWSF0IS_NUMBER	Number	Identifier for each supplying water system facility that is unique when combined with TINWSF0ST_CODE.

Field Name	Data Type	Description
TINWSF0ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.
T6YWSF0ID	Number	Unique identifier for each supplying water system facility.

Exhibit C.15: Description of T6YWSFIND (Water system facility indicators table)

Field Name	Data Type	Description
T6YWSFIND_ID	Number	Unique identifier for each water system facility indicator record.
T6YWSF_ID	Number	Identifier that relates each record to the unique record in the T6YWSF table.
TINWSFIN_IS_NUMBER	Number	Identifier for each water system facility indicator that is unique when combined with T6YWSF_ID
WSF_IND_NAME	Text	The water system facility indicator name.
WSF_IND_DESC	Text	The description of the water system facility indicator name.
WSF_IND_VALUE_CD	Text	The value of the indicator established by the primacy agency.
WSF_IND_DATE	Date/Time	The date associated with the indicator.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that is unique when combined with TINWSF_ST_CODE.
TINWSF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.

Exhibit C.16: Description of T6YWSIND (Water system indicators table)

Field Name	Data Type	Description
T6YWSIND_ID	Number	Unique identifier for each water system indicator record.
T6YWS_ID	Number	Identifier that relates each record to the unique record in the T6YWS table.
TINWSIN_IS_NUMBER	Number	Identifier for each water system indicator that is unique when combined with. T6YWSF_ID.
TINWSYS_IS_NUMBER	Number	Identifier for each water system that is unique when combined with TINWSYS_ST_CODE.
TINWSYS_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the system.
WS_IND_NAME	Text	The water system indicator name.
WS_IND_DESC	Text	The description of the water system indicator name.
WS_IND_VALUE_CD	Text	The value of the indicator established by the primacy agency.
WS_IND_DATE	Date/Time	The date associated with the indicator.

Exhibit C.17: Description of T6YWSPURCH (Water system buyers and sellers)

Field Name	Data Type	Description
T6YWSPURCH_ID	Number	Unique identifier for each water system buyer and seller record.
T6YWS_ID	Number	Identifier that relates each record to the unique record in the T6YWS table.
TINWSYS0IS_NUMBER	Number	Identifier for each supplying water system that is unique when combined with TINWSYS0ST_CODE.
TINWSYS0ST_CODE	Text	Two-digit code that identifies the State that submitted data for the supplying water system.
TINWPURC_IS_NUMBER	Number	Identifier for each water system purchase record that must be combined with TINWSYS0ST_CODE when used.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that must be combined with TINWSF_ST_CODE when used.

Field Name	Data Type	Description
TINWSF_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the facility.
TINWSF0IS_NUMBER	Number	Identifier for each supplying water system facility record that must be combined with TINWSF0ST_CODE when used.
TINWSF0ST_CODE	Text	Two-digit code that identifies the State that submitted data for the supplying facility.
T6YWS0ID	Number	Unique identifier for each supplying water system.
TINWSYS_IS_NUMBER	Number	Identifier for each water system that is unique when combined with TINWSYS_ST_CODE.
TINWSYS_ST_CODE	Text	Two-digit code that identifies the State that submitted data for the system.
T6YWSF_ID	Number	Unique identifier for each water system facility record.
T6YWSF0ID	Number	Unique identifier for each supplying water system facility.

Exhibit C.18: Description of T6YSAR_TRANSACTION (Sample analytical result transaction table)

Field Name	Data Type	Description
T6Y_TRANSACTION_ID	Number	Unique identifier for each transaction. (Note: Some records will be listed more than once if they were flagged for more than one reason such as being greater than 4*MCL and greater than 10*MCL.)
T6YSAR_ID	Number	Unique identifier for each sample analytical result (enables linking to T6YSAR).
TSASAR_IS_NUMBER	Number	Identifier for each sample analytical result that is unique when combined with TSASAR_ST_CODE.
TSASAR_ST_CODE	Text	Two-digit code that identifies the State that submitted data.
QA_FLAG_ID	Number	A coded value (1 through 11) that identifies the reason that the record was flagged. Values have the following descriptions: 1 = flagged a s a potential duplicate; 2 = flagged as a transient sample for an analyte for which transient systems are not required to sample; 3 = flagged as a non-compliance sample; 4 = flagged as a non-routine sample; 5 = flagged as 4 times greater than the MCL; 6 = flagged as 10 times greater than the MCL; 7 = flagged as less than the MDL; 8 = flagged as less than 1/10th of the MDL; 9 = flagged for having abnormal units; 10= DBP samples flagged as taken outside the distribution system/entry point; and 11 = Utah nitrate or nitrite records flagged as being assigned an inaccurate analyte code.
ACTION_ID	Number	A coded value (1 through 3) that identifies the reason that the record was flagged. Values have the following descriptions: 1 = no change; 2 = one of the record's fields was changed; 3 = record excluded (or a duplicate).
ANALYZE	Text	Field contains "yes" or "no," identifying whether or not the record will be included in the occurrence analysis.
REMARK	Text	Text describing the QA issues, as well as other notes related to the record.
STATERESPONSE	Text	Verbatim response from the State on the flagged record (when available).
ACTIONDETAIL	Text	Additional detail on the record's "action" such as why the record was excluded or changed.
CREATEDATE	Date/Time	Date the transaction was entered into the database.
LASTMODIFIEDDATE	Date/Time	Date the transaction record was last modified.
ACTION_ID_CLEAN	Number	A coded value (1 through 4) that identifies the reason that the record was flagged. Values have the following descriptions: 1 = no change; 2 = one of the record's fields was changed; 3 = record excluded; 4 = duplicate record (which may or may not be excluded as one copy of the duplicate is retained).

Field Name	Data Type	Description
NEW_COLUMN	Text	Field indicating which column in “T6YSAR” should be modified by the transaction record.
NEW_VALUE_DATE	Date	New value to replace the existing value in “T6YSAR” that should be modified by the transaction record. Only stores values if they are in Date format.
NEW_VALUE_TEXT	Text	New value to replace the existing value in “T6YSAR” that should be modified by the transaction record Only stores values if they are in Text format.
NEW_VALUE_NUMERIC	Number	New value to replace the existing value in “T6YSAR” that should be modified by the transaction record. Only stores values if they are in Number format.
COLUMN_TYPE	Number	A coded value (1 through 3) that identifies the column that stores the value that will replace the existing value in “T6YSAR” that should be modified by the transaction record. 1 = NEW_VALUE_DATE, 2 = NEW_VALUE_TEXT, 3 = NEW_VALUE_NUMERIC.
NUMBER0	Text	Public water system identification number (PWSID) derived from T6YSAR.
COLLECTION_END_DT	Date	The latest date the samples represented in the sample summary were collected derived from T6YSAR.
CONCENTRATION_MSR	Number	A numeric value that represents the result obtained from a sample analysis derived from T6YSAR.
LAB_ASGND_ID_NUM	Text	An identifier used for reconciliation with the state data system or sample identification number assigned by the laboratory derived from T6YSAR.
ANALYTE_CODE	Text	4-digit EPA Analyte code
QA_TRANSACT_ID	Number	Unique identifier for QA of each transaction.

Exhibit C.19: Description of T6YWS_TRANSACTION (Water system transaction table)

Field Name	Data Type	Description
T6YWS_TRANSACTION_ID	Number	Unique identifier for each transaction. (Note: Some records will be listed more than once if they were flagged for more than one reason such as being greater than 4*MCL and greater than 10*MCL.)
T6YWS_ID	Number	Unique identifier for each sample analytical result (enables linking to T6YSAR).
TINWSYS_IS_NUMBER	Number	Identifier for each sample analytical result that is unique when combined with TSASAR_ST_CODE.
TINWSYS_ST_CODE	Text	Two-digit code that identifies the State that submitted data.for the system
QA_FLAG_ID	Number	A coded value (1 through 11) that identifies the reason that the record was flagged. Values have the following descriptions: 1 = flagged a s a potential duplicate; 2 = flagged as a transient sample for an analyte for which transient systems are not required to sample; 3 = flagged as a non-compliance sample; 4 = flagged as a non-routine sample; 5 = flagged as 4 times greater than the MCL; 6 = flagged as 10 times greater than the MCL; 7 = flagged as less than the MDL; 8 = flagged as less than 1/10th of the MDL; 9 = flagged for having abnormal units; 10= DBP samples flagged as taken outside the distribution system/entry point; and 11 = Utah nitrate or nitrite records flagged as being assigned an inaccurate analyte.
ACTION_ID	Number	A coded value (1 through 3) that identifies the reason that the record was flagged. Values have the following descriptions: 1 = no change; 2 = one of the record’s fields was changed; 3 = record excluded (or a duplicate).
ANALYZE	Text	Field contains "yes" or "no," identifying whether or not the record will be included in the occurrence analysis.
REMARK	Text	Text describing the QA issues, as well as other notes related to the record.
CREATEDATE	Date/Time	Date the transaction was entered into the database.

Field Name	Data Type	Description
LASTMODIFIEDDATE	Date/Time	Date the transaction record was last modified.

Appendix D: Occurrence data for the Aircraft Drinking Water Rule (ADWR)

In May 2021, EPA downloaded compliance monitoring data from its Aircraft Reporting and Compliance System (ARCS) for evaluation under SYR 4. ARCS is a centralized web-based data collection and management system that provides accountability and regulatory oversight and is used to facilitate the reporting of aircraft public water system (PWS) data. This data is also publicly available on the ADWR Compliance Reports website: https://www.epa.gov/dwreginfo/adwr-compliance-reports. Air carriers subject to the ADWR must report to EPA and conduct, as appropriate, the following actions in ARCS, unless an alternative reporting method has been approved (https://www.epa.gov/dwreginfo/aircraft-drinking-water-rule):

- A complete inventory of aircraft PWS fleet;
- PWS activity details, such as whether the aircraft is currently in an active or inactive status.
- The date the Operations and Maintenance plan was developed;
- The date the Coliform Sampling plan was developed;
- The date the aircraft PWS Sampling plan(s) was incorporated into the aircraft water system Operations and Maintenance plan;
- The date the Operations and Maintenance plan(s) was incorporated into FAA-accepted air carrier Operation and Maintenance program;
- The frequency for routine disinfection and flushing, and the corresponding routine total coliform sampling frequency; and
- The date for routine disinfection and flushing, routine coliform sampling dates and results, and corrective actions (when applicable).

Approximately 212,937 records[9] of aircraft PWS compliance monitoring data for total coliform (TC) and *E. coli* (EC) samples were available in ARCS from February 2011 through May 2021, including results reported for more than 70 different makes/models of aircraft. These results were used to characterize the positivity rates of TC and EC in aircraft PWSs on an annual basis, as well as for all the years that data were available (2011-2021) and for the subset of years 2012 through 2019. The evaluation of data for years 2012 through 2019 was performed to allow for a comparison with similar data for stationary PWSs as described in Section 5.5. In addition, this approach removes potentially confounding considerations associated with evaluating data for calendar year 2020 when a large number of aircraft PWS were inactive due to COVID-19, as well as years 2011 and 2021 for which the ARCS data evaluated at this time only represents partial years.

Aircraft inventory data, including manufacturer, model, and disinfection and flushing frequency,

[9] The number of records presented here is greater than the number of rows of data downloaded from ARCS (70,979 at the time of download in support of the SYR 4 analysis) because it counts all samples within each row of data (i.e., Sample 1, Sample 2, and Sample 3). Note that Sample 3 is related to the ability to have third sample collected, which is not a requirement of ADWR and is not often used. Typically there is no data for Sample 3 fields.

were linked to the monitoring results by public water system identification number (PWSID). Aircraft PWS were categorized as small, medium, or large based on the seat capacity (small = ≤130 seats; medium = >130 – 250 seats; large = >250 seats). Note that these categories were developed specifically for this analysis, based on the dataset and do not represent regulatory categories. ADWR does not categorize aircraft PWS based on size. In addition, the first three digits of the model number were used to summarize the make/model of each aircraft. For example, inventory data showing model numbers for Boeing as 737800, 737-823, and 7377BD all were captured in this analysis as 737.

A number of quality assurance (QA) steps were applied to the ADWR dataset to identify the TC and EC records suitable for analysis. Data were excluded via the following QA steps:

- Records where [Location] was "-" were excluded (72,406 records)
- Remaining records where [Total Coliform] was "-" or "from" were excluded (4).
- Remaining records where [Sample Taken On] date was incorrectly entered were excluded. These dates were as follows: 12/08/0014 00:00", "09/26/0201 03:52", "09/13/0019 03:59", "09/09/0201 03:35", "07/22/0204 05:17", "07/16/0018 01:35", "06/21/0018 01:40", and "02/02/0017 16:10" (16 records).
- Remaining records where [Total Coliform] result was entered as "absent" but [*E. coli*] was positive (9 records).

The ADWR analyses were stratified in a variety of ways to summarize results, including the number of TC samples and public water systems by aircraft size, manufacturer, model, air carrier, sample type, and more. It is important to note that all EC positivity rates were calculated twice, under two different sets of assumptions:

1. An EC sample was included in the analysis only if the EC result was listed as "Present" or "Absent."
2. An EC sample was included if the EC result was listed as "Present" or "Absent" (i.e., the same as the first set of assumptions) but with an added consideration of assuming that an EC sample was "Absent" if the associated TC result was reported as "Absent" and there was no EC result provided. These results are labeled in the file as "*E. coli* (Alternative Approach)."

After the QA steps were applied, there were 140,502 TC results used in this evaluation, provided by 8,093 PWSs and covering the full range of years for which ARCS data were collected (i.e., February 2011 – May 2021). Of those results, 7,250 results (5.2 percent) were positive for TC. Under the first approach for calculating EC positivity rates listed above, there were 92,994 EC results provided by 7,091 PWSs (i.e., 66 percent of the number of TC results and 88 percent of the aircraft PWSs), with a total of 241 results (0.26 percent) positive for EC. Under the second approach for calculating EC positivity rates listed above, there were 140,485 EC results provided by 8,093 aircraft PWSs, with 241 results (0.17 percent) positive for EC.

Considering only the 8-year period from 2012-2019, there were 118,070 TC results used in this evaluation, provided by 7,816 PWSs. Of those results, 6,448 results (5.5 percent) were positive

for TC. Under the first approach for calculating EC positivity rates listed above, there were 78,114 EC results provided by 6,776 PWSs (i.e., 66 percent of the number of TC results and 87 percent of the PWSs), with a total of 201 results (0.26 percent) positive for EC. Under the second approach for calculating EC positivity rates, there were 118,056 EC results provided by 7,816 PWSs, with 201 results (0.17 percent) positive for EC.

Data users will find a difference between the number of FAA Corporate Names in the inventory versus the samples file. The difference is due to the inventory FAA Corporate Names covering the last year of data collection and the samples file covering all the years of data collection. Some of the additional air carriers listed in the sampling file have since merged or gone out of business. For more on ADWR analyses, see *Six-Year Review 4 Technical Support Document for Microbial Contaminant Regulations* (USEPA, 2024b).

Appendix E: User Guide to Downloading and Using Six-Year Review 4 and Related Data from EPA's Website

This appendix includes a user guide for downloading and using the Six-Year Review 4 (SYR 4) and related data from EPA's website. This document is also posted online with the data. In addition, instructions on importing the SYR 4 datasets are included in this Appendix (see Section 10). The data dictionary for all datasets is also included in Appendix C above.

Several of the contaminant occurrence datasets that are posted online were not analyzed as part of the SYR 4 effort. These contaminants were not subject to detailed review in SYR 4 due to recent, ongoing, or pending regulatory action (e.g., lead, copper, DBPs). These datasets passed the same QA procedures as those analyzed in SYR 4.

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 varies. 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 Exhibit E.1.

The remainder of this document is organized as follows:

Section 1: Background Information on SYR 4 Data Records

To support the national contaminant occurrence and exposure assessments performed under the fourth Six-Year Review process (SYR 4), EPA collected compliance monitoring data and treatment technique information from public water systems (PWSs) for regulated drinking water contaminants. This analysis allows EPA to characterize the frequency of occurrence, the levels found, and the geographic distribution of contaminants and related data to help the agency determine if there may be a meaningful opportunity to improve public health protection. EPA conducted a voluntary data request from states, primacy agencies, territories, and tribes (referred to as "States" throughout the remainder of this Appendix) to obtain compliance monitoring data and treatment technique information necessary to analyze national contaminant occurrence in support of SYR 4. This data request was conducted through the Information Collection Request (ICR) process. EPA requested States to submit their Safe Drinking Water Act (SDWA) compliance monitoring data and treatment technique information collected between January 2012 and December 2019. 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).

Through extensive data management efforts, quality assurance evaluations, and communications and consultations with State's data management staff, EPA established a single contaminant occurrence dataset that consists of compliance monitoring data and treatment technique information from 59 out of 66 jurisdictions (46 states plus Washington, D.C., American Samoa, Navajo Nation, Commonwealth of the Northern Mariana Islands, and other tribes). This dataset is referred to as the National Compliance Monitoring ICR dataset for the fourth Six-Year Review (SYR 4 ICR dataset). The 59 States that provided data for the SYR 4 ICR dataset comprise 88 percent of all PWSs and 92 percent of the total population served by PWSs nationally, and are geographically representative of PWSs nationwide. The SYR 4 ICR dataset was used to estimate a variety of occurrence measures to characterize the national occurrence of regulated contaminants in public water systems to support the Six-Year Review process.

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 States, to extract the requested data identified in the SYR 4 ICR from a SDWIS/State database. In all, 46 states and 13 other jurisdictions 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 not using SDWIS/State submitted their compliance monitoring data and treatment technique "as is," resulting in a variety of formats, including dBase, Excel, XML, 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 online via EPA's Central Data Exchange. All data were conformed to a similar format with consistent units of measurement for consistency. For more details about the collection and formatting of SYR 4 ICR data, see the main chapters of this document.

EPA conducted a quality assurance and control evaluation of these data submitted by States and assembled these data into the SYR 4 ICR database, which includes more than 83 million records

from approximately 142,000 public water systems, serving approximately 303 million people nationally. The dataset includes the results of all compliance monitoring data (i.e., all sample analytical detections and non-detections) from January 2012 to December 2019 for regulated chemical phase contaminants, radionuclides, disinfectants and disinfection byproducts (D/DBPs), DBP precursors, microbial contaminants/indicators, disinfectant residuals, and other related data including treatment information. As noted in the main chapters, only the data that passed the QA/QC process are posted online.

Exhibit E.1, Six-Year Review 4 Data Field Names and Definitions, contains a list of the data elements, column names and a brief description of the data for each data element included in the SYR 4 ICR data text files.

Exhibit E.1: Six-Year Review 4 Data Field Names and Definitions

Data Element	Column Name	Description
Contaminant Identification Code	ANALYTE_CODE	4-digit Safe Drinking Water Information System (SDWIS) contaminant identification number for which the sample is being analyzed.
Contaminant Name	ANALYTE_NAME	Common name of contaminant for which the sample is being analyzed.
Primacy Code	PRIMACY_CODE	2- digit code identifying the primacy agency (i.e., State) for the water system.
State Code	STATE_CODE	2-digit code identifying the U.S. state or territory in which the water system is located.
Public Water System Identification Number (PWSID)	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.
Federal Public Water System Type Code	SYSTEM_TYPE	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_POPULATIO N_SERVED	Retail population served by a system.
Adjusted Total Population-served	ADJUSTED_TOTAL_ POPULATION_ SERVED	Adjusted total population served (retail plus adjusted wholesale population served as not to double-count buyer systems that purchase from multiple seller 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).
Facility Identification Code	WATER_FACILITY_ID	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;

Data Element	Column Name	Description
		• OT = Other; • PC = Pressure Control; • PF = Pumping Facility; • 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 Identification Code	SAMPLING_POINT_ID	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 Identification Number	LABORATORY_ASSIGNED_ID	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 Identification Number	SAMPLE_ID	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 Value	DETECTION_LIMIT_VALUE	Limit below which the specific lab indicated they could not 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.)
Sample Analytical Result - Sign	DETECT	The sign indicates whether the sample analytical result was: • (0) "less than" means the contaminant was not detected or was

Data Element	Column Name	Description
		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."
Sample Analytical Result - Value	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.
Sample Analytical Result - Unit of Measure	UNIT	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	RESIDUAL_FIELD_ FREE_CHLORINE_M G_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 Chlorine	RESIDUAL_FIELD_ TOTAL_CHLORINE_ MG_L	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).

Section 2: SYR 4 Data Records Posted for Phase Chemicals, Lead, Copper and Radionuclides

Exhibit E.2 provides a count of States, total number of sample records and systems for each phase chemical, lead, copper, and radionuclide collected for SYR 4. Contaminant occurrence data are grouped into zip files, which are indicated in the final column of Exhibit E2.

Exhibit E.2: Number of Six-Year Review 4 Data Records for Phase Chemicals, Lead, Copper, and Radionuclides and Zip Filename(s)

Contaminant	Analyte ID	Number of States	Total Number of Sample Records	Total Number of Systems	Zip Filename
Phase Chemicals					
1,1,1-Trichloroethane	2981	58	491,411	52,207	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
1,1,2-Trichloroethane	2985	58	482,294	52,200	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
1,1-Dichloroethylene	2977	58	508,764	52,206	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
1,2,4-Trichlorobenzene	2378	58	480,039	52,201	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
1,2-Dibromo-3-chloropropane	2931	57	244,298	37,153	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
1,2-Dichloroethane	2980	58	493,514	52,209	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
1,2-Dichloropropane	2983	58	481,065	52,197	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
2,3,7,8-TCDD	2063	42	38,934	6,222	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
2,4,5-TP	2110	58	187,025	40,954	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
2,4-D	2105	58	191,658	41,519	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
Alachlor	2051	58	215,965	42,822	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
Antimony, Total	1074	59	230,942	51,063	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
Arsenic	1005	59	452,852	52,505	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
Asbestos	1094	48	24,124	13,772	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
Atrazine	2050	58	225,827	43,763	SYR4_PhaseChem_1 (1,1,1-Trichloroethane to Atrazine).zip
Barium	1010	59	232,216	52,488	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Benzene	2990	58	487,631	52,207	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Benzo(A)pyrene	2306	58	190,003	35,877	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Beryllium, Total	1075	59	229,630	50,225	SYR4_PhaseChem_2 (Barium to Cyanide).zip
BHC-Gamma	2010	58	195,775	38,843	SYR4_PhaseChem_2 (Barium to Cyanide).zip

Contaminant	Analyte ID	Number of States	Total Number of Sample Records	Total Number of Systems	Zip Filename
Cadmium	1015	59	230,098	50,989	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Carbofuran	2046	58	176,608	37,375	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Carbon Tetrachloride	2982	58	510,599	52,205	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Chlordane	2959	58	189,512	38,310	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Chlorobenzene	2989	58	479,909	52,184	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Chromium	1020	59	238,413	51,357	SYR4_PhaseChem_2 (Barium to Cyanide).zip
cis-1,2-Dichloroethylene	2380	58	495,228	52,210	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Cyanide	1024	57	163,373	38,760	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Dalapon	2031	58	232,471	40,062	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Di(2-Ethylhexyl) Adipate	2035	58	192,447	36,369	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Di(2-Ethylhexyl) Phthalate	2039	58	202,419	36,486	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Dichloromethane	2964	58	487,166	52,222	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Dinoseb	2041	58	186,403	40,854	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Diquat	2032	54	110,637	22,215	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Endothall	2033	51	98,015	18,624	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Endrin	2005	58	192,869	38,483	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Ethylbenzene	2992	58	487,555	52,200	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Ethylene Dibromide	2946	57	243,161	38,371	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Fluoride[1]	1025	59	435,466	52,202	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Glyphosate	2034	55	105,084	21,744	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Heptachlor	2065	58	193,927	38,640	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Heptachlor Epoxide	2067	58	193,623	38,638	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Hexachlorobenzene	2274	58	195,150	38,311	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Hexachlorocyclopentadiene	2042	58	196,236	38,471	SYR4_PhaseChem_3 (Dalapon to Hexachlorocyclopentadiene).zip
Mercury	1035	59	226,418	50,990	SYR4_PhaseChem_4 (Hybrid Nitrate to Nitrate).zip
Methoxychlor	2015	58	196,131	38,834	SYR4_PhaseChem_4 (Hybrid Nitrate to Nitrate).zip
Nitrate	1040	59	1,404,609	105,202	SYR4_PhaseChem_4 (Hybrid Nitrate to Nitrate).zip
Nitrate (Hybrid)[2]	1040/ 1038	59	1,635,300	127,904	SYR4_PhaseChem_4 (Hybrid Nitrate to Nitrate).zip

Contaminant	Analyte ID	Number of States	Total Number of Sample Records	Total Number of Systems	Zip Filename
Nitrite	1041	59	512,234	73,442	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Nitrate-Nitrite	1038	51	561,314	76,530	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
o-Dichlorobenzene	2968	58	480,075	52,200	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Oxamyl	2036	58	175,728	37,235	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
p-Dichlorobenzene	2969	58	480,247	52,203	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Pentachlorophenol	2326	58	201,636	41,094	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Picloram	2040	58	188,833	41,375	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Selenium	1045	59	232,598	51,317	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Simazine	2037	58	220,013	43,211	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Styrene	2996	58	479,601	52,187	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Tetrachloroethylene	2987	58	544,460	52,210	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Thallium, Total	1085	59	229,685	51,007	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Toluene	2991	58	488,192	52,348	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Total Polychlorinated Biphenyls (PCB)	2383	49	116,454	23,262	SYR4_PhaseChem_5 (Nitrate-Nitrite to Total Polychlorinated Biphenyls (PCB)
Toxaphene	2020	58	183,765	37,419	SYR4_PhaseChem_6 (Toxaphene to Xylenes, total).zip
trans-1,2-Dichloroethylene	2979	58	488,716	52,194	SYR4_PhaseChem_6 (Toxaphene to Xylenes, total).zip
Trichloroethylene	2984	58	540,777	52,222	SYR4_PhaseChem_6 (Toxaphene to Xylenes, total).zip
Vinyl Chloride	2976	58	482,672	52,021	SYR4_PhaseChem_6 (Toxaphene to Xylenes, total).zip
Xylenes, Total	2955	56	412,436	46,720	SYR4_PhaseChem_6 (Toxaphene to Xylenes, total).zip
Lead and Copper					
Lead	1030	54	1,552,995	53,058	SYR4_PhaseChem_4 (Hybrid nitrate to nitrate).zip

Contaminant	Analyte ID	Number of States	Total Number of Sample Records	Total Number of Systems	Zip Filename
Copper	1022	55	1,579,728	54,224	SYR4_PhaseChem_2 (Barium to Cyanide).zip
Radionuclides					
Gross Alpha, Excl. Radon & U	4000	55	64,413	16,925	SYR4_Rads.zip
Gross Beta Particle Activity	4100	50	48,520	11,261	SYR4_Rads.zip
Combined Radium (-226 & -228)	4010	53	86,594	21,972	SYR4_Rads.zip
Combined Uranium	4006	55	97,663	18,491	SYR4_Rads.zip

[1] Includes records that passed the QA/QC procedures described in this document. See USEPA (2024c) for additional information on procedures conducted for the occurrence analysis.
[2] Includes all sampling results for nitrate and sampling results for total nitrate plus nitrite for systems for which there were no SYR 4 nitrate (only) data.

Section 3: SYR 4 Data Records Posted for Disinfection Byproducts

Exhibit E.3 provides a count of States, total number of sample records and systems for each regulated disinfection byproduct collected for SYR 4, and the zip files names that the data files can be located. These data records were not analyzed under SYR 4 because of the ongoing considerations of potential revisions of the Stage 1 and Stage 2 DBP Rules.

Exhibit E.3: Number of Six-Year Review 4 Data Records for Disinfection Byproducts and Zip filename(s)

Contaminant	Analyte ID	Number of States	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
Monobromoacetic 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

Note: The speciation data is higher for TTHM than HAA5 (90+% vs 70+%). There were two more States that provided speciated THM results as compared to speciated HAA results. About 11,000 systems had speciated THM data but not speciated HAA data. There are only about 200 systems with speciated HAA data but no speciated THM data. In addition, the number of PWSs providing speciated TTHM data is higher than number of PWSs providing TTHM. There are about 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.

Section 4: SYR 4 Data Records Posted for Disinfection Byproduct Related Parameters

This DBP-related data includes total organic carbon (TOC), alkalinity, pH, dissolved organic carbon (DOC), specific UV-absorbance (SUVA), and UV-absorbance. Full datasets for TOC and alkalinity (i.e., text files including all individual sample analytical results for TOC and alkalinity) are included. In addition to the full datasets for TOC and alkalinity, a paired TOC-alkalinity dataset was created that included, for each treatment plant (listed in Exhibit E.2 as a water system facility), the average monthly concentrations of TOC and alkalinity in 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 SYR 4 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. EPA produced these datasets to support the ongoing considerations of potential revisions of the Stage 1 DBP Rule (85 FR 61680, USEPA, 2020). EPA did not analyze these data records under the SYR 4 effort. Historical efforts to evaluate the paired TOC-alkalinity data are described in *Six-Year Review 3 Technical Support Document for Disinfectants/Disinfection Byproducts Rules* (USEPA, 2016).

Exhibit E.4 provides a count of States, total number of sample records and systems for TOC (raw and finished), alkalinity, pH, DOC, SUVA, and 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. The "full" TOC dataset contains only the raw/finished water designations from the original data provided by the State (see SOURCE_TYPE_CODE). However, for the "paired" TOC-alkalinity dataset, EPA applies the following logic to assign raw/finished water designations to records that were missing it. Raw samples are identified as samples taken at source water sampling points. Records are marked as raw if SOURCE_TYPE_CODE equals "RW" or SOURCE_TYPE_CODE is NULL but the water system facility type code equals "IG", "IN", "RS", "SP", "WL", or "CC". Records are marked as finished if SOURCE_TYPE_CODE equals "FN" or SOURCE_TYPE_CODE is NULL but the water facility type code equals "CW", "DS", "PF", "ST", "TM", "TP". Exhibit E.5 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 Exhibit E.1.

Exhibit E.4: Number of Six-Year Review 4 Data Records for TOC, Alkalinity, pH, DOC, SUVA, and UV-absorbance and Zip Filename(s)

Contaminant	Analyte ID	Number of States	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

Contaminant	Analyte ID	Number of States	Total Number of Sample Records	Total Number of Systems	Zip Filename
					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-254	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 – Paired Dataset					
Paired TOC-alkalinity record	N/A	33	92,666	1,192	SYR4_DBP_Related Parameters.zip

Exhibit E.5 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.
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.

Section 5: SYR 4 Data Records Posted for Microbial Contaminants, Microbial Related Parameters, and Disinfectant Residuals

Data for three microbial contaminants (total coliforms (TC), *Escherichia coli* (EC), and fecal coliform (FC)) were collected from 2012 to 2019 for SYR 4. The TC datasets are separated into individual files by each year of data collected because of the large volume of data collected. Unlike the TC records which are provided separately by year, the EC and FC are contained in one file. The EC dataset is one large file intended for use in Access or R. Systems are required under the Surface Water Treatment Rule to monitor for disinfectant residuals at the same time and locations as TC under TCR/RTCR. Most States submitted data from systems that included free and total residual chlorine results paired with TC records. However, some States provided the residual monitoring data in separate datafiles or did not submit that information under the SYR 4 ICR.

Exhibit E.6 provides a count of States, total number of sample records and systems for TC, EC, FC, and records of disinfectant residuals. Exhibit E.6 also shows that some States submitted chlorine residual monitoring results separately under different analyte codes (e.g., Chlorine (Analyte ID 0999), Residual Chlorine (Analyte ID 1012), and Free Residual Chlorine (Analyte ID 1013)). To maximize the number of paired total coliform and chlorine residual records, EPA took additional steps to add records from States reporting residual data records separately (see Section 5.5.2 of the main text for details on pairing and the analytes used). The "full" datafiles in Exhibit E.6 contain these paired records as well as records for systems with reported microbial indicator presence and absence but no associated disinfection residual information.

To assist the user, EPA produced the "paired" TC, EC, and FC datafiles (Exhibit E.6), which contain only the records for systems in the "full" versions of those datasets that include paired residual information. The "paired" datafiles were not analyzed under SYR 4 because of the ongoing considerations of potential revisions of the Surface Water Treatment Rules.

Note that the TC, EC, and FC datasets contain the monitoring records under TCR/RTCR for systems with all source water types. The HPC, *Giardia*, disinfectant residual, and paired TC/EC/FC disinfectant residual files contain the monitoring records under the SWTRs. See Exhibit E.1 for a description of field names.

Exhibit E.6: Number of Six-Year Review 4 Data Records for Microbial Contaminants, Microbial Related Parameters, and Disinfectant Residuals and Zip Filename(s)

Contaminant	Analyte ID	Number of States/ Entities with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Microbial Contaminants and Disinfectant Residual – Full Datasets					
Total Coliform (2012)	3100	54	2,349,687	102,423	SYR4_TC.zip

Contaminant	Analyte ID	Number of States/ Entities with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
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
E. coli (EC) In Raw Water[1]	3014	43	65,805	19,515	SYR4_EC_FC_HPC_Giardia.zip
E. coli (EC) In Distribution Systems[2]	3014	49	6,346,973	90,607	SYR4_EC_FC_HPC_Giardia.zip
E. coli (EC) In Unknown Sampling Location[3]	3014	54	762,585	24,486	SYR4_EC_FC_HPC_Giardia.zip
Fecal Coliform (FC)	3013	40	16,818	1,835	SYR4_EC_FC_HPC_Giardia.zip
Coliphage	3028	2	3	3	SYR4_EC_FC_HPC_Giardia.zip
Enterococci	3002	3	8	4	SYR4_EC_FC_HPC_Giardia.zip
Cryptosporidium	3015	29	19,542	740	SYR4_EC_FC_HPC_Giardia.zip
Heterotrophic Bacteria (HPC)	3001	16	135,081	595	SYR4_EC_FC_HPC_Giardia.zip
Giardia Lamblia	3008	15	4,628	229	SYR4_EC_FC_HPC_Giardia.zip
Legionella		0	0	0	N/A
Chlorine[4]	0999	19	6,100,133	4,438	SYR4_Disinfectant_ Residuals.zip
Total Chlorine[4]	1000	1	125,788	741	SYR4_Disinfectant_ Residuals.zip
Chloramine[4]	1006	9	78,664	198	SYR4_Disinfectant_ Residuals.zip
Residual Chlorine[4]	1012	4	179,599	572	SYR4_Disinfectant_ Residuals.zip
Free Residual Chlorine[4]	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 – Paired Datasets[5]					
E. coli (EC) with Associated Disinfectant Residuals	3014	49	3,079,032	28,091	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
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
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] Includes results with a sample type code of "TG" (i.e., triggered monitoring). Note that these record counts are subsets of the *E. coli* records included in the *E. coli* data set.
[2] Includes results not marked as triggered but had a sample point type of "DS", "FC", "FN", "LD", "MD", or "MR" or records with water facility type of "CC", "DS", "TP", or "TM" and sample point type of "WS" or null. Note that these record counts are subsets of the *E. coli* records included in the *E. coli* data set.
[3] Includes remaining *E. coli* results not identified as coming from raw water or the distribution system. Note that these record counts are subsets of the *E. coli* records included in the *E. coli* data set.
[4] Reported independently of the coliform sample results.
[5] Refer to Section 5.5.2 for more details on the paired disinfectant residual and total coliform records.

Section 6: SYR 4 Data Records Posted for Aircraft Drinking Water Rule (ADWR)

EPA downloaded compliance data from the Agency's Aircraft Reporting and Compliance System (ARCS) for the period from February 2011 to May 2021. This dataset includes aircraft compliance monitoring data for TC and EC for aircraft drinking water systems (Exhibit E.7). The Aircraft PWS Inventory file includes records for 8,627 unique aircraft drinking water systems. Details on the QA/QC procedure for this data can be found in Appendix D of this document. Note that the number of sample records presented below and included in the posted data reflect counts prior to the QA/QC procedures were applied for the SYR 4 analyses as presented in USEPA (2024b). After the QA/QC steps described in Appendix D are applied, there are 140,502 total coliform and 92,994 *E. coli* records remaining.

Exhibit E.7: Number of Aircraft Drinking Water Rule (ADWR) Data Records and Zip filename

Contaminant	Total Number of Sample Records	Total Number of Systems	Zip Filename
Aircraft PWS Sample by Air Carrier and Results			
Total Coliform	212,937	8,094	SYR4_ADWRCompliance Data.zip
E. coli[2]	93,011	7,091	SYR4_ADWR Compliance Data.zip

[1] The number of records presented here is greater than the number of rows of data downloaded from ARCS (70,979 at the time of download in support of the SYR 4 analysis) because it counts all samples within each row of data (i.e., Sample 1, Sample 2, and Sample 3). Note that Sample 3 is related to the ability to have third sample collected, which is not a requirement of ADWR and is not often used. Typically there is no data for Sample 3 fields.
[2] The count of *E. coli* records and systems is based on all *E. coli* samples listed as either "present" or "absent." It does not include samples listed as "not speciated" or "not analyzed."

Exhibit E.8, Data Dictionary Aircraft Drinking Water Rule (ADWR) Dataset, contains a list of the data elements, column names and a brief description of the data for each data element included in the ADWR data text files.

Exhibit E.8: Data Dictionary Aircraft Drinking Water Rule (ADWR) Dataset

Data Element	Column Name	Description
PWS Inventory		
Official FAA Corporate Name	FAA Corporate Name	The name of the air carrier or operator as registered with the FAA.
FAA Designator	FAA Designator	The four-character designator assigned to the air carrier by the FAA.
PWSID	PWS ID	The aircraft public water system identification number (PWS_ID) used by EPA to uniquely identify the aircraft public water system (PWS).
FAA Aircraft Registry No	FAARegistry No.	The number for the aircraft that is registered with the Federal Aviation Administration (FAA), commonly referred to as the N-number or tail number.
Aircraft Activity Status Code	Status	The activity status of the aircraft. It is selectable from the drop-down list. Permissible values are [Active] or [Inactive].
Routine Disinfection and Flushing Frequency	D&FFrequency	The frequency of routine disinfection and flushing scheduled for this aircraft.
Routine Sample Frequency	SamplingFrequency	The frequency of routine coliform sampling scheduled for this aircraft.
Aircraft Manufacturer	Manufacturer	The manufacturer of the aircraft.
Aircraft Model	Model	The manufacturer's model of the aircraft.
Seating Capacity	Seat Capacity	The number of passenger seats configured for the aircraft. It has a maximum value of 999.
Samples by Air Carrier		
Official FAA Corporate Name	FAA Corporate Name	The name of the air carrier or operator as registered with the FAA.

Data Element	Column Name	Description
FAA Aircraft Registry No	FAA Registry No.	The number for the aircraft that is registered with the Federal Aviation Administration (FAA), commonly referred to as the N-number or tail number.
PWSID	PWS ID	The aircraft public water system identification number (PWS_ID) used by EPA to uniquely identify the aircraft public water system (PWS).
Routine Sample Frequency	Routine Sample Frequency	The frequency of routine coliform sampling scheduled for this aircraft.
Sample Type	Sample Type	Indicates the type of individual sample: routine, repeat, follow-up, special.
Date and Time Collected	Sample Taken On	The date and time the sample was collected. When the galley and lavatory samples are collected on the same day, the date and time the first sample was collected is used. The required format is MM/DD/YYYY with time reported on a 24-hour clock as H:MI (e.g., 12/01/2014 15:00).
Date and Time Results Received	Samples Results On	The date and time the sample analysis results were received from the laboratory (e.g., phone message, USPS delivery date, office date and time stamp, e-mail receipt date and time). The required format is MM/DD/YYYY with time reported on a 24-hour clock as Hours:Minutes (e.g., 12/01/2014 15:00).
Sample Collection Location (Sample 1)	Sample1 Location	The location on the aircraft from where the first sample was collected. The options are [galley] or [lavatory].
Total Coliform Result (Sample 1)	Sample1 Total Coliform	The reported lab result that indicates the presence or absence of total coliform in the first sample analyzed. The drop-down list options are [Present] or [Absent].
E. coli Result (Sample 1)	Sample1 E.Coli	The lab analytical result that indicates the presence or absence of *E. coli* in the first sample analyzed. The drop-down list options are [Present] or [Absent] or [Did not speciate]. “Did not speciate” is used when the lab did not analyze a TC+ sample (or “present” sample result) for *E. coli*. Note: certified labs are required to analyze all TC+ samples for *E. coli*, but it is the carrier’s responsibility to make sure the lab completed the speciation.
Sample Collection Location (Sample 2)	Sample2 Location	The location on the aircraft from where the second sample was collected. The options are [galley] or [lavatory].
Total Coliform Result (Sample 2)	Sample2 Total Coliform	The reported lab result that indicates the presence or absence of total coliform in the second sample analyzed. The drop-down list options are [Present] or [Absent].
E. coli Result (Sample 2)	Sample2 *E. coli*	The lab analytical result that indicates the presence or absence of *E. coli* in the second sample analyzed. The drop-down list options are [Present] or [Absent] or [Did not speciate]. “Did not speciate” is used when the lab did not analyze a TC+ sample (or “present” sample result) for *E. coli*. Note: certified labs are required to analyze all TC+ samples for *E. coli*, but it is the carrier’s responsibility to make sure the lab completed the speciation.
Sample Collection Location (Sample 3)	Sample3 Location	The location on the aircraft from where the third sample was collected. The options are [galley] or [lavatory].
Total Coliform Result (Sample 3)	Sample3 Total Coliform	The reported lab result that indicates the presence or absence of total coliform in the third sample analyzed. The drop-down list options are [Present] or [Absent].
E. coli Result (Sample 3)	Sample3 *E. coli*	The lab analytical result that indicates the presence or absence of *E. coli* in the third sample analyzed. The drop-down list options are [Present] or [Absent] or [Did not speciate]. “Did not speciate” is used when the lab did not analyze a TC+ sample (or “present” sample result) for *E. coli*. Note: certified labs are required to analyze all TC+ samples for *E. coli*, but it is the carrier’s responsibility to make sure the lab completed the speciation.

Section 7: Additional Data Collected under SYR 4 ICR

Additional data relating to certain microbial rules were collected under the SYR 4 ICR request, including calculated compliance values and corrective actions information. Note that these data did not undergo the same quality assurance evaluations as the rest of the data.

Calculated Compliance Values

Exhibit E.9 provides a summary of the data elements included in the calculative compliance values related to *Cryptosporidium* binning information from SYR 4 ICR database. Exhibit E.10 provides a summary of the systems and states that provided SYR4 *Cryptosporidium* binning data.

Exhibit E.9: Data Dictionary of *Cryptosporidium* Binning Information Included as part of the Calculated Compliance Values Table (Filename: SYR4_CryptoBinning)

Data Element	Column Name	Description
Contaminant Identification Code	ANALYTE_CODE	4-digit Safe Drinking Water Information System (SDWIS) contaminant identification number for which the sample is being analyzed.
Contaminant Name	ANALYTE_NAME	Common name of contaminant for which the sample is being analyzed.
Public Water System Identification Number (PWSID)	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.
Facility Identification Code	WATER_FACILITY_ID	A unique identifier for each water system facility.
Compliance Period Begin Date	CP_PRD_BEGIN_ DT	Compliance Period Begin Date.
Compliance Period End Date	CP_PRD_END_DT	Compliance Period End Date.
Bin Number	BIN_NUMBER	The BIN assignment for the period of time covered by the average.

Exhibit E.10: Six-Year Review 4 Data Summary for Calculated Compliance Values Related to *Cryptosporidium* Binning Information

Number of States	Total Number of Sample Records	Total Number of Systems	Zip Filename
23	27,812	486	SYR4_CryptoBinning.zip

Corrective Actions

Exhibit E.11 provides a summary of the data elements included in the corrective actions table within the SYR 4 ICR database. Exhibit E.12 provides a summary of the corrective action data collected as part of SYR. Note, however, that EPA did not evaluate the specific types of corrective actions (e.g., those related to sanitary surveys) as part of SYR 4.

Exhibit E.11: Corrective Actions Data Dictionary (Filename: SYR4_CorrectiveActions)

Data Element	Column Name	Description
Corrective Action ID	CORACT_ID	Unique identifier for each corrective action.
Public Water System Identification Number (PWSID)	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.
State Code	STATE_CODE	State in which the system is located using the State's two letter abbreviation.
Date Issue Identified	DATE_ISSUE_IDENTIFIED	Date the corrective action was identified.
Schedule Type	SCHEDULE_TYPE	Type of schedule for the corrective action.
Schedule Description	SCHEDULE_DESCRIPTION	Schedule for the corrective action.
Corrective Action Category Code	CORACT_CAT_CODE	Category code for the corrective action.
Corrective Action Name	CORACT_NAME	Name of the corrective action.
Due Date	DUE_DATE	Due date for the required corrective action.
Achieved Date	ACHIEVED_DATE	The date that the water system achieved the corrective action required.

Exhibit E.12: Six-Year Review 4 Data Summary for Corrective Actions

Number of States	Total Number of Sample Records	Total Number of Systems	Zip Filename
41	69,821	15,984	SYR4_Corrective_Actions.zip

Section 8: Treatment Data

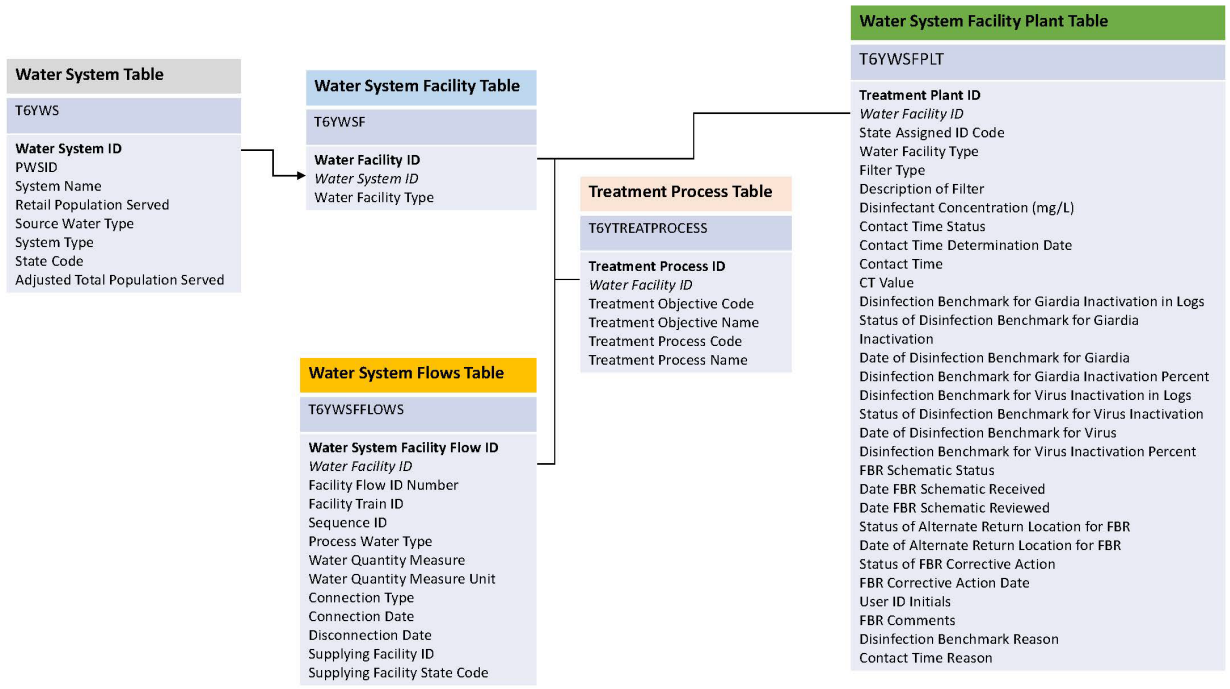
Exhibits E.13 and E.14 provide a comprehensive summary of the data elements included in the treatment information within the SYR 4 ICR database. EPA has posted these data online; however, it is important to note that the treatment information did not undergo the same quality assurance evaluations as the SYR 4 occurrence data. Exhibit E.13 identifies the data elements used in the treatment information tables and a description of each data element. However, the majority of these data elements are not populated in the SYR 4 ICR dataset. Exhibit E.14 represents the database relationships between tables in the SYR 4 ICR treatment database. This diagram shows how the treatment tables relate to one another. Bolded field names are primary keys, or unique fields, designated to identify all table records. Primary keys contain a unique number for each row of data. Italicized field names are foreign keys that serve as the link (i.e., connection) between two or more related tables. Relationships between key fields in different tables are illustrated by the lines connecting the tables.

Exhibit E.13: Treatment Data Dictionary (Filename: SYR4_Treatment)

Data Element	Description
Water system facility plant table (T6YWSFPLT)	
Treatment Plant ID	Unique identifier for each treatment plant water system facility record.
Water Facility ID	Identifier that relates each record to the unique record in the T6YWSF table.
State Assigned ID Code	A State-assigned value which identifies the treatment plant water system facility.
Water Facility Type	The value extracted from SDWIS/State will be “TP” (treatment plant). The values from non-SDWIS States include “TM” (transmission manifold) and “ST” (storage).
Filter Type	Unfiltered (UF), Conventional Filtration (CF), Direct Filtration (DF), Diatomaceous Earth (DE), Other (OT), and other permitted values that the System Administrator may add.
Description of Filter	A description of the filter.
Disinfectant Concentration (mg/L)	Disinfectant Concentration in mg/L.
Contact Time Status	Contact Time Status. Permitted values are: RQD – Required; NRQD - Not Required; REQT – Requested; RECV – Received; URVW - Under Review; RVWD – Reviewed; APVD – Approved; DTMD – Determined; DENY – Denied; RESB – Resubmitted.
Contact Time Determination Date	Date the Contact Time was determined.
Contact Time	Contact Time in minutes – the number of minutes the water was in contact with the disinfectant in order to be properly disinfected. The range of values is 0001 to 2400.
CT Value	CT value in mg x min/liter.
Disinfection Benchmark for *Giardia* Inactivation in Logs	The disinfection profile benchmark for *Giardia* inactivation in Logs.
Status of Disinfection Benchmark for *Giardia* Inactivation	The status of the disinfection profile benchmark for *Giardia* inactivation. See CONTACT_TIME_STAT for permitted values and description.
Date of Disinfection Benchmark for *Giardia*	The date the disinfection virus benchmark was determined.
Disinfection Benchmark for *Giardia* Inactivation Percent	The disinfection profile benchmark for *Giardia* inactivation percent.
Disinfection Benchmark for Virus Inactivation in Logs	The disinfection profile benchmark for virus inactivation in Logs.
Status of Disinfection Benchmark for Virus Inactivation	The status of the disinfection profile benchmark for Virus inactivation. See CONTACT_TIME_STAT for permitted values and description
Date of Disinfection Benchmark for Virus	The date the disinfection virus benchmark was determined.
Disinfection Benchmark for Virus Inactivation Percent	The disinfection profile benchmark for virus inactivation percent.
FBR Schematic Status	Under the Filter Backwash Rule, a water system is required to submit a schematic of this treatment plant to the primacy agency for review to demonstrate the percentage of filter backwash that is returned to the treatment plant influent. See CONTACT_TIME_STAT for permitted values and description.
Date FBR Schematic Received	Date primacy agency received treatment plant schematic to demonstrate the percentage of filter backwash that is returned to the treatment plant influent.
Date FBR Schematic Reviewed	Date primacy agency completes review of treatment plant schematic and determines the percentage of filter backwash that is returned to the treatment plant influent.
Status of Alternate Return Location for FBR	The status of a request from the water system to request an alternate location for return of the filter backwash
Date of Alternate Return Location for FBR	The date that the water system requested an alternate location for return of the filter backwash.
Status of FBR Corrective Action	The status of corrective action by the water system as required by the primacy agency after review of the schematic of the filter backwash flow in the treatment plant.

Data Element	Description
FBR Corrective Action Date	The date that the water system achieved the corrective action required for the filter backwash.
User ID Initials	The User ID of the person who created this record.
FBR Comments	A memo field into which a user may enter comments about the Filter Backwash Recycling Rule.
Disinfection Benchmark Reason	Text description associated with the Disinfection Benchmark Reason.
Contact Time Reason	Text description associated with the Contact Time.
Treatment process table (T6YTREATPROCESS)	
Treatment Process ID	Unique identifier for each treatment record.
Water Facility ID	Identifier that relates each record to the unique record in the T6YWSF table.
Treatment Objective Code	A coded value that categorizes the treatment objective.
Treatment Objective Name	The name of the treatment objective.
Treatment Process Code	A coded value that categorizes the treatment process.
Treatment Process Name	The name of the treatment process.
Water system flows table (T6YWSFFLOWS)	
Water System Facility Flow ID	Unique identifier for each water system facility flow record.
Water Facility ID	Identifier that relates each record to the unique record in the T6YWSF table.
Facility Flow ID Number	Identifier for each water system facility flow entry that is unique when combined with T6YWSFT6YWSF_ID.
Facility Train ID	This attribute identifies the water system facilities that are part of the same flow.
Sequence ID	This attribute identifies the order of the water system facilities in a specific flow.
Process Water Type	A system administrator controlled code of the type of water flowing between the facilities.
Water Quantity Measure	A value that represents the number of gallons of water purchased.
Water Quantity Measure Unit	A coded value which specifies the unit of measurement for the quantity of water purchased.
Connection Type	Categorizes the type of connection between the water system facilities.
Connection Date	The date of the connection of the water system facility to another water system facility.
Disconnection Date	The date of the disconnection of the water system facility from another water system facility.
Supplying Facility ID	Identifier for each supplying water system facility that is unique when combined with TINWSF0ST_CODE.
Supplying Facility State Code	Two-digit code that identifies the State that submitted data for the facility

Exhibit E.14: Treatment Data Diagram



Section 9: SYR 4 Data Considerations

The SYR 4 ICR data has undergone appropriate quality assurance evaluation and enough States provided compliance monitoring data and treatment technique information to be representative for national-scale analyses. EPA used the data in analytical activities informing decisions for SYR 4. The data include sufficient information for users to be able to reproduce the SYR 4 analyses. There are a few limitations of the final SYR 4 ICR dataset that should also be acknowledged. There may be different levels of completeness for different contaminants within the dataset. In some cases, the number of records per State ranged from less than one hundred records up to more than a million records for a given contaminant. States might not have submitted data for certain contaminants if they have monitoring waivers for the contaminant. States may grant waivers to PWSs to reduce monitoring frequencies, and it is possible that no samples were collected by systems during the SYR 4 period of review. Other States may have submitted data for these contaminants under the ICR; however, the data were not in a format compatible with the SYR 4 ICR dataset. Furthermore, there were four States and three additional tribes or territories whose data are missing entirely from the analysis. A thorough QA/QC process was undertaken to evaluate these SYR 4 ICR data used for analyses. However, it is possible that data entry errors may still exist in the final SYR 4 ICR dataset. The QA/QC review focused only on the data elements essential for analysis as part of SYR 4. For a complete discussion of the SYR 4 ICR dataset, including a description of the quality assurance/quality control review, refer to the main text of this document and USEPA (2024a). For more detailed information on the microbial contaminants' occurrence analysis, refer to USEPA (2024b).

Section 10: Instructions on Importing SYR 4 Datasets

These text files are tab delimited and have no text qualifier. Field names are included in the first row of each file. The complete SYR 4 ICR dataset is too large to be imported into Excel as well as certain individual files, these files include individual years of TC and EC files, free chlorine, total chlorine and paired datasets of TC/EC/FC with residual disinfectant. The data are available for download for each parameter and should be imported into a data management system that supports large datasets for analysis.

10A: Downloading Data Files (Note that instructions may vary depending on the version and software used to import data.)

1. Begin by reviewing the SYR 4 ICR Dataset Summary (Exhibit E.2) and in particular note the table of Data Field Names and Definitions (Exhibit E.1).

2. Access the SYR 4 ICR data by going to the Six-Year Review homepage. Click on the link for "Six-Year Review 4."

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

10B: 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 microbial and disinfection byproduct 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.

10C: 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 the format, 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.

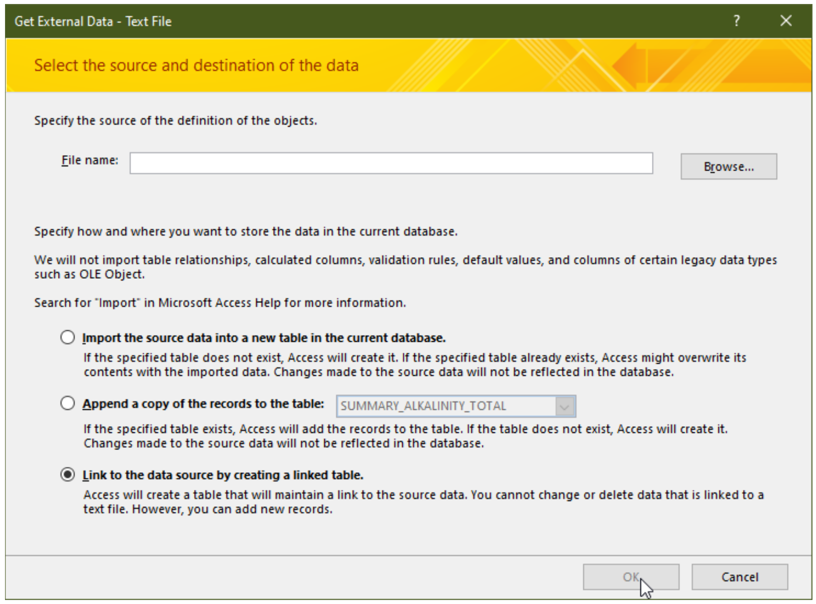
10D: Importing Data into Microsoft Access

1. Open a blank database in Microsoft Access.

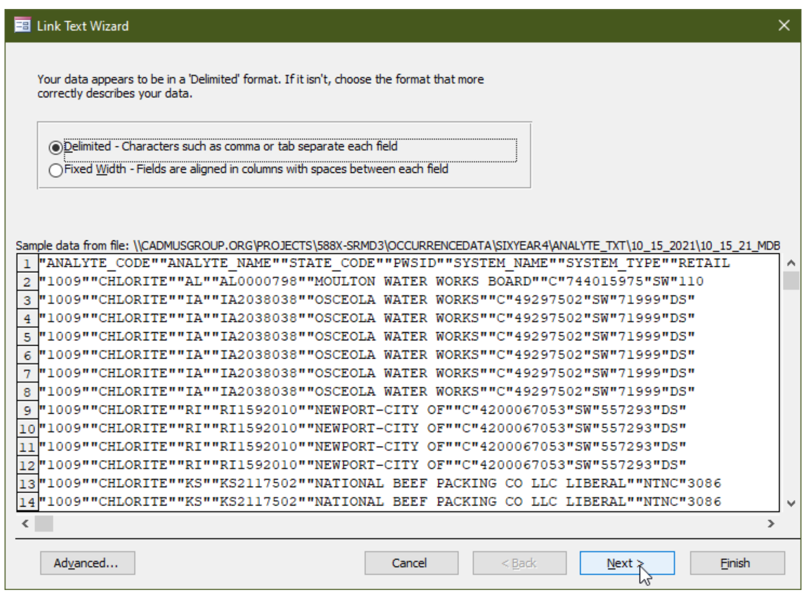
2. In the database, select External Data among the tabs at the top of the page.

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

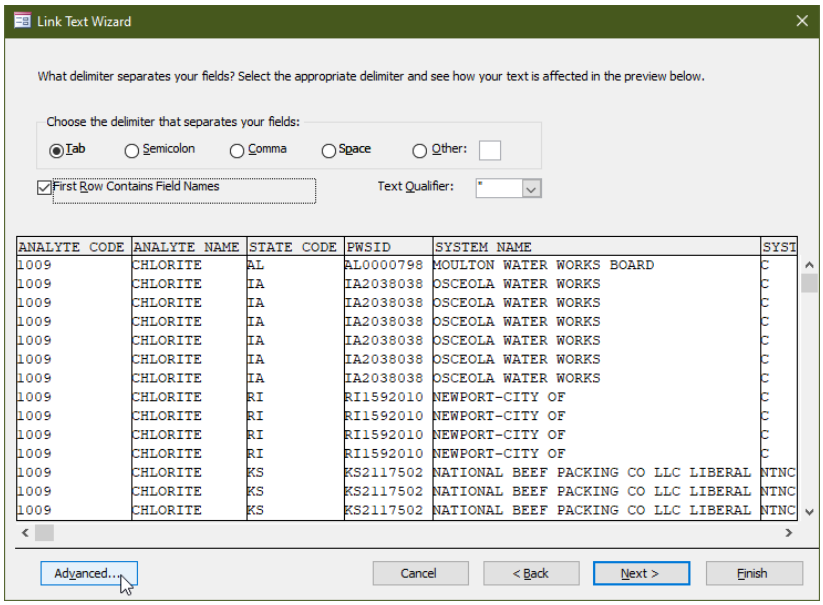
4. You will be prompted to select a text file. Locate the text files you extracted in Step 4, and with the following options: “import the source data into a new table in the current database”, or “Link to the data source by creating a linked table”. You can choose either method, but note that linking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



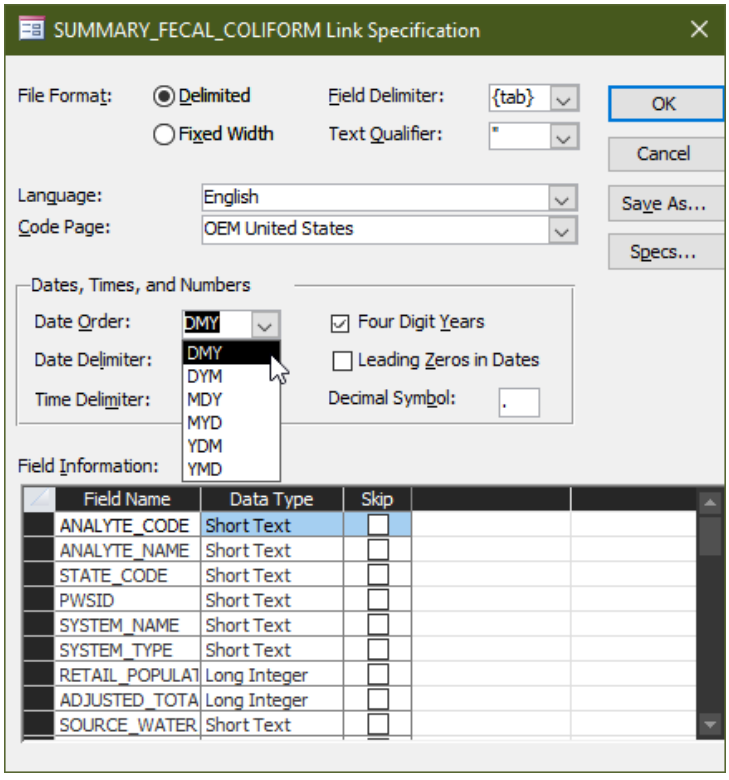
5. The Link (or Import) Text Wizard will appear. The default settings will be displayed and should have Delimited selected as the data format. Select Next>.



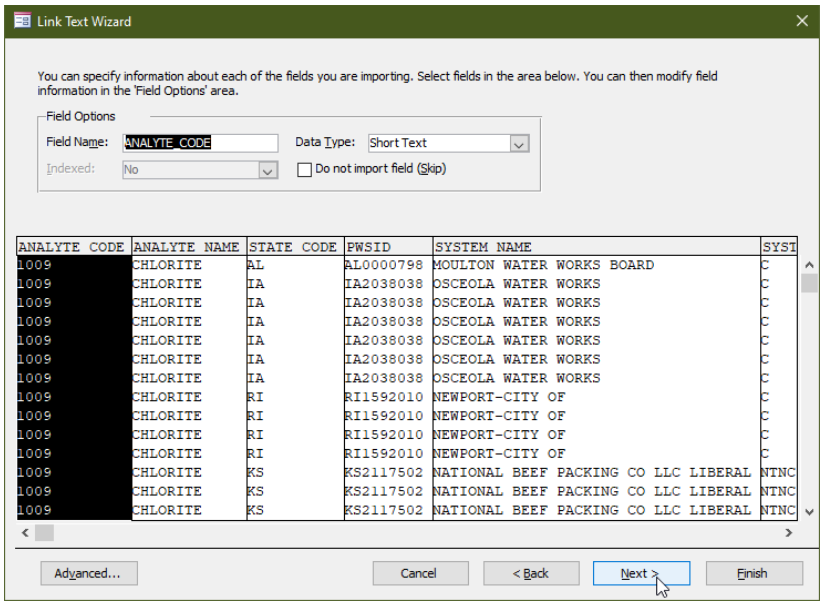
6. Default settings will display next and should have “Tab” selected as the delimiter. Select the checkmark box next to “First Row Contains Field Names.” Next, click “Advanced…”.



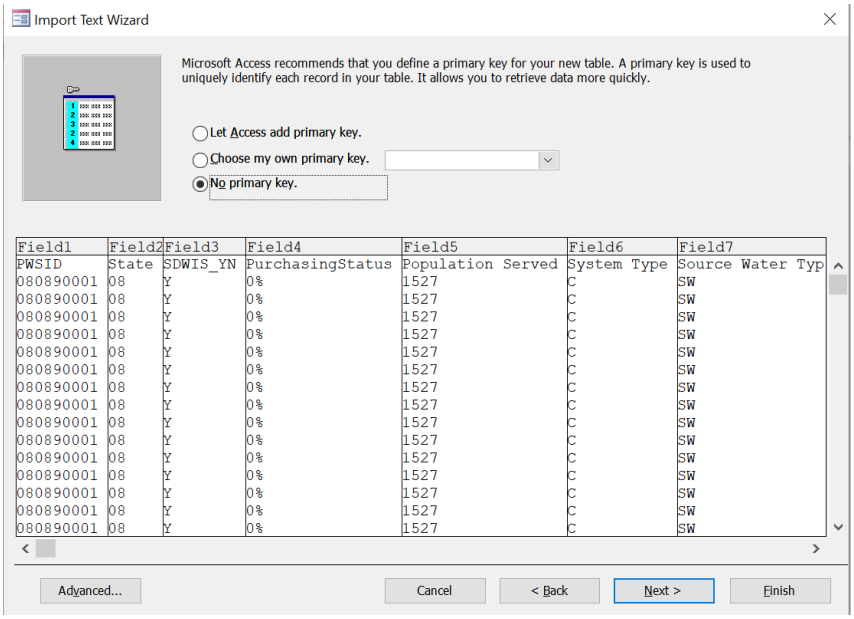
7. The Link (or Import) Specification window will appear. In the Dates, Times, and Numbers section, set the Date Order value to "DMY."



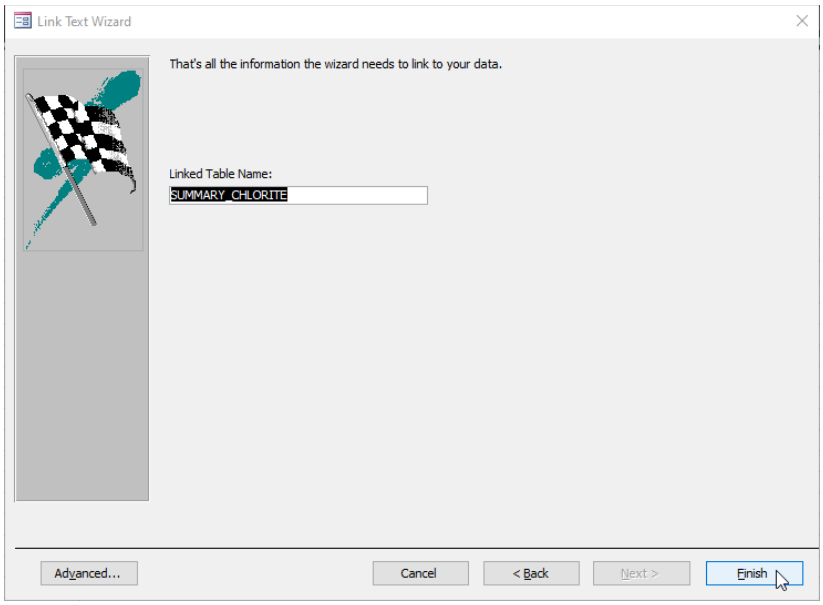
8. On the screen that follows, keep the default settings shown below and click Next>.



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

References

United States Environmental Protection Agency (USEPA). 2016. Six-Year Review 3 Technical Support Document for Disinfectants/Disinfection Byproducts Rules. EPA-810-R-16-012. December 2016.

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.