The Data Management and Quality Assurance/Quality Control Process for the Third Six-Year Review Information Collection Rule Dataset

Office of Water (4607M)
EPA-810-R-16-015
December 2016

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, at least once every six years, review and revise, as appropriate, each National Primary Drinking Water Regulation (NPDWR)." The NPDWRs are often referred to as the national drinking water contaminant regulations or drinking water standards. The purpose of the review, called the Six-Year Review, 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.

This report describes how the compliance monitoring data for EPA's third Six-Year Review 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 Third Six-Year Review to establish a high quality, national compliance monitoring dataset consisting of data from 54 states/primacy agencies (46 states plus Washington, D.C. and the tribal data). The compliance monitoring data for these 54 states/primacy agencies comprise almost 13 million analytical records from approximately 139,000 public water systems (PWSs), which serve approximately 290 million people nationally. This dataset, the Third Six-Year Review (SYR3) ICR Dataset for the third Six-Year Review (or "SYR3 ICR Dataset"), is the largest and most comprehensive compliance monitoring dataset ever compiled and analyzed by EPA's Drinking Water Program.

Information regarding the acquisition, storage and management of the SYR3 ICR data is presented in Section 2 through 4 of this report. Detailed descriptions of the QA/QC evaluations and data preparation for analyses are presented in Section 5 and Section 6, respectively. Additional technical information related to the SYR3 ICR database is presented in the appendices to this report.

For the national contaminant occurrence assessments for the chemical phase rules and radionuclides rules conducted in support of EPA's third Six-Year Review of NPDWRs, refer to the USEPA (2016a) report entitled *The Analysis of Regulated Contaminant Occurrence Data from Public Water Systems in Support of the Third 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 (2016b). For more detailed information on the occurrence analysis of contaminants/parameters regulated under the D/DBPRs, refer to USEPA (2016c). The final SYR3 ICR datasets are posted online at: https://www.epa.gov/dwsixyearreview.

Table of Contents

Appendices

Exhibits

Acronyms

CAS	Chemical Abstracts Service
CHEMID	Four Digit SDWIS Code
CO	Confirmation
cVOC	Carcinogenic Volatile Organic Chemical
CWS	Community Water System
DBCP	1,2-Dibromo-3-chloropropane
DBP	Disinfection Byproduct
DBPR	Disinfection Byproduct Rule
D/DBPR	Disinfectants and Disinfection Byproducts Rules
DEHA	Di(2-ethylhexyl)adipate
DEHP	Di(2-ethylhexyl)phthalate
EDB	Ethylene dibromide
eDWR	Electronic Drinking Water Report
EPA	Environmental Protection Agency (United States)
FBRR	Filter Backwash Recycling Rule
FTP	File Transfer Protocol
GAC	Granular Activated Carbon
GW	Ground Water
GWR	Ground Water Rule
GWUDI	Ground Water Under Direct Influence (of Surface Water)
HAA5	Haloacetic Acids
HPC	Heterotrophic Plate Count
IESWTR	Interim Enhanced Surface Water Rule
ICR	Information Collection Request
IOC	Inorganic Chemical
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
MDL	Method Detection Limit
MFL	Million Fibers per Liter
mg/L	Milligrams per Liter
MOR	Monthly Operating Report
mrem/yr	Millirem per year
MR	Maximum Residence
MRL	Minimum Reporting Level
MS	Microsoft
NCOD	National Contaminant Occurrence Database
ND	Non-detect or Non-detection
NPDWR	National Primary Drinking Water Regulation
NTNCWS	Non-Transient Non-Community Water System
OMB	Office of Management and Budget
PCBs	Polychlorinated Biphenyls
pCi/L	Picocuries per Liter
PQAPP	Programmatic Quality Assurance Project Plan

PWS	Public Water System
PWSID	Public Water System Identification Number
QA	Quality Assurance
QC	Quality Control
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 Chemical
SW	Surface Water
SWP	Purchased Surface Water
SWTR	Surface Water Treatment Rule
SYR3	Third Six-Year Review
TCR	Total Coliform Rule
TNCWS	Transient Non-Community Water System
TOC	Total Organic Carbon
TTHM	Total Trihalomethane
USEPA	United States Environmental Protection Agency
µg/L	Micrograms per Liter
VOC	Volatile Organic Chemical

1 Introduction

This document describes how the compliance monitoring data for the third Six-Year Review were obtained, evaluated, and formatted, where necessary, to enable national contaminant occurrence estimates in support of EPA's third Six-Year Review (SYR3) 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 below.

The 1996 Amendments to the Safe Drinking Water Act (SDWA) require that the Environmental Protection Agency (EPA) "shall, at least once every six years, review and revise, as appropriate, each National Primary Drinking Water Regulation (NPDWR)." The NPDWRs are often referred to as the national drinking water contaminant regulations or drinking water standards. The purpose of the review, called the Six-Year Review, 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.

National contaminant occurrence assessments were conducted in support of EPA's SYR3, using data from National Compliance Monitoring ICR Dataset for the third Six-Year Review (or "SYR3 ICR dataset"). These compliance monitoring data were provided to EPA by the states via the Information Collection Request (ICR) process. The report *The Analysis of Regulated Contaminant Occurrence Data from Public Water Systems in Support of the Third Six-Year Review of National Primary Drinking Water Regulations: Chemical Phase Rules and Radionuclides Rules* (USEPA, 2016a) 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 SYR3. Included in that report are detailed descriptions of the national contaminant compliance monitoring dataset compiled and the statistical analytical methods employed (using the national dataset) to generate national estimates of regulated contaminant occurrence in public drinking water systems.

The NPDWRs for the microbial contaminant regulations and disinfectants/disinfection byproducts rules (D/DBPRs) were also included under SYR3. For more detailed information on the microbial contaminants' occurrence analysis, refer to USEPA (2016b). For more detailed information on the occurrence analysis of contaminants regulated under the D/DBPRs, refer to USEPA (2016c).

SDWA compliance monitoring data for some of the regulated contaminants are assessed separately under other regulatory actions and were not evaluated under the SYR3. Data for lead and copper, as well as carcinogenic Volatile Organic Compound (cVOCs), were not subject to a detailed review because of recently completed, ongoing or pending regulatory actions. In addition, compliance monitoring data was not collected for epichlorohydrin and acrylamide because there are currently no acceptable laboratory analytical methods for detecting these contaminants in drinking water. Furthermore, no states submitted SYR3 data for these two contaminants. For the technical analysis for these two contaminants, see *Support Document for*

Third Six Year Review of Drinking Water Regulations for Acrylamide and Epichlorohydrin (U.S. EPA, 2016d).

The SYR3 ICR data were received from the states and primacy agencies in a variety of formats and data structures, and 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 and other primacy agencies, and 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 SYR3 contaminant occurrence evaluations. Specifically, this document describes the compliance monitoring data 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 and D/DBP data are described in USEPA (2016b) and USEPA (2016c), respectively.

2 Data Acquisition

Compliance monitoring data provide information critical to Six-Year 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 PWSs and their violations of EPA's drinking water regulations. However, SDWIS/Fed does not receive or store compliance monitoring data (called parametric data), which includes non-detections (NDs) 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 used to support national occurrence assessments for SYR3, 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 by the states, see the third Six-Year Review ICR renewal (75 FR 6023, USEPA, 2010).

Similar to the second Six-Year Review, EPA contacted each primacy agency via a letter for SYR3 to request the voluntary submission of their compliance monitoring data for regulated chemical and radiological contaminants that were collected between January 2006 and December 2011. See Appendix A for the compliance monitoring data request letter. In addition, for SYR3 EPA requested compliance monitoring and parametric data for the Ground Water Rule (GWR); Surface Water Treatment Rules (SWTR); the Interim Enhanced Surface Water Treatment Rule (IESWTR); the Long-Term 1 Enhanced Surface Water Treatment Rule (LT1ESWTR); the Long-Term 2 Enhanced Surface Water Treatment Rule (LT2ESWTR); Disinfectants and Disinfection Byproducts Rules (D/DBPRs); and the Filter Backwash Recycling Rule (FBRR).

EPA requested only information stored electronically as structured data (no paper records) and that represented routine compliance monitoring and treatment technique information. Exhibit 2.1 shows the regulated contaminants for which EPA requested data, and Exhibit 2.2 shows the requested data elements for each sample result. See Appendix B: Crosswalk of Data Elements Requested for SYR3 ICR and the SDWIS Data Element Names for a cross-walk table between the data elements requested and the actual data element names as they appear in SDWIS. Note that there were cases where EPA did not receive data on all of the data elements and/or analytes requested. Furthermore, there were situations (such as with coliphage) where the only data received did not pass QA/QC and thus were not evaluated further.

Exhibit 2.1: List of Contaminants/Parameters Identified in SYR3 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)
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	
Microbiological Contaminants and Surface Water Treatment Rules (SWTRs)[1]		
Total coliforms	Fecal coliforms	*Escherichia coli (E. coli)*
Chlorine	*Cryptosporidium*	Heterotrophic Plate Count (HPC)
Chloramines	*Giardia lamblia*	

Disinfectants and Disinfection Byproducts Rules (D/DBPRs)[2]		
Total Trihalomethanes (TTHMs):	Haloacetic Acids (HAA5):	Bromate
Chloroform	Monochloroacetic acid	Chlorite
Bromodichloromethane	Dichloroacetic acid	Chlorine
Dibromochloromethane	Trichloroacetic acid	Chloramines
Bromoform	Monobromoacetic acid	Chlorine dioxide
	Dibromoacetic acid	
Ground Water Rule (GWR)		
Escherichia coli (E. coli)	Enterococci	Coliphage
Filter Backwash Recycling Rule (FBRR)		
No specific occurrence data collected; see Exhibit 2.2 for data elements for FBRR.		

Source: Attachment A to letter EPA sent contacting each Primacy Agency 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.
[1] Including: Surface Water Treatment Rule (June 1989); Interim Enhanced SWTR (December 1998); Long-Term 1 Enhanced SWTR (January 2002); and Long-Term 2 Enhanced SWTR (January 2006).
[2] Including both Disinfectants/Disinfection Byproducts Rules: Stage 1 (December 1998) and Stage 2 (January 2006).

Exhibit 2.2: Data Elements Requested by EPA for the Third 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 two-character postal state abbreviation or Region code; the remaining seven 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 purchased ground water; or • Surface water or purchased surface water; or • Ground water under the direct influence of surface water (GWUDI) or purchased GWUDI. (Note: Some states may not distinguish GWUDI from surface water sources. In those states, a GWUDI source should be reported as surface water.)
Sanitary Survey Information	Site visit information for Total Coliform Rule (TCR), Ground Water Rule (GWR), and Surface Water Treatment Rules (SWTRs), including: site visit type, date completed, associated deficiencies identified, and corrective actions taken.

Data Category	Description
Treatment Information	
Water System Facility	System facility data including: treatment plant identification number, treatment plant information, treatment unit process/objectives, facility flow and treatment train (train or flow of water through treatment units within the treatment plant).
Filtration Type	Information relating to system filtration, including filtration status and 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 include: coagulant/coagulant aid type and dose, disinfectant concentration (amounts, types, primary and secondary types of disinfection, disinfection profile/bench mark data), log of viral inactivation/removal, contact time, contact value, pH, and 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 allows for 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 was 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, four-digit SDWIS contaminant identification number or Chemical Abstracts Service (CAS) Registry Number for which the sample is being analyzed.
Sample Analytical Result - Sign	Sign indicating 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.” (Not required for TCR data)
Sample Analytical Result - Value	Numeric (decimal) analytical result, or the MRL if the analytical result is less than the contaminant’s MRL. (For the TCR, results will indicate presence/absence)
Sample Analytical Result - Unit of Measure	Unit of measurement for the analytical results reported (usually expressed in µg/L or mg/L for chemicals, or pCi/L or mrem/yr for radiological contaminants). (Not required for TCR data)
Sample Analytical Method Number	EPA identification number of the analytical method used to analyze the sample for a given contaminant.

Data Category	Description
Minimum Reporting Level (MRL) - Value	MRL refers to the lowest concentration of an analyte that may be reported. (Not required for TCR data)
MRL - Unit of Measure	Unit of measure to express the concentration value of a contaminant's MRL. (Not required for TCR 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 (MOR)) summary records, alternative compliance criteria.
Sample Summary Reports	Sample summaries for Disinfectants and Disinfection Byproducts Rules (D/DBPRs), SWTRs, TCR, and 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, Heterotrophic Plate Count (HPC), etc.]

Source: Attachment A to letter EPA sent contacting each Primacy Agency 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.
[1] These are the data elements requested in the SYR3 ICR. Note that the "Data Category" and "Description" Columns were intentionally descriptive rather than prescriptive. This allowed the states that do not SDWIS/State flexibility to provide as much information as possible. EPA accepted all data "as is" without prescribing structure or format.

About 75 percent of all states currently store and manage at least portions of their compliance monitoring data in the Safe Drinking Water Information System/State Version (SDWIS/State). EPA developed SDWIS/State in collaboration with state primacy agencies to manage drinking water information and provide a common structure for the development of reusable components and shared applications. The SDWIS/State structure is flexible enough to support the most complex primacy agency program implementation while maintaining a common core of data elements required for reporting to SDWIS/Fed. In an attempt to make the SYR3 data submittal process as easy for states as possible, EPA developed a SDWIS/State Extract Tool, which runs a customized query to pull the requested data from a SDWIS/State database. States that used SDWIS/State for data storage and management and were interested in using the SDWIS/State Extract Tool sent an email to EPA to request instructions and a link to download the extraction tool. Nearly all of the states using SDWIS/State that submitted data to EPA for SYR3 used the SDWIS/State Extract Tool to extract and compile the EPA-requested compliance monitoring data.

SDWIS/State supports the eDWR (Electronic Drinking Water Report) XML Schema used by laboratories throughout the nation to electronically report sample analytical results as structured data to SDWIS/State. As a result, primacy agencies receive high quality data from laboratories that is batch-processed into SDWIS/State rather than manually entered. Consequently, states have a substantial amount of high-quality structured data available in SDWIS/State. In all, 46 states and eight other primacy agencies provided compliance monitoring data that included parametric records. The four states that did not provide data were Colorado, Delaware, Georgia, and Mississippi. Exhibit 2.3 lists the states that did and did not use the SDWIS/State Extract Tool. 33 states and three tribes used the SDWIS/State Extract Tool to extract all or some of their chemical data; therefore, those datasets were all submitted in a similar format. The 18 states/entities not using SDWIS/State submitted their compliance monitoring data "as is," resulting in a variety of formats, including dBase, Microsoft (MS) Access, comma-delimited,

tab-delimited, text and Excel. With the exception of one state that shipped a CD/DVD of their data, all states submitted their data over the Internet via file transfer protocol (FTP).

Exhibit 2.3: Summary of States and Other Entities that Provided Compliance Monitoring Data for SYR3

	State/Entity Name		
States/Tribes that DID use the SDWIS/State Extract Tool	Alabama Alaska Arizona Arkansas Connecticut Idaho Illinois Indiana Iowa Kansas Kentucky Louisiana	Maine Missouri Montana Nebraska Nevada New Jersey[1] New Mexico New York North Carolina[1] North Dakota Ohio Oklahoma	Oregon Region 4 tribes Region 5 tribes Region 8 tribes Rhode Island South Carolina Texas[1] Utah Vermont Virginia West Virginia Wyoming
States/Tribes that DID NOT use the SDWIS/State Extract Tool	American Samoa California Florida Hawaii Maryland Massachusetts	Michigan Minnesota Navajo Nation New Hampshire Pennsylvania Region 1 tribes	Region 9 tribes South Dakota Tennessee Washington Washington, D.C. Wisconsin

[1] North Carolina, New Jersey, and Texas submitted their SDWIS/State data in an Oracle database. EPA applied the SDWIS/State Extract Tool to their databases to extract and compile the compliance monitoring data requested by EPA for SYR3.

3 Data Storage

EPA created an enterprise-level database (the SYR3 ICR SQL database) designed similarly to SDWIS/State to house the data that primacy agencies sent in response to the SYR3 ICR data request. The SYR3 ICR database is a Microsoft SQL Server relational database which consists of tables, views, 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 SYR3 SQL database contain a unique record identifier (also known as 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 SYR3 SQL database was designed to ensure information was not duplicated between tables and to maintain the logical relationships inherent to the data.

Exhibit 3.1 presents a description of the tables included in the SYR3 ICR SQL database. The database includes 17 "primary" tables (i.e., those listed in the table below with the prefix "tbl"). The primary tables include SDWIS data elements, codes and the compliance monitoring data. Three additional tables related 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 the table below 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 SYR3 SQL Database.

Exhibit 3.1: Description of Tables Included in SYR3 ICR SQL Database

Table Name	Brief Description	Description of Contents of Table
tblSixYrWs	Water system (Ws) table	Inventory information: PWSID, source water type, system type, population, etc.
tblSixYrWsf	Water system facility (Wsf) table	Facility identification information: facility ID, facility type, etc.
tblSixYrSpt	Sample point (Spt) table	Sample point identification information: sample point type, source type, etc.
tblAnalyte	Analyte table	Analyte identification information: contaminant name, 4-digit chemical IDs, etc.
tblSixYrSar	Sample analytical result (Sar) table	Monitoring records: sample date, sample type code, analyte, concentration, reporting level, method, etc.
tblSixYrDbpSum	Disinfectant By-Product summaries table	Summary used to enter sampling requirements and collection information in support of the SWTR/IESWTR and DBP rules.
tblSixYrFanls	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.
tblSixYrSampSum	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.

Table Name	Brief Description	Description of Contents of Table
tblSixYrSaniSur	Sanitary survey table	Includes information on sanitary surveys, such as the date of the site visit, if there were any deficiencies, etc.
tblSixYrSanSurvDef	Sanitary survey deficiency table	Includes information on sanitary survey deficiencies, such as the type of deficiency, the severity, etc.
tblSixYrSSCorAct	Sanitary survey corrective actions table	Includes information on sanitary survey corrective actions.
tblSixYrWsfPlt	Treatment plant water system facilities table	Includes information on treatment plant facilities.
tblTreatProcess	Treatments associated to treatment plants table	Includes information pertaining to the treatment processes and objectives.
tblWsfFlows	Water system facility flows table	Includes information on the relationship or connection between the different water system facilities of a water system.
tblWsfInd	Water system facility indicators table	Includes information on the recording of an indicator for a Water System Facility.
tblWsInd	Water system indicators table	Includes information on the recording of an indicator for a Water System.
tblWsPurch	Water system buyers and sellers	Includes information on the purchase of water between water systems.
lkp_SixYrSar_Transaction_QAFlag	Transaction QA Flag – Lookup Table	Includes lookup information on the QA flag codes and definitions related to the flagged Sample Analytical Results in tblSixYrSar_Transaction
lkp_SixYrSar_Transaction_Action	Transaction Action – Lookup Table	Includes lookup information on the action identification codes and definitions related to the flagged Sample Analytical Results in tblSixYrSar_Transaction
tblSixYrSar_Transaction	Transaction Table	Flagged monitoring records: reason why 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 necessary to prepare the SYR3 datasets for QA/QC review and, ultimately, for data analysis. The SDWIS/State Extract Tool pulled the SDWIS/State data into Microsoft Access. States that did not use the SDWIS/State Extract Tool were restructured into a format similar to the SDWIS/State Extract Tool's output. The two groups of datasets (the extract states and the non-extract States (referred to for the remainder of this document as the "SDWIS states" and the "non-SDWIS states," respectively) were managed separately, ultimately getting all datasets into the same format.

A status documentation file was maintained for each state. Specifically, the status documentation described the state datasets received as well as the date received, file type, whether the SDWIS/State Extract Tool was used and the date range of the data. The status documentation also described any state-specific notes, issues or concerns. Upon receipt of each state dataset, EPA created state-specific directories for each raw dataset. Original datasets were saved and maintained exactly as received. 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 Dataset Content

Similar to the second Six-Year Review, the first assessment of the submitted SYR3 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 (see Exhibit 2.2), specifically those necessary for the SYR3 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 very useful when trying to interpret the data submitted by the non-SDWIS states. Data dictionary and supporting information files were reviewed for definitions of the various data elements, row and column headings, codes, and acronyms. If any fields were missing or if there were fields that were not recognizable, EPA included a question to the state in their "flagged record report" email. (See Section 5.2 for a more detailed description of the "flagged record report.") In addition, many of the non-SDWIS states submitted datasets with more data elements than necessary. In those cases, EPA determined which data elements were and were not specific to the SYR3 data request.

It was also necessary to confirm that all of the requested contaminants were included in each state dataset (See Exhibit 2.1). As a first step for the non-SDWIS states, EPA reviewed the CHEMIDs (i.e., four-digit SDWIS codes) and/or contaminant names within each state's dataset. Many states included only CHEMIDs or contaminant names. A few other states only included CAS numbers or state-specific codes. EPA populated missing information using a variety of sources including a list of SDWIS codes from the SDWIS/Fed database as well as the ChemIDPlus website (if only CAS numbers were included). There were three states that submitted at least some data for a contaminant or contaminants for which a four-digit SDWIS code could not be determined. Other times, the state appeared to be using an incorrect four-digit SDWIS code for a particular contaminant. EPA compiled a list of questions for states related to issues such as missing contaminants or undetermined CHEMIDs to be included in the "flagged record reports." States were asked questions such as if there was a statewide waiver for missing

contaminants, if certain contaminant data were stored in a separate database, or if there had been a typo with a particular CHEMID.

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

4.2 Restructuring Non-SDWIS State Data

Datasets received from the non-SDWIS states were restructured into a format similar to the data structure of the SDWIS states to allow for the construction of a unified database for the SYR3 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. For example, EPA considered information such as "The state sent in 5 files – one with chems, one with GWR data, one with LT2 data…."

A few states submitted their data as a single flat file. However, the SYR3 ICR SQL database was designed as a relational database so the structure of that flat file had to be modified, or "mapped," into the structure of the relational database. The various data elements had to be mapped from the single flat file table into three separate inventory tables for water systems, facilities, and sample points (tblSixYrWs, tblSixYrWsf, and tblSixYrSpt, respectively). As an example, a flat file from a state may have contained columns for PWSID, population served, and system type for each and every sample analytical result. However, in the final SYR3 SQL ICR database the sample analytical result table (tblSixYrSar) stores the sample analysis results with a water system ID to link it to a single record in a separate water system table (tblSixYrWs) 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 tblSixYrWs. The same procedure was followed with the sample point and facility information. Note that there were cases where a state provided sample point information but not facility information. Within the SYR3 ICR SQL 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.

A few states submitted datasets with incomplete sample non-detection records. Some states aggregate or summarize non-detection results for multi-analyte laboratory methods. For these states, records contained a single record with "0" or "ND" for all contaminants not detected and individual numeric detection records for those contaminants with a positive result. Special processing was required to create individual non-detection records for all contaminants analyzed with the multi-analyte method. For example, EPA-certified laboratory method 502.2 can analyze for 21 different VOCs. If none of the 21 VOCs are detected, a state may create a single record with a code such as "21 VOCs" in the contaminant identification field and a "0" or "ND" in the results field. In these cases, the single reported non-detection record was expanded to 21 separate records, each assigned the appropriate unique contaminant identification code and was identified as a non-detection result. If one or more of the 21 VOCs were detected, the state entered the individual detected contaminants in the contaminant identification field and the concentration detected in the analytical results field as individual observations, but the remaining VOCs with non-detections were again aggregated into a single record with a "0" or "ND" result. To address

this, the specific contaminants with non-detections had to be identified and a separate record was created with a unique contaminant identification code and each record was identified as a non-detection result.

One state submitted some of its data in a vertical format (i.e., contaminant concentrations for different sample dates were included as separate columns of data rather than rows of data). It was necessary to create a single VALUE and single DATE column. The dataset was transposed into the standard horizontal row format (one row per system per contaminant per sample) by appending the various value and date columns to one another.

A few states store their xylenes data not as total xylenes but as separate analytes: m-xylene, o-xylene and p-xylene. For the SYR3 analyses, a single "total xylenes" sample was desired. Thus, a single "total xylenes" record was created for each unique PWSID, sample ID, and date. (In cases where there was not a corresponding m-xylene and p-xylene record for every o-xylene record, the affected records were excluded from the dataset.) The remaining xylenes data needed to have three records for every unique PWSID, SAMPLE ID, DATE combination (one for m-xylene, one for p-xylene and one for o-xylene). When all records were non-detects, the maximum detection limit was used for the newly created "total xylenes" non-detection record. When all records were detections, the three detection values were summed. When one or two xylenes were detected and the other(s) was/were not, only the detected values were summed (essentially setting the non-detections to zero).

For each non-SDWIS state, EPA compiled a list of all tables and data elements, as well as each data element's set of permitted values and a description of each value. From this, the state values were matched to the corresponding values within SDWIS/Fed for the federally reportable data elements. The remaining data elements and permitted values were matched (or "mapped") to the corresponding SDWIS/State values where possible. (For example, the source water type column in the state dataset could be called "PSource"; EPA created a crosswalk table indicating that "PSource" should be mapped to the SDWIS/Fed field "D_FED_PRIM_SRC_CD.") Generally, the states that did not use the Query Extraction Tool provided enough information in data dictionaries or other documentation for EPA to accurately organize the data in the SDWIS/Fed format.

Prior to populating the SYR3 ICR SQL database, EPA standardized the data reported by each non-SDWIS state to reflect the appropriate SDWIS codes. For example, in the source water type field (i.e., "D_FED_PRIM_SRC_CD"), all instances of "surface water" or "S" were changed to "SW." In the system type field (i.e., "D_PWS_FED_TYPE_CD"), all instances of "CWS" or "community" were changed to "C" for community water systems. All PWSIDs had to be put in the federal format of the two-character postal state abbreviation or Region code followed by a seven-digit number, unique to each PWS in the state.

After the various state-specific formatting and transformations were completed, EPA imported all datasets into Access. 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 SYR3 ICR SQL database. All queries were created and saved in Access to document the transformation, ensuring that this process was reproducible.

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

When preparing the data for the occurrence analysis, even prior to the review for potential outliers, etc., it was necessary to get the following three data elements into a consistent format: the sample analytical result sign, sample analytical result value and sample analytical result unit of measure. Many of the state datasets included analytical results signs (e.g., "<" for non-detections or "=" for detections), detection limits and analytical results data in multiple fields. EPA added a "DETECT" field to the SYR3 ICR dataset to identify the results sign and to more easily conduct analyses. Wherever the analytical result was greater than zero and the result sign indicated a detection, then DETECT was set equal to 1, representing a detection. When the analytical result was equal to zero and/or the result sign indicated a non-detection, then DETECT was set equal to 0 (i.e., a non-detect).

Finally, data were received in a variety of units of measure. It was important that all data for each individual contaminant be expressed in a single unit in order to facilitate analysis. Chemical monitoring data were received in both milligrams per liter (mg/L) and micrograms per liter (µg/L). For this analysis, all data for IOCs were converted to mg/L, while all data for the SOCs, VOCs, and uranium were converted to µg/L. Data for alpha particles, beta particles[1], and combined radium-226/228 were analyzed in picocuries per liter (pCi/L). Note that with the exception of asbestos and the radionuclides, all thresholds and concentrations in this report are expressed in µg/L. As described in Section 5.2.7, all records with missing or unusual units in the SYR3 ICR dataset were sent back to states for input.

[1] 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 the state datasets were converted into a consistent format, a significant effort was undertaken to ensure the quality of the data submitted. Data quality, completeness, and representativeness were key considerations for the dataset. Given the size, scope, and variety of formats of the datasets received from the states, EPA conducted extensive data management and QA/QC assessments on the data to be included in the SYR3 ICR dataset. This QA/QC effort encountered a range of data quality across the different contaminants and different states. Included below is 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 below. For additional QA/QC measures performed for the MDBP data, refer to USEPA (2016b) and USEPA (2016c).

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

The final SYR3 ICR dataset consists of compliance monitoring data received from 54 out of 67 states/primacy agencies. It represents a very large sample and the largest compliance monitoring dataset ever compiled and analyzed by EPA's Drinking Water Program. The 54 states/primacy agencies that provided data for the SYR3 ICR dataset comprise 95 percent of all PWSs and 92 percent of the total population served by PWSs nationally, and are geographically representative of PWSs nationwide.

The absence of data from the 4 states and 9 primacy agencies in the final SYR3 ICR dataset could potentially bias the dataset's representation of the national occurrence of particular contaminants. The four states, representing about 5 percent of PWSs and 8 percent of 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 four states that did not provide compliance monitoring contaminant occurrence data (Colorado, Delaware, Georgia, and Mississippi) are generally geographically distributed across the United States and reflect a diverse mix of urban, agricultural, and industrial areas. No regional geologic terrain, climatic or hydrologic zone, geography, or socio-economic activity is unrepresented in the dataset. Although two states in the southeastern U.S., Georgia and Mississippi, did not provide data, all other southeast states provided data, allowing for substantial regional coverage, especially from a population-based perspective. All other regions had at most one state not included in the dataset. The SYR3 ICR dataset, with 46 of the 50 states represented, is therefore considered reasonably complete and nationally representative as the basis of the contaminant occurrence estimates for this Six-Year Review. To further address the issue of potential bias, though, EPA conducted an assessment for the chemical phase and radionuclides by comparing occurrence in the 4 states to that in the 46 states.

Because a complete compliance monitoring dataset of all 50 states was not available to EPA, it is not possible to know the true national occurrence for a particular contaminant or how occurrence rates for a particular contaminant in the 4 missing states compare to occurrence in the other 46 states. 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 all 50 states. EPA compiled SDWIS/Fed records of MCL violations for the chemical phase and radionuclide rules only, used here as an indicator of contaminant occurrence, by state

for the same years (2006-2011) as the SYR3 ICR dataset.[2] The MCL violation records were used to determine if the violation rate in the 4 missing states was significantly different than the violation rate in the 46 states in the dataset, or if the violation rate in the 46 states could be considered representative (from the same statistical population). EPA conducted this assessment for the IOCs, SOCs, VOCs, and radionuclides evaluated under SYR3.

The mean MCL violation rate for each contaminant (i.e., the percentage of systems with at least one MCL violation) was calculated for the 46 states in the dataset and separately for the 4 states not in the SYR3 ICR dataset. For each contaminant, a statistical t-test was used to determine whether these two estimated mean MCL violation rates (46-state vs. 4-state) were significantly different; the t-test had an alpha (α) level of 0.05 and assumed unequal variance.[3] If the p-value resulting from the t-test was less than 0.05, EPA rejected the null hypothesis that the two mean MCL violation rates were from the same population and accepted the alternative hypothesis that they were from different populations.

Of the 61 contaminants evaluated, only nine contaminants had at least one MCL violation listed in the SDWIS/Fed database for the 2006-2011 time period; thus, t-tests were conducted on only these nine contaminants. For five contaminants (fluoride, nitrate, gross alpha, uranium, and combined radium), the t-test resulted in a p-value > 0.05 (EPA failed to reject the null hypothesis). This suggests, but does not prove, that the mean MCL violation rates for the 46 states and the 4 states were not statistically different (were from the same population). For three additional contaminants, only one of the four states had MCL violations, and so the t-test could not be applied.

Arsenic was the only contaminant for which the t-test resulted in a p-value < 0.05 (EPA rejected the null hypothesis); thus, the mean arsenic MCL violation rate for the 46 states appears to be statistically different (come from a different population) than the mean arsenic MCL violation rate for the 4 states. This suggests that the absence of system compliance monitoring data from the four states might result in some amount of over-estimation of occurrence for that contaminant. These findings, however, are most appropriately used as context or background for the quantitative occurrence findings presented in USEPA (2016a).

To further evaluate the completeness of each state's dataset, EPA used the SDWIS/Fed database as a reference and compared the number of water systems by state in the SYR3 ICR dataset to the number of systems by state in the SDWIS/Fed database (frozen fourth quarter 2011). Only the SDWIS/Fed database records from the 46 states also in the SYR3 ICR dataset were included. As described in Section 6.2 purchased water systems (systems that purchase 100 percent of their water) are accounted for differently than non-purchased water systems. To simplify this comparison of number of systems by state, only non-purchased systems were included in the counts. Although the system inventory information represented in the two data sources is very

[2] 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 data were extracted from SDWIS/Fed by EPA in March 2014.

[3] The t-test calculation used considered the variance, mean, and sample size of each of the two groups of states to estimate the probability that the observed difference in sample means represents an actual difference in compliance monitoring and not just a statistical inconsistency resulting from low sample sizes.

similar, it is not equivalent. The main difference is that the SYR3 ICR dataset counts reflect the total number of active water systems with compliance monitoring data in any of the six years represented in the dataset (2006-2011), while the SDWIS/Fed 2011 fourth quarter data freeze counts reflect the total number of active water systems in a single year (2011). 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 5.1 presents this comparison between the SDWIS/Fed and SYR3 ICR datasets. In order to be consistent with the SDWIS/Fed counts, the population values listed for the SYR3 ICR dataset include only the populations directly served by non-purchased systems (retail populations); total adjusted populations are discussed in Section 6.2.

Exhibit 5.1 compares the number of systems and population served by these systems in the December 2011 SDWIS/Fed freeze and the SYR3 ICR dataset by state. The comparison between the counts of systems in the two data sources indicate that the data in the SYR3 ICR dataset are reasonably complete. Overall, there is an approximately 11 percent difference between the number of systems listed in a December 2011 SDWIS/Fed freeze compared to the number of systems in the SYR3 ICR dataset. (The percent difference is calculated by subtracting the number of systems in SDWIS/Fed from the number in SYR3 ICR, and then dividing by the number of systems in the SYR3 ICR dataset.) In Exhibit 5.1, positive values for percent difference indicate that more systems are reported in the SYR3 ICR dataset, while negative values indicate that more systems are reported in the 2011 SDWIS/Fed Freeze. Comparing the number of systems for each state, the absolute percentage difference between SDWIS/Fed and the SYR3 ICR dataset ranges from a 0 percent difference (e.g., Region 1 Tribes and Utah) to an approximately 26 percent difference (e.g., Region 5 Tribes) in the number of systems. Based on the population served by systems, there is a three percent difference between the total population-served by systems listed in SDWIS/Fed and the population served by systems listed in the SYR3 ICR dataset. Comparing individual state population served values, the absolute percentage differences between SDWIS/Fed and the Six-Year states ranges from less than a 1 percent difference (e.g., Alabama and New Mexico) to approximately a 20 percent difference (e.g., Nebraska). Based on the comparisons presented in Exhibit 5.1, the SYR3 ICR dataset is representative of national PWSs and population served and suitable for use as the basis of national contaminant occurrence estimates.

Exhibit 5.1: Comparison of the Total Number of Non-Purchased Systems and Retail Population Served in SDWIS/Fed and the SYR3 ICR Dataset, By State

State	Total Number of Non-Purchased Systems[1]			Retail Population Served by Non-Purchased Systems		
	2011 SDWIS/Fed Freeze	SYR3 ICR Dataset	Percent Difference[2]	2011 SDWIS/Fed Freeze	SYR3 ICR Dataset	Percent Difference[2]
Alabama	399	415	4%	4,270,460	4,269,317	< -0.1%
Alaska	1,429	1,403	-2%	718,776	762,190	6%
American Samoa	19	17	-11%	60,958	61,309	1%
Arizona	1,511	1,493	-1%	6,414,815	6,431,456	0.3%
Arkansas	643	639	-1%	1,808,219	1,782,034	-1%

State	Total Number of Non-Purchased Systems[1]			Retail Population Served by Non-Purchased Systems		
	2011 SDWIS/Fed Freeze	SYR3 ICR Dataset	Percent Difference[2]	2011 SDWIS/Fed Freeze	SYR3 ICR Dataset	Percent Difference[2]
California	7,215	7,540	5%	28,781,357	28,528,121	-1%
Connecticut	2,523	2,971	18%	2,676,429	2,716,577	2%
Florida	5,295	6,350	20%	16,742,435	17,383,116	4%
Hawaii	108	118	9%	1,421,758	1,452,737	2%
Idaho	1,936	1,907	-1%	1,315,860	1,360,791	3%
Illinois	4,097	4,625	13%	8,228,681	8,296,918	1%
Indiana	4,012	4,397	10%	4,886,097	4,946,190	1%
Iowa	1,660	1,763	6%	2,365,619	2,380,108	1%
Kansas	647	642	-1%	2,281,561	2,292,280	0.5%
Kentucky	261	257	-2%	3,268,613	3,299,397	1%
Louisiana	1,287	1,390	8%	4,844,307	4,868,351	0.5%
Maine	1,851	2,198	19%	903,130	964,872	7%
Maryland	3,390	3,886	15%	5,022,871	5,711,914	14%
Massachusetts	1,545	1,674	8%	7,154,525	7,117,276	-1%
Michigan	10,873	13,078	20%	4,809,937	5,087,202	6%
Minnesota	6,943	7,753	12%	4,617,552	4,689,328	2%
Missouri	2,458	2,768	13%	4,463,766	4,515,797	1%
Montana	1,899	1,856	-2%	894,851	902,225	1%
Navajo Nation	146	152	4%	131,031	140,818	7%
Nebraska	1,155	1,283	11%	1,545,502	1,861,572	20%
Nevada	531	584	10%	942,651	984,355	4%
New Hampshire	2,394	2,610	9%	1,124,928	1,156,828	3%
New Jersey	3,686	4,295	17%	7,428,858	7,534,923	1%
New Mexico	1,109	1,089	-2%	1,899,344	1,896,614	-0.1%
New York	8,206	8,945	9%	16,731,989	18,127,928	8%
North Carolina	5,684	6,806	20%	6,945,228	7,131,934	3%
North Dakota	301	279	-7%	513,800	508,028	-1%
Ohio	4,543	5,363	18%	9,056,572	9,232,856	2%
Oklahoma	960	1,102	15%	3,002,063	3,091,513	3%
Oregon	2,484	2,705	9%	2,831,651	2,767,113	-2%
Pennsylvania	8,779	10,128	15%	10,699,485	10,814,930	1%
Region 1 - Tribes	5	5	0%	49,031	49,031	0%
Region 4 - Tribes	31	32	3%	28,387	27,889	-2%
Region 5 - Tribes	100	126	26%	139,916	154,489	10%

State	Total Number of Non-Purchased Systems[1]			Retail Population Served by Non-Purchased Systems		
	2011 SDWIS/Fed Freeze	SYR3 ICR Dataset	Percent Difference[2]	2011 SDWIS/Fed Freeze	SYR3 ICR Dataset	Percent Difference[2]
Region 8 - Tribes	103	101	-2%	91,321	92,432	1%
Region 9 - Tribes	284	314	11%	367,252	353,335	-4%
Rhode Island	459	487	6%	775,182	778,796	0.5%
South Carolina	1,104	1,064	-4%	2,681,749	2,683,477	0.1%
South Dakota	447	463	4%	603,361	609,007	1%
Tennessee	700	673	-4%	5,616,106	5,704,724	2%
Texas	5,635	5,528	-2%	16,682,616	17,119,034	3%
Utah	892	892	0%	1,443,051	1,470,928	2%
Vermont	1,273	1,414	11%	489,778	503,324	3%
Virginia	2,519	2,917	16%	4,769,127	5,340,030	12%
Washington	3,902	4,309	10%	5,038,297	5,149,128	2%
Washington, D.C.	1	1	0%	0	0	0%
West Virginia	822	988	20%	1,292,503	1,314,496	2%
Wisconsin	11,345	12,563	11%	4,468,486	4,576,227	2%
Wyoming	698	682	-2%	380,269	378,901	-0.4%
Total	**132,299**	**147,040**	**11%**	**225,722,111**	**231,374,166**	**3%**

[1] More than half of the water systems with data in the SYR3 ICR dataset are transient non-community water systems. Because only the nitrate/nitrite regulations require compliance monitoring by these transient systems (see Exhibit 5.3), 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 'percent difference' was calculated by subtracting the 2011 SDWIS/Fed Freeze total number of non-purchased systems (or retail population served by systems) from the SYR3 ICR dataset total number of non-purchased systems (or retail population served by systems). That difference was then divided by the total number of non-purchased systems (or retail population served by systems) from the SYR3 ICR dataset. The 'percent difference' is less than zero if the SYR3 ICR dataset indicated a smaller number of systems (or retail population served by systems).

Exhibit 5.2 compares the number of systems and population served by these systems in the December 2011 SDWIS/Fed freeze and the SYR3 ICR dataset stratified by source water type and system type. (Only non-purchased systems and their retail population served are included in this comparison.) The overall national 46 state totals indicate about 11 percent more systems and a 3 percent greater population served is reported in the SYR3 ICR dataset than is represented in SDWIS/Fed. For community water systems (CWSs), there is about a four percent difference based on the number of systems and a two percent difference based on the population served by systems. Percentage differences were larger for ground water systems than surface water systems. For non-transient non-community water systems (NTNCWSs), there is about a 13 percent difference based on the number of systems and an 8 percent difference based on the population served by systems. For transient non-community water systems (TNCWSs), there is about a 12 percent difference based on the number of systems and a 7 percent difference based on the population served by systems. CWSs account for approximately 93 percent of the total population served by systems in the United States. Despite these percent differences apparent

between the SDWIS/Fed data and the SYR3 ICR data, the SYR3 ICR dataset is suitable for use as the basis of national contaminant occurrence estimates. As is stated earlier in this report, the 54 states/primacy agencies that provided data for the SYR3 ICR dataset comprise 95 percent of all PWSs and 92 percent of the total population served by PWSs nationally, and are geographically representative of PWSs nationwide.

Exhibit 5.2: Comparison of the Total Number of Systems and Retail Population Served in SDWIS/Fed and the SYR3 ICR Dataset, By Source Water Type and System Type

Source Water Type	2011 SDWIS/Fed Freeze				SYR3 ICR Dataset				
	CWS	NTNCWS	TNCWS	Total	CWS	NTNCWS	TNCWS	Unknown[1]	Total
Number of Non-Purchased Systems									
Ground Water (GW)	33,247	16,325	77,221	126,793	34,576	18,802	87,816	123	141,317
Surface Water (SW)	4,226	322	958	5,506	4,327	335	1,058	3	5,723
Total	37,473	16,647	78,179	132,299	38,903	19,137	88,874	126	147,040
Retail Population Served									
Ground Water (GW)	77,175,728	4,734,551	9,552,196	91,462,475	79,082,376	5,148,753	10,332,691	2,573	94,566,393
Surface Water (SW)	133,813,746	153,948	291,942	134,259,636	136,398,900	137,898	270,751	224	136,807,773
Total	210,989,474	4,888,499	9,844,138	225,722,111	215,481,276	5,286,651	10,603,442	2,797	231,374,166

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

EPA conducted supplementary evaluations of the completeness and representativeness for microbial contaminant regulations and D/DBPRs. For more detailed information on evaluation of the microbial contaminants' SYR3 ICR data, refer to USEPA (2016b). For more detailed information on the evaluation of SYR3 ICR data for contaminants regulated under the D/DBPRs, refer to USEPA (2016c).

5.2 Quality Assurance Measures

Before analyzing contaminant occurrence, EPA performed a rigorous QA/QC evaluation of the data from each state. EPA sent emails to each state, asking specific questions about its dataset, as appropriate. Question topics included descriptions of non-intuitive data element names, definitions of field headings, or non-standard codes that were not described in any documentation files from the state. EPA also confirmed that all of the requested contaminants were included in each state dataset. When a state was missing data for any of the contaminants listed in Exhibit 2.1, EPA asked the state to identify the reason for the omission, such as a state-wide waiver of the requirement to monitor for the contaminant(s).

Exhibit 5.3 lists the systems that are required to sample for the contaminants within each chemical group. All data that passed the QA/QC process from these systems were included in the SYR3 analyses. Data from systems that were not required to sample for a given contaminant (e.g., SOC data from transient systems or radionuclide data from transient or non-transient non-community systems) were excluded from the SYR3 analyses.

Exhibit 5.3: Chemical Group Monitoring Requirements

Chemical Group	System Types Required to Sample (sample data included in analyses)	System Types Not Required to Sample (sample data excluded from analyses)
Inorganic Chemicals (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.
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 Chemicals (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 Chemicals (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.

EPA created several automated data QA checks within the SYR3 ICR dataset. These QA checks identified (or "flagged") records of potential data quality concerns. EPA sent out a detailed report to each state describing their flagged records. These reports included the counts of flagged records by category, as well as specific questions related to each of these categories. In addition, an attachment identified the specific records that were flagged. EPA requested that each state provide the appropriate disposition (delete, make corrections, etc.) of these flagged records. EPA documented all changes made to the compliance monitoring data and suggested to the states that they make corrections in their data system as well, if appropriate. To resolve data quality issues that required significant corrections to the raw data, such as identifying outliers or identifying and changing incorrect units, state data management staff were consulted when appropriate before data corrections were completed.

The sections below (5.2.1 through 5.2.12) provide a description of the various QA measures that were used to identify records of potential data quality concern. For all flagged records, input from the states was always used 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. Note: No records were deleted from the SYR3 ICR dataset. When a determination was made to exclude records from the occurrence analyses, a code was added to the "transaction

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

Note that Section 5.2.1 through Section 5.2.6 describe the QA measures that were applied to the entire database (i.e., were relevant to all regulated contaminant monitoring data in the SYR3 ICR dataset). The QA measures described in Section 5.2.7 through Section 5.2.12 relate specifically to the 61 contaminants regulated under the Phase I, II, IIb, and V Rules, the Arsenic Rule and the Radionuclides Rule whose occurrence analyses are described in USEPA (2016a). (The Phase I, II, IIb, and V Rules and the Arsenic Rule are described as the "Chemical Phase Rules" in subsequent sections.) Exhibit 5.4 and Exhibit 5.5 below provide a visual for the overall flow of the QA/QC process. Exhibit 5.4 presents the QA measures that were applied to all contaminants in the SYR3 ICR dataset. Exhibit 5.5 presents the QA measures that were applied only to the "Chemical Phase Rule" contaminants. Details on additional QA/QC measures specific to the microbials and DBPs (including QA/QC measures applied to TOC) can be found in USEPA (2016b) and USEPA (2016c). Note that additional QA/QC measures were also taken to identify and exclude fluoride samples from fluoridated water systems. See Appendix D for more information on additional QA/QC measures for fluoride data.

Exhibit 5.4: Flow Chart of QA Measures Applied to Entire SYR3 ICR Dataset

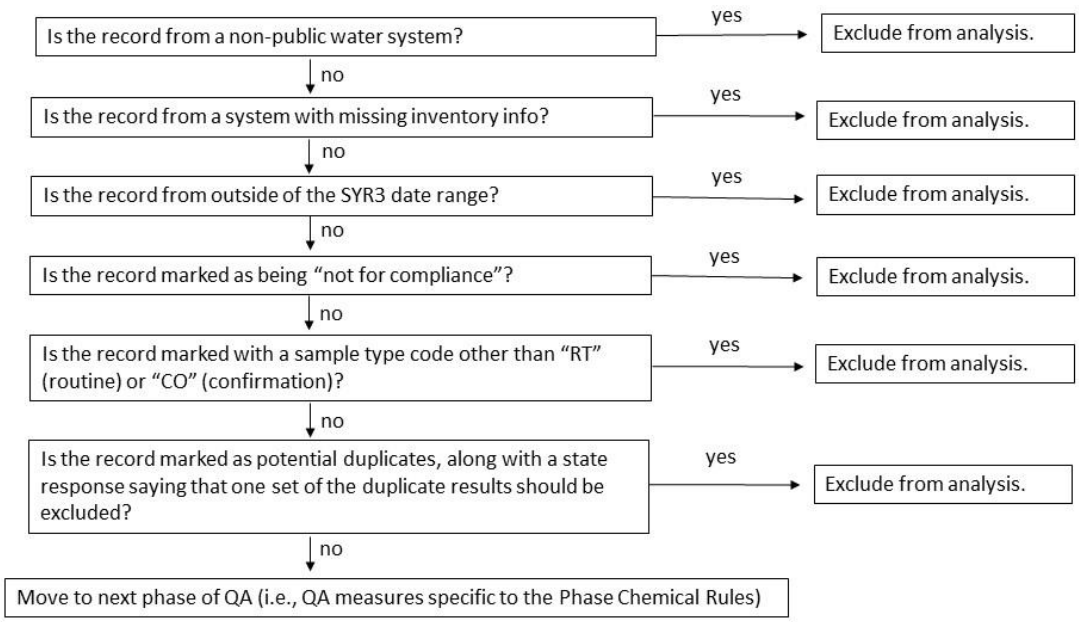
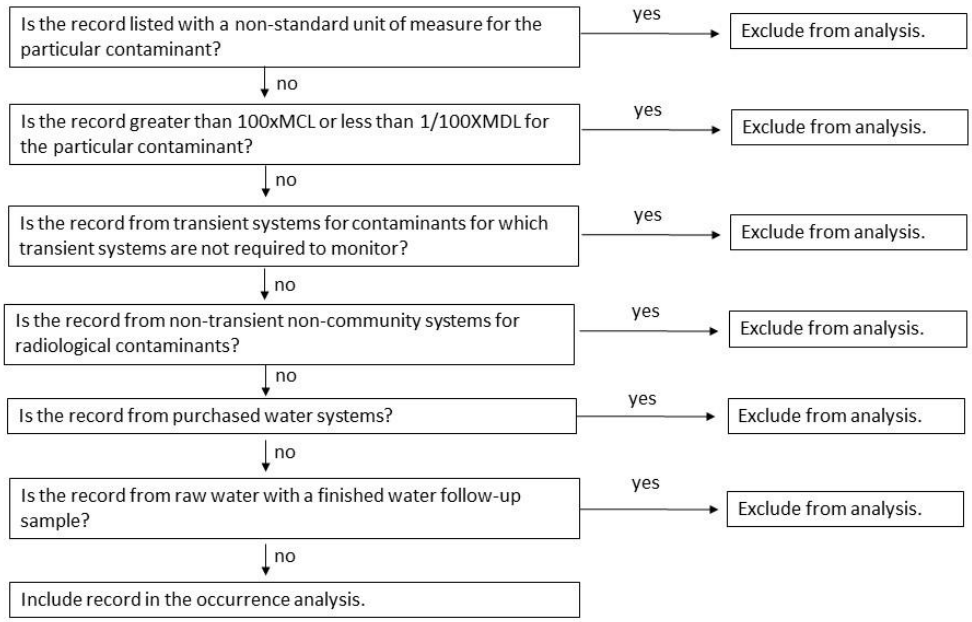


Exhibit 5.5: Flow Chart of QA Measures Applied to Chemical Phase and Radionuclide Rules' Contaminants Only



After applying the various QA measures to more than 13 million SYR3 ICR records for the Chemical Phase and Radionuclide Rules' contaminants, almost 95 percent of the records remained in the final dataset that was used for conducting occurrence analyses. Most of the records were removed in either Step 9, removal of records from transient water systems for contaminants for which transient water systems aren't required to sample, or in Step 11, removal of records from purchased water systems (systems that are not required to sample for the Chemical Phase or Radionuclide Rule contaminants). Exhibit 5.6 documents the specific counts of records included and excluded in each QA step.

Exhibit 5.6: Summary of the Count of Records Removed via the QA Measures Applied to Chemical Phase and Radionuclide Rules' Contaminants

QA Step	Chemical Phase and Radionuclide Rule Contaminants	
	Included	**Excluded**
Original number of records	13,263,466	
Step 1: Removal of records from non-public water systems	13,234,811	28,655
Step 2: Removal of records from systems with missing inventory data	13,230,314	4,497
Step 3: Removal of records from outside the SYR3 date range	13,165,136	65,178
Step 4: Removal of records marked as non-compliance	13,102,451	62,685
Step 5: Removal of records marked with a sample type code other than routine or confirmation	13,048,326	54,125

QA Step	Chemical Phase and Radionuclide Rule Contaminants	
	Included	Excluded
Step 6: Removal of duplicate records	13,041,190	7,136
Step 7: Removal of records with non-standard units	13,041,042	148
Step 8: Removal of records that are potential high or low outliers	13,036,947	4,095
Step 9: Removal of records from transient water systems for contaminants for which transients are required to sample	12,726,735	310,212
Step 10: Removal of records from non-transient water systems for radionuclides	12,718,035	8,700
Step 11: Removal of records from purchased water systems	12,598,568	119,467
Step 12: Removal of raw water records without a follow-up finished water sample	12,552,409	46,159
Final number of records	**12,552,409**	
Percent Included	**94.6%**	

5.2.1 Non-Public Water Systems

Some primacy agencies require water systems that do not meet the criteria to be classified as public water systems to submit sample results that are "routine" or "for compliance." The primacy agency's information system usually identifies these water systems as "non-public" or uses another method to differentiate them from public water systems. Non-public water systems have fewer than 15 service connections and serve fewer than 25 people. All records from non-public water systems were excluded from the occurrence analysis.

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 or no population served). When the data were missing, EPA included a question to the state in their "flagged record report" to ask if they meant to include these data and/or informed the state that those data would be acquired from the 4th quarter 2011 SDWIS/Fed data freeze unless they preferred to send the information themselves. When inventory data were incomplete or missing and the states did not respond to inquiries, the missing data were populated with data from the SDWIS/Fed freeze from December 2011. All cases where SDWIS/Fed data were used to populate inventory data fields in the state's dataset were documented. 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 SYR3 ICR requested data from 1/1/2006 through 12/31/2011. The SDWIS/State Extract Tool only extracted 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

There are several scenarios where 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 SDWIS/State Extract Tool only extracted 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." Two, non-SDWIS states (California and Michigan) did not make a designation as to whether their data were for compliance. For all occurrence analyses, EPA assumed that all data from these two states were for compliance. All sample results flagged as "not for compliance" were excluded from the occurrence analysis.

5.2.5 Non-Routine

Some primacy agencies have regulations that are more stringent than the NPDWRs and require water systems to submit more sample results than federally required. Primacy agencies also may require laboratories to report *all* sample results from water systems including results from contaminants that are not regulated. Usually non-routine sample results that are specifically listed as "special request" in the database are also identified as being "non-compliance" samples. Most other types of non-routine sample results, such as confirmation, repeat or maximum residence time sample results are "for compliance." While the SDWIS/State Extract 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 (2016a); all other sample results for those contaminants were considered to be "non-routine" and were excluded from the occurrence analysis. See USEPA (2016b) and USEPA (2016c) for more details on the sample type codes that were excluded from the microbial and DBP occurrence analyses, respectively.

5.2.6 Duplicate Records

In the SYR3 analysis, potential duplicates were identified as all detection records with the same PWSID, Sample Point ID, analyte, sample collection date, and concentration. To be consistent with the second Six-Year Review, all records identified as potential duplicates were included in the occurrence analysis unless the state responded to say that the records were indeed duplicates and one set should be excluded from the analysis.

5.2.7 Units of Measure (Chemical Phase and Radionuclide Rules Only)

EPA identified all detection records where the units of measure reported were not one of the standard units used for the particular contaminant (i.e., not equal to "MG/L," "UG/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. All records in non-standard units were excluded from the occurrence analyses unless there was strong evidence of the correct standard unit to use (e.g.,

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).

5.2.8 Potential Outliers (Chemical Phase and Radionuclide Rules Only)

To identify potential high outliers, EPA flagged all detected concentrations that were greater than four times the contaminant's MCL and all detected concentrations that were greater than ten times the contaminant's MCL. To identify potential low outliers, EPA flagged all detected concentrations that were less than the contaminant's minimum Method Detection Limit (MDL[4]) and all detected concentrations that were less than one-tenth the minimum MDL. See Exhibit 5.7 for a list of all MCL values relevant to the Chemical Phase and Radionuclide Rules' contaminants only. (See USEPA, 2016b and USEPA, 2016c for values relevant to the microbials and DBPs).

EPA included questions to the state on each of these potential high and low outliers in their "flagged record report." Any changes suggested by the states were implemented for these records. For example, some states wrote back to say there were "no errors" in their high detect concentrations or that they had "no reason or evidence to show these data to be invalid." Other states stated that "all of the high results were due to using mg/L when they should have been µg/L." For the states that did not respond, all detected 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 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 MDL were included in the analysis. The values of 100XMCL and 1/100XMDL were chosen as conservative high-end and low-end cut-offs, respectively.

Exhibit 5.7: 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 Chemicals				
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

[4] The Method Detection Limit, MDL, is defined as the minimum concentration of a substance that can be measured with 99 percent confidence, based on an analyte concentration being greater than zero as determined from analysis of a sample in a given matrix containing the analyte. In other words, the MDL is the concentration at which presence or absence of an analyte can be dependably determined. This contrasts with the Minimum Reporting Level (MRL), which is a concentration above the MDL, typically set two to ten times the MDL, and allows for reporting at specified levels of precision and accuracy of the actual concentration of the analyte present in the sample.

Contaminant	Maximum Contaminant Level (MCL)		Method Detection Limit (MDL)	
	Value	Unit of Measure	Value	Unit of Measure
Cadmium	5	µg/L	0.05	µg/L
Chromium (Total)	100	µg/L	0.08	µg/L
Cyanide	200	µg/L	5	µg/L
Fluoride	4,000	µg/L	0.01	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 Chemicals				
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
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

Contaminant	Maximum Contaminant Level (MCL)		Method Detection Limit (MDL)	
	Value	Unit of Measure	Value	Unit of Measure
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 Chemicals				
1,2-Dichlorobenzene	600	µg/L	0.02	µg/L
1,4-Dichlorobenzene	75	µg/L	0.01	µg/L
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
Ethylbenzene	700	µg/L	0.01	µg/L
Monochlorobenzene	100	µg/L	0.01	µg/L
Styrene	100	µg/L	0.01	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
Xylenes (Total)	10,000	µg/L	0.01	µg/L
Radiological Contaminants				
Alpha Particles	15	pCi/L	--	--
Beta Particles [1]	50	pCi/L	--	--
Combined Radium-226 & -228	5	pCi/L	--	--
Uranium	30	µg/L	--	--

[1] 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.

5.2.9 Transient Water Systems (Chemical Phase and Radionuclide Rules Only)

Transient non-community water systems operate for at least 60 days per year and serve at least 25 people per day. Transient water systems are usually identified by system type “transient, non-community” or something similar. As such, transient water systems are only required to submit nitrate, nitrite and combined 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 (2016a), except for rules that transients are required to monitor.

5.2.10 Non-Transient Water Systems (Radionuclides Only)

Transient non-community water systems and non-transient non-community water systems are not required to submit radiological sample results. Unless a state responded to say that the system in question used to be a CWS at the time of sampling (and thus the records should be included), all data from transient and non-transient water systems were excluded from the occurrence analyses for the radionuclides.

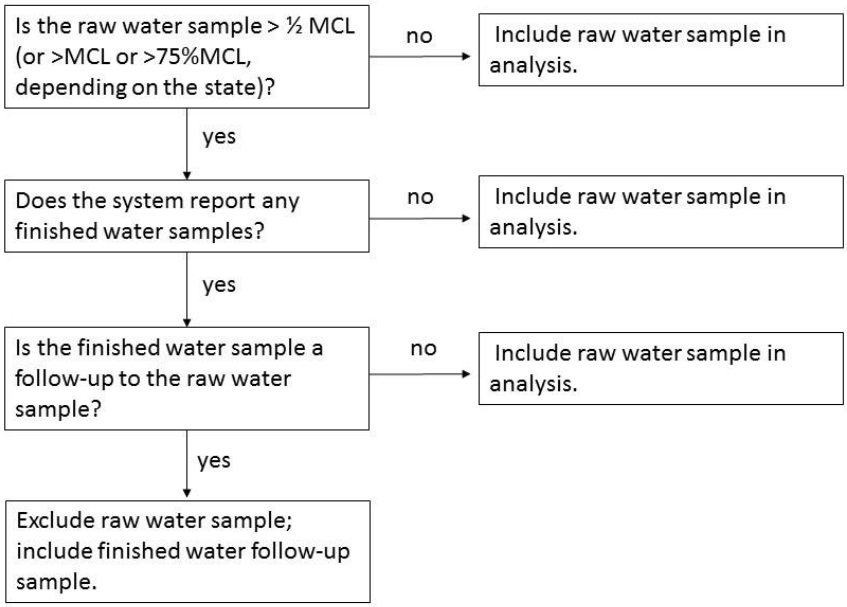
5.2.11 Purchased Water Systems (Chemical Phase and Radionuclide Rules Only)

Purchased water systems buy all their water from one or more water systems. These systems do not have sources that require entry point monitoring for the Chemical Phase or Radionuclide rules. All results from purchased systems were excluded from the occurrence analyses presented in USEPA (2016a). Population-served values and occurrence estimates in USEPA (2016a) were generated using the total (adjusted) population served. (See Section 6.2 for a description of the adjustments of the population served by public water systems for the wholesaler and retail systems.)

5.2.12 Samples in Source/Raw Water (Chemical Phase and Radionuclide Rules Only)

The water source type (i.e., raw or finished) of all potential outliers was investigated since in some states, systems are allowed to monitor at source (raw) water sampling points. If a contaminant is detected in a source water sample, the system is required to collect a follow-up sample at the entry point to the distribution system, unless there was no treatment. EPA developed a protocol for handling the raw (untreated or unfinished) samples related to the Chemical Phase and Radionuclide Rules' contaminants (see Exhibit 5.8).

Exhibit 5.8: Flow Chart of Protocol for the Inclusion of Raw Water Sample Results[1]



[1] Some states have different thresholds that, when exceeded by source water monitoring, would require follow-up monitoring from the entry point. For some states, this threshold is the MCL; for other states, the threshold is ½ the MCL or 75 percent of the MCL.

5.3 System Inventory Updates

For the SYR3 analyses, each system must have a single source water type and population-served designation to define each system in a unique source water type/population size strata. Systems using both ground water and surface water, and systems using ground water under direct influence of surface water, were considered surface water systems for analysis. Systems with more than one specified value of their population served in the original data were included using their most frequently occurring population served value.

For the Chemical Phase and Radionuclide Rule analyses, an additional adjustment to 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 intake or other SW facilities but do purchase some SW; however, in addition, they do have some of 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 very small portion of their water and the rest of the water comes from GW wells. Based on the QA criteria described in Section 5.2.11, data from these systems should be excluded from the SYR3 data analyses presented in USEPA (2016a) since data from purchased water systems were excluded. However, the GW sources from these systems did provide legitimate (and required) compliance monitoring data. Thus, it was necessary in the SYR3 analyses to consider these SWP systems as GW systems since the compliance monitoring data that were provided by these systems were from GW sources.

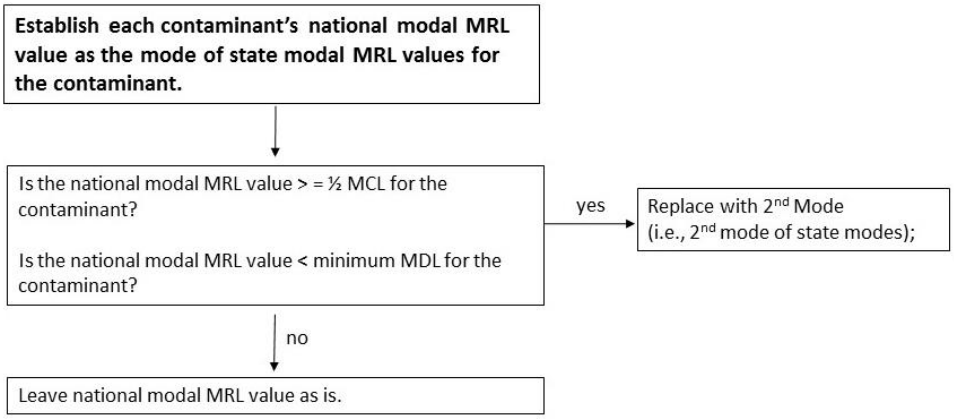
6 Data Preparation for Analyses

6.1 Non-detection record replacement (Chemical Phase and Radionuclide Rules only)

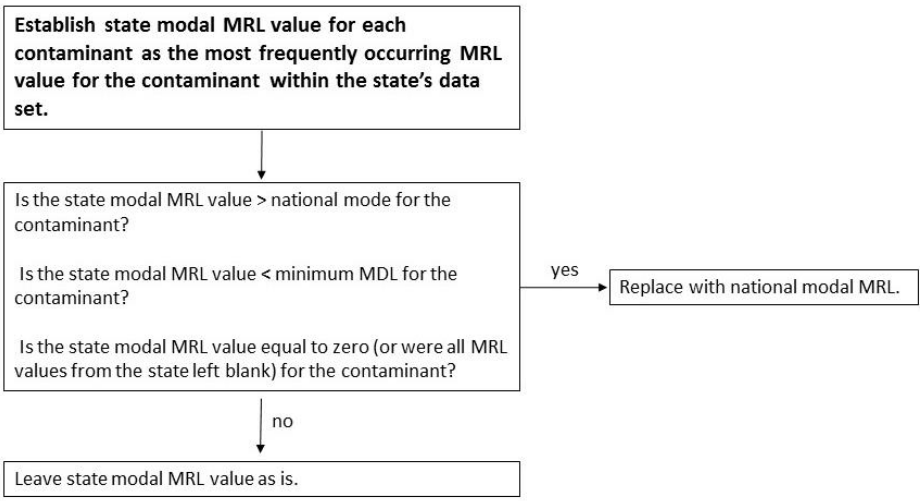
Within the SYR3 ICR dataset, each sample analytical result specifies a sample analytical result value and a sample analytical result sign to indicate whether that result is a detection (i.e., greater than or equal to the MRL) or a non-detection. Sample records reported as non-detections tended to be less uniform and less complete than sample records for analytical detections. For some of the states that did report MRL data for systems, 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, system mean concentrations were calculated 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. For additional details on how non-detections were handled for the DBP data, refer to USEPA (2016c).

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 (derived directly from the PWS compliance monitoring data submitted to EPA in the SYR3 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 modal MRL values for that contaminant. If state-modal MRL values were extremely low or high, a process was developed to identify and replace such values with more reasonable MRL values. A description of the three steps in this process is below.

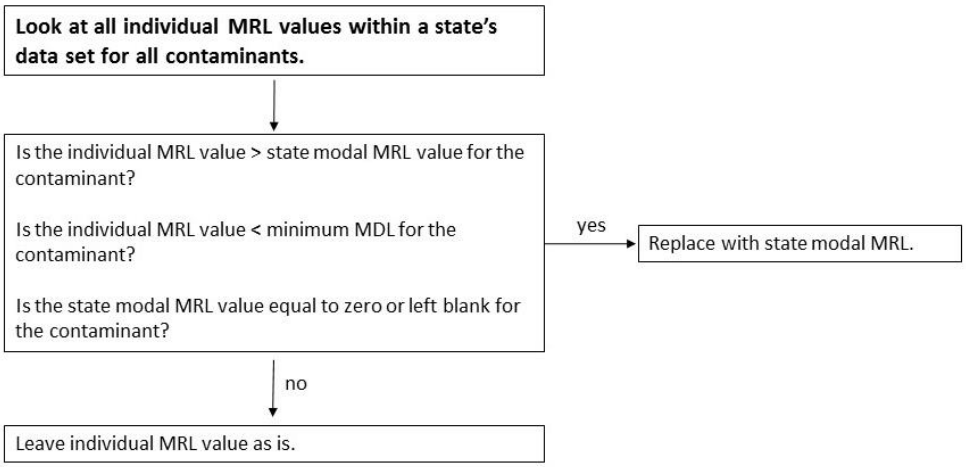
Step 1: Establish a national modal MRL value for each contaminant



Step 2: Establish a state modal MRL value for each contaminant



Step 3: Review individual MRL values for potential replacement



6.2 Adjustments of Population Served by Public Water Systems

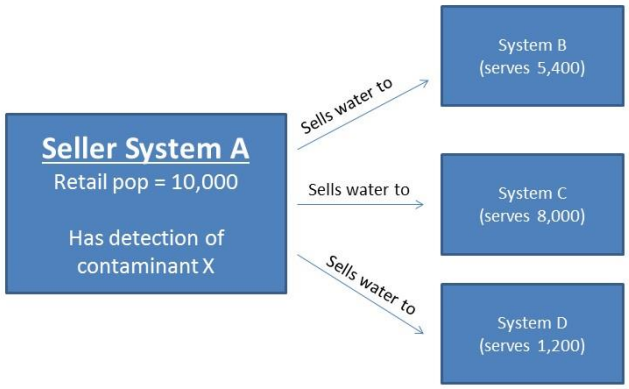
"Purchased" water systems are the systems that purchase 100 percent of their water from other systems ("seller" or "wholesaler" systems). Compliance monitoring requirements are different for purchased water systems compared to non-purchased systems because purchased water systems do not have their own water sources (e.g., wells or intakes). For the occurrence analyses presented in USEPA (2016a) of the Chemical Phase and Radionuclide Rules' contaminants, EPA excluded data from systems that purchase 100 percent of their water, as those systems are not required to sample for those contaminants.[5] However, EPA did adjust the population values of the wholesale systems to include the population of the systems that they sell to (the purchased

[5] Note that consecutive (or "purchasing") water systems do their own sampling for microbial contaminants and DBPs; thus, the data from these systems were not excluded from the microbial and DBP occurrence analyses (see USEPA, 2016b and USEPA, 2016c).

water systems) for those analyses. The population served directly by these wholesale systems is known as the "retail population," while the population served indirectly through the purchased systems is known as the "wholesale population." This adjustment ensured that the entire relevant population was included in the exposure estimates.

Exhibit 6.1 below helps illustrate a simple example of these adjustments. In the diagram, Systems B, C and D (the purchased systems) buy 100 percent of their water from System A (the wholesale system). System A is required to monitor for contaminant X; however, Systems B, C, and D are not. If there is a detection of contaminant X and population values were not adjusted, the exposure estimates would not take into account the populations served by System B, System C, and System D, even though these populations would indeed be exposed to contaminant X. To correct for this, EPA uses the total population served (retail plus wholesale population) for System A for all population-served estimates, which is equal to 24,600 people.

Exhibit 6.1: Simple Illustration of the Total (Retail plus Wholesale) Population Served by Selling Systems



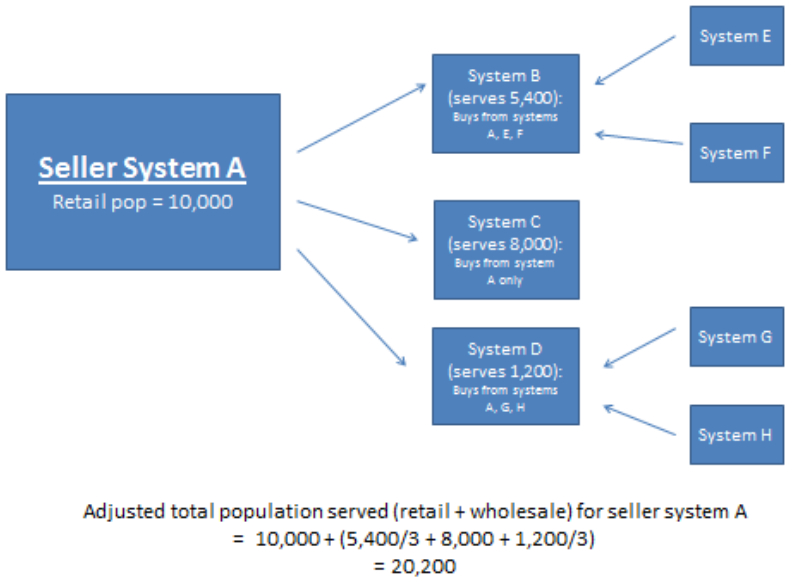
Total population served by seller system A exposed to detection of contaminant X
= retail population + wholesale population
= 10,000 + (5,400 + 8,000 + 1,200)
= 24,600

For some systems, a slightly more complicated adjustment to the wholesalers' total population served values was required. Many purchased water systems actually 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. There are no data available on the actual relative quantities of water purchased from the different wholesalers; therefore, in the cases of multiple wholesalers, the population served by the purchased system was assumed to be uniformly distributed across the wholesalers.

Exhibit 6.2 below illustrates the complete population adjustment for System A, including the uniform distribution of the purchased 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 System A, System E, and System F.

Exhibit 6.2: Illustration of the Allotment of Wholesale Population to the Selling System



Adjusted total population served (retail + wholesale) for seller system A
= 10,000 + (5,400/3 + 8,000 + 1,200/3)
= 20,200

To make adjustments across the SYR3 ICR data, EPA compiled a list of all wholesale and purchased systems. This list of buyer-wholesaler relationships was from SDWIS/Fed, fourth quarter of 2010. EPA then created a crosswalk linking the purchased systems to the wholesale systems from which they purchased 100 percent of their water. The population served by each purchased system was then distributed evenly across the relevant wholesale system populations, according to the calculations described above. As a result, the contaminant occurrence measures are associated with the total (retail plus wholesale) population served by these non-purchased systems included in the Six-Year Review data.

7 Public Access to SYR3 ICR Data

Through extensive data management efforts and quality assurance evaluations, as well as through communications and consultations with state data management staffs, EPA established a high quality compliance monitoring dataset (the SYR3 ICR dataset) that consists of data from 54 states and primacy agencies (46 states plus data from Washington, D.C. and the tribes). The initial SYR3 ICR dataset included more than 47 million analytical records from approximately 167,000 PWSs that serve approximately 290 million people nationally.[6] More than two-thirds of these records (more than 33 million) were for contaminants (such as lead, copper and cVOCs) that were not analyzed as part of the SYR3 because of recent, ongoing or pending regulatory actions. More than 13 million analytical Chemical Phase Rule contaminants records underwent QA/QC review in order to be included in the SYR3 ICR dataset to support the SYR3 analyses in USEPA (2016a). 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 SYR3 ICR dataset comprise almost 13 million analytical records from approximately 139,000 PWSs that serve approximately 290 million people nationally.[7] (For details on the number of records removed via the QA/QC review for microbials or DBPs, refer to USEPA (2016b) and USEPA (2016c).)

EPA maintains the final SYR3 ICR compliance monitoring data online at: https://www.epa.gov/dwsixyearreview. The public can download the final SYR3 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 SYR3 ICR compliance monitoring and related data from EPA's website.

[6] This count of 167,000 PWSs represents all water systems with any SYR3 data (including purchased water systems). In this case, 290 million is the population served directly (retail) by these purchased and non-purchased systems.

[7] This count of 139,000 PWSs represents non-purchased systems only. The population served remains at 290 million; however, the number now reflects the total population served directly (retail) and indirectly (wholesale) by non-purchased systems only.

8 References

United States Environmental Protection Agency (USEPA). 2010. Agency Information Collection Activities; Submission to OMB for Review and Approval; Contaminant Occurrence Data in Support of EPA's Third Six Year Review of National Primary Drinking Water Regulations (Renewal). Notice: February 5, 2010, Volume 75, Number 24, Page 6023-6024.

USEPA. 2016a. The Analysis of Regulated Contaminant Occurrence Data from Public Water Systems in Support of the Third Six-Year Review of National Primary Drinking Water Regulations: Chemical Phase Rules and Radionuclides Rules. EPA-810-R-16-014. December 2016. December 2016.

USEPA. 2016b. Six- Year Review 3 Technical Support Document for Microbial Contaminant Regulations. EPA-810-R16-010. December 2016.

USEPA. 2016c. Six- Year Review 3 Technical Support Document for Disinfectants/Disinfection Byproducts Rules. EPA-810-R-16-012. December 2016.

USEPA. 2016d. Support Document for Third Six-Year Review of Drinking Water Regulations for Acrylamide and Epichlorohydrin. EPA- 810-R-16-019. December 2016.

The Data Management and Quality Assurance/Quality Control Process for the Third Six-Year Review Information Collection Rule Dataset: Appendices

List of Appendices

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

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

OFFICE OF WATER

Suffix First Last Name (Drinking Water Admin)
Title
Organization
Street 1
Street 2
City, State, Zip

Dear *Suffix Last Name*,

The 1996 Safe Drinking Water Act (SDWA) 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 third round of the Six-Year Review (Six-Year 3).

As was done for the second Six-Year Review, 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. This request for Six-Year 3 includes the following rules that were not part of Six-Year 2: the Ground Water Rule (GWR); Surface Water Treatment Rules (SWTRs); Disinfection Byproduct Rules (DBPRs); and, Filter Backwash Recycling Rule (FBRR). We are requesting data reflecting monitoring conducted between January 2006 and December 2011. The Office of Management and Budget (OMB) has approved the information collection request for EPA's third Six-Year Review under the provisions of the Paperwork Reduction Act, 44 U.S.C. 3501 et seq., and has assigned OMB control number 2040-0275.

These data are an important component in supporting EPA's Six-Year Review of NPDWRs. We are encouraging each State to submit its occurrence and treatment technique information, because these data will contribute directly to EPA's understanding of national contaminant occurrence, the population exposed to regulated contaminants, and exposure reductions associated with the current regulations. EPA is requesting your voluntary submission by October 31, 2012.

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. Attachment A, Exhibit 1 of this letter provides a list of the regulated contaminants for which EPA is requesting data. In Exhibit 2 of Attachment A, we identify the critical data elements needed for each sample result. To make your voluntary reporting as easy as possible, your State can transmit its compliance monitoring dataset to EPA by whatever electronic means is most convenient (see Attachment A for the data submission options). Attachment A 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.

Through our previous work on the Six-Year Review data collections, we have worked closely with data managers to work through data transfer and to answer questions. It is our understanding that <insert contact> is the current data manager in your program and, therefore, is copied on this request. Soon after October 22, 2012 we will begin contacting data managers and coordinating directly with them by phone and/or email. Please let us know if you prefer we work with another staff person.

Thank you for your consideration of this request. Many of you voluntarily submitted your data for Six-Year 2. 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 Karen Wirth at 202-564-5246 or wirth.karen@epa.gov.

Sincerely,

Pamela S. Barr
Acting Director, Office of Ground Water and Drinking Water

cc: <<data contact>>
Enclosure: Attachment A

ATTACHMENT A

I. Details Regarding EPA's Request for Occurrence 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. For Six-Year 3, this request also includes data collected for the following rules not included in Six-Year 2: the GWR, SWTRs, DBPRs, and FBRR. Exhibit 1, below, lists the specific contaminants for which EPA is requesting monitoring data. 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.

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

Exhibit 1: Occurrence Data Requested		
Gross alpha	Strontium-90	
Microbiological Contaminants & Surface Water Treatment Rules (SWTRs)[1]		
Total coliforms	Fecal coliforms	*Escherichia coli (E. coli)*
Chlorine	*Cryptosporidium*	Heterotrophic Plate Count (HPC)
Chloramines	*Giardia lamblia*	
Disinfectants and Disinfection Byproducts Rules (DBPRs)[2]		
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
Filter Backwash Recycling Rule (FBRR)		
No specific occurrence data collected; see Exhibit 2 for data elements for FBRR		

1. Including: Surface Water Treatment Rule (SWTR) (June 1989); Interim Enhanced SWTR (December 1998); Long-Term 1 Enhanced SWTR (January 2002); and, Long-Term 2 Enhanced SWTR (January 2006).
2. Including both Disinfection Byproducts/Treatment Rules: Stage 1 (December 1998) and Stage 2 (January 2006).

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

EPA is requesting the voluntary submission of occurrence data for regulated chemical, radiological, and microbiological contaminants (Exhibit 1) that reflect monitoring conducted between January 2006 and December 2011. 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 (pages A-3 to A-5) 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 dataset is incomplete, perhaps due to waivers and exemptions, etc.

Voluntary submission of your regulated drinking water contaminant occurrence and treatment technique data is the most critical step in this national occurrence assessment.

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.

Exhibit 2: Requested Data Categories	
Data Category	**Description**
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.)
Sanitary Survey Information	Site visit information for TCR, GWR, and SWTRs, including: site visit type, date completed, associated deficiencies identified, corrective actions taken.
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: coagulant/coagulant aid type and dose, disinfectant concentration (amounts, types, primary and secondary types of disinfection, disinfection profile/benchmark data), log of viral inactivation/removal, contact time, contact value, pH, temperature.
Filter Backwash Information	Information about filter backwash that is returned to the treatment plant influent (e.g., information on: recycle/schematic status, alternative return location, corrective action requirements, and recycle flows and frequency).
Sample-Specific Information	
Sampling Point Identification Code	A sampling point identifier established by the State, unique within each applicable facility, for each applicable sampling location (e.g., entry point to the distribution system). This information enables occurrence assessments that address intra-system variability.
Sample Identification Number	Identifier assigned by State or the laboratory that uniquely identifies a sample.
Sample Collection Date	Date the sample is collected, including month, day and year.
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.

Exhibit 2: Requested Data Categories	
Data Category	**Description**
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." *(Not required for TCR data)*
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, results will indicate presence/absence.*
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 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 data)*
MRL - Unit of Measure	Unit of measure to express the concentration value of a contaminant's MRL. *(Not required for TCR 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 (MOR)) summary records, alternative compliance criteria.
Sample Summary Reports	Sample summaries for DBPRs, SWTRs, TCR, and 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.]

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 occurrence and treatment technique data. EPA developed and refined a SDWIS/State extract 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 and transmittal for those States using some or all of SDWIS/State. Currently, some States do store and manage their data in more than one database. For data that is not stored in SDWIS/State, options also include submission through electronic file transfer protocol (FTP) or by mailing/shipping CDs/DVDs (see section D, below, for details).

1. Extracting data that is stored in SDWIS/State:

SDWIS/State Extract Tool: EPA has developed the SDWIS/State Extract Tool to pull the relevant data (specified in Exhibit 2, pages A-3 to A-5) from a SDWIS/State database. 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 is stored in SDWIS/State. For the data transfer step, please see the FTP paragraph within section D, below.

Note: If you have not migrated all drinking water monitoring data for the applicable period (January 2006 to December 2011) to SDWIS/State, a separate data submission to include all

data back to January 2006 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 Extract Tool:* EPA has built in several automated data QA checks with this extract tool. For example, the extract tool will check for duplicate data, and analytical results that are >10 times the MCL. Before the data is extracted from SDWIS/State, the extract 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 is 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 is submitted.

2. Format for Non-SDWIS/State data:

Virtually any electronic file format is acceptable. It would be ideal for States to submit their datasets in one of the following file formats: dBase™ (.dbf); Microsoft Access tables (.mdb); 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 they are currently stored in, and submitting that copy of the electronic data to EPA. If it is easier for you to provide your entire electronic dataset, 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 dataset, and definitions of all data elements, column/row headings, codes, acronyms, etc., used in each dataset. (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 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 were included with your transmittal.

D. How do I send my data to EPA?

For data that is not stored in SDWIS/State, options for sending your data to EPA include submission through electronic file transfer protocol (FTP) or by mailing/shipping CDs/DVDs.

1. FTP:

To ensure security of your data, each State will only have access to its own data on the FTP site, and, for further security, will be given usernames and passwords. In addition, datasets uploaded to the FTP site will be downloaded and removed within one working day of when they are uploaded and stored on a secured file server that is not accessible via the FTP site. For added security, you can zip the files with a password (so that they can only be unzipped with the password). If possible, please scan all files for viruses before uploading them.

If you would like to transfer your data via the FTP site, please email: SixYearData@cadmusgroup.com to receive instructions to access the FTP site and a username and password.

2. Shipping:

If you choose to send CDs/DVDs of your data, this can be sent via U.S. Postal Service or commercial air carrier (such as FedEx or UPS) to:

Six-Year Data Coordinator
The Cadmus Group, Inc.
100 Fifth Ave., Suite 100
Waltham, MA 02451-8727
Phone: 617-673-7000

E. When do these data need to be submitted?

To help EPA meet its Six-Year Review 3 statutory timeframe and to allow EPA time to compile, analyze and document the results of its review, EPA is asking that you please provide the requested datasets by October 31, 2012.

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 occurrence and treatment technique data for NPDWRs to support the third Six-Year Review. Through the Six-Year Review process, EPA reviews and assesses risks to human health posed by regulated drinking water contaminants, and drinking water occurrence and treatment technique data are critical to these assessments. Without an understanding of where and at what levels these contaminants are occurring in public drinking water, EPA cannot assess any potential risk to public health.

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 Office of Ground Water and Drinking Water 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 those 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 Six-Year Review 3, these data will be made publically available. The procedures used to analyze these data will reflect those established and refined for the first and second Six-Year Reviews. Copies of EPA's first and second Six-Year Review occurrence findings and methodology reports (*Occurrence Estimation Methodology and Occurrence Findings Report for the Six-Year Regulatory Review of Existing National Primary Drinking Water Regulations* (EPA 815-R-03-006) and *The Analysis of Regulated Contaminant Occurrence Data from Public Water Systems in Support of the Second Six-Year Review of National Primary Drinking Water Regulations* (EPA 815-B-09-006)) can be obtained at: http://water.epa.gov/lawsregs/rulesregs/regulatingcontaminants/sixyearreview/index.cfm. These documents contain the first and second 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 3 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 3 statistical occurrence analyses. A copy of the resulting final, QA/QCd datasets for your State will be made available to the public.

D. What will happen once the data are submitted?

EPA will conduct uniform QA/QC assessments on each dataset. 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 occurrence information provided by States for the first and second 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 3 decisions.

Treatment information – being collected for the first time under Six-Year Review 3 -- will also be compiled and assessed to support Six-Year Review 3 decisions. However, the format of this information does not lend itself to analogous quantitative analysis and national summaries. Rather, assessment of this information will be conducted and 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 assessments.

Appendix B: Crosswalk of Data Elements Requested for SYR3 ICR and the SDWIS Data Element Names

The table below is a crosswalk of the data elements requested in the SYR3 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 SYR3 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
Sanitary Survey Information	TINVISIT.Reason_CD; TINVISIT.Visit_Date; TINVISIT.HIGHEST_DEFICIENCY; TINVISIT.* (TENACTIV.NAME, TENCSHAT.ACHIEVED_DATE)
Treatment Information	
Water System Facility	tblSixYrWsf; [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

Data Category	SDWIS Mapping ([Table Name].[Data Element])
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 SYR3 SQL Database

This appendix contains 20 tables presenting the various tables and their data elements in the SYR3 Relational Database, along with all permitted values in those tables.

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

Field Name	Data Type	Description
SixYrWS_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	State in which the system is located using the states' two letter abbreviation.
NUMBER0	Text	Public water system identification number (PWSID)
NAME	Text	Water system name
D_POPULATION_COUNT	Number	Retail population served by the water system.
D_FED_PRIM_SRC_CD	Text	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
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
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
ACTIVITY_DATE	Text	For SDWIS/State states, the ACTIVITY_DATE is the date of the ACTIVITY_STATUS_CD. For non SDWIS/State states, it's the date that the water system was deactivated (if applicable).
STATE_CODE	Text	This field is used to identify the states in which tribal systems are located. State in which the system is located using the states' two letter abbreviation.
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.

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

Field Name	Data Type	Description
SixYrWsf_ID	Number	Unique identifier for each water system facility record.
SixYrWS_ID	Number	Identifier matching each record to tblSixYrWs
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that is unique when combined with TINWSF_ST_CODE.
TINWSF_ST_CODE	Text	State in which the facility is located using the states' two letter abbreviation.
ACTIVITY_STATUS_CD	Text	Activity status of the water system facility. A = Active; I = Inactive
ACTIVITY_DATE	Date/Time	For SDWIS/State states, the ACTIVITY_DATE is the date of the ACTIVITY_STATUS_CD. For non SDWIS/State 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.
TINWSF_NAME	Text	Name of the water system facility.
TYPE_CODE	Text	Type of the water system facility. CC = Consecutive Connection; CH = Common Headers; CW = Clear Well; DS = Distribution System; IG = Infiltration Gallery; IN = Intake; OT = Other; PC = Pressure Control; PF = Pumping Facility; RS = Reservoir; SI = Surface Impoundment; SP = Spring; SS = Sampling Station; ST = Storage; TM = Transmission Main (Manifold); TP = Treatment Plant; WH = Well Head; WL = Well; XX = unknown
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 tblSixYrSpt (sample point table)

Field Name	Data Type	Description
SixYrSpt_ID	Number	Unique identifier for each sample point record.
SixYrWsf_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWsf table.
SixYrWS_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWs table.
TINWSF0IS_NUMBER	Number	Identifier for each water system facility that is unique when combined with TINWSF_ST_CODE.
TINWSF0ST_CODE	Text	State in which the facility is located using the states' two letter abbreviation.
TSASMPPT_IS_NUMBER	Number	Identifier for each sample point that is unique when combined with TSASMPPT_ST_CODE.
TSASMPPT_ST_CODE	Text	Identifies the state in which the sample was taken using the states' two letter abbreviations.
TSASMPPT_TYPE_CODE	Text	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; WS = Water System Facility Point
SOURCE_TYPE_CODE	Text	The type of water source, based on whether treatment has taken place. FN = Finished, treated; RW = Raw, untreated; x = unknown

Field Name	Data Type	Description
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 tblAnalyte (analyte table)

Field Name	Data Type	Description
Analyte_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
AlternateName	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 tblSixYrSar (sample analytical result table)

Field Name	Data Type	Description
SixYrSar_ID	Number	Unique identifier for each sample analytical result record.
SixYrWS_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWs table.
SixYrWsf_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWsf table.
SixYrSpt_ID	Number	Identifier that relates each record to the unique record in the tblSixYrSpt table.
Analyte_ID	Number	Identifier that relates each record to the unique record in the tblAnalyte table.
TSASAR_IS_NUMBER	Number	Identifier for each sample analytical result that is unique when combined with TSASAR_ST_CODE.
TSASAR_ST_CODE	Text	State from which the data came using the states' two letter abbreviation.
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	State from which the data came using the states' two letter abbreviation.
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	State from which the data came using the states' two letter abbreviation.
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.

Field Name	Data Type	Description
TSASAMPL0ST_CODE	Text	State from which the data came using the states' two letter abbreviation.
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 CO = Confirmation; DU = Duplicate; FB = Field Blank; MR = Maximum Residence Time; MS = Matrix Spike; OT = Other; RP = Repeat; RT = Routine; RW = Raw Water; SB = Shipping Blank; SP = Special; TE = Technical Evaluation
REPEAT_LOC_TYP_CD	Text	The location of the repeat/check/confirmation sample with respect to the location of the original routine sample.
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 centrigrade or "F" for fahrenheit degrees.

Field Name	Data Type	Description
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]

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

Field Name	Data Type	Description
SixYrDbpSum_ID	Number	Unique identifier for each DBP summary record.
SixYrWS_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWs table.
SixYrSpt_ID	Number	Identifier that relates each record to the unique record in the tblSixYrSpt table.
SixYrFanls_ID	Number	Identifier that relates each record to the unique record in the tblSixYrFanls table.
TSAMDBPS_IS_NUMBER	Number	Identifier for each MDBP summary that must be combined with TSAMDBPS_ST_CODE when used.
TSAMDBPS_ST_CODE	Text	State in which the MDBP summary occurred using the states' two letter abbreviation.
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.

Field Name	Data Type	Description
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 in order 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

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

Field Name	Data Type	Description
SixYrFanls_ID	Number	Unique identifier for each facility analyte level record.
Analyte_ID	Number	Identifier that relates each record to the unique record in the tblAnalyte 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	State in which the system is located using the states' two letter abbreviation.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that must be combined with TINWSF_ST_CODE when used.
TINWSF_ST_CODE	Text	State in which the facility is located using the states' two letter abbreviation.
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.
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.
SAMPLE_RQT_PER_DAY	Number	The number of samples that must be collected during a twenty four 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.

Field Name	Data Type	Description
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.

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

Field Name	Data Type	Description
SixYrSampSum_ID	Number	Unique identifier for each sample summary record.
Analyte_ID	Number	Identifier that relates each record to the unique record in the tblAnalyte 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	State for each sample summary result using the states' two letter abbreviation.
TSASMPSM_IS_NUMBER	Number	Identifier for each sample summary that must be combined with TSASMPSM_ST_CODE when used.
TSASMPSM_ST_CODE	Text	State for each sample summary using the states' two letter abbreviation.
TINWSYS_IS_NUMBER	Number	Identifier for each water system that must be combined with TINWSYS_ST_CODE when used.
TINWSYS_ST_CODE	Text	State in which the system is located using the states' two letter abbreviation.
TINWSF_IS_NUMBER	Number	Identifier for each water system facility that must be combined with TINWSF_ST_CODE when used.
TINWSF_ST_CODE	Text	State in which the facility is located using the states' two letter abbreviation.
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.
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 above 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
MEASURE	Number	The calculated value of the results represented in the sample summary defined by the sample summary's TYPE_CODE.
UOM_CODE	Text	The unit of measure (UOM) that is associated with the value reported for the sample summary measure.

Exhibit C.9: Description of tblSixYrSaniSur (sanitary survey table)

Field Name	Data Type	Description
SixYrSaniSur_ID	Number	Unique identifier for each sanitary survey record.
SixYrWS_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWs table.

Field Name	Data Type	Description
TINVISIT_IS_NUMBER	Number	Identifier for each site visit that must be combined with TINVISIT_ST_CODE when used.
TINVISIT_ST_CODE	Text	State in which the site visit occurred using the states' two letter abbreviation.
STATUS	Text	Status: C = completed; P = planned
VISIT_DATE	Date/Time	The date on which the Site Visit was made to the water system.
DUE_DATE	Date/Time	The anticipated date by which this site visit should occur.
REASON_CD	Text	Code that represents the reason for which a Site Visit was made to a public water system. SNSV = Sanitary Survey
FREQUENCY_NUMBER	Number	Frequency for the specified period.
FREQUENCY_PERIOD	Text	Period associated with the specified frequency number. DY = Day(s); MN = Month(s); WK = Week(s); YR = Year(s)
NEXT_DUE_DATE	Date/Time	Date the next Site Visit is due.
HIGHEST_DEFICIENCY	Text	Highest level of deficiency for the Site Visit SIG = Significant; NON = No deficiencies; REC = Recommendation made; MIN = Minor
SS_EL_SOURCE	Text	Source -- one of the eight elements in EPA/State Joint Guidance on Sanitary Surveys.
SS_EL_TREATMENT	Text	Treatment -- one of the eight elements in EPA/State Joint Guidance on Sanitary Surveys.
SS_EL_DISTRIB_SYS	Text	Distribution System -- one of the eight elements in EPA/State Joint Guidance on Sanitary Surveys.
SS_EL_FIN_WTR_STRG	Text	Finished Water Storage -- one of the eight elements in EPA/State Joint Guidance on Sanitary Surveys.
SS_EL_PUMPS	Text	Pumps (facilities, controls, etc.) -- one of the eight elements in EPA/State Joint Guidance on Sanitary Surveys.
SS_EL_MR_DV	Text	Monitoring and Reporting (M&R) and Data Verification (DV) -- one of the eight elements in EPA/State Joint Guidance on Sanitary Surveys.
SS_EL_WS_MGT_OPS	Text	Water System Management and Operations -- one of the eight elements in EPA/State Joint Guidance on Sanitary Surveys.
SS_EL_OP_COMP_EVAL	Text	Operator Compliance Evaluation -- one of the eight elements in EPA/State Joint Guidance on Sanitary Surveys.
SS_EL_SECURITY	Text	Security -- a coded value that describes in summary, the outcome of evaluating this category during the Site Visit. The "Public Health Security and Bioterrorism Preparedness and Response Act of 2002" requires primacy agencies to review the security and preparedness of water system to respond to emergencies. Permitted values will be the same as the existing categories, including spaces.
SS_EL_FINANCIAL	Text	Financial -- a coded value that describes in summary, the outcome of evaluating this category during the Site Visit. The Safe Drinking Water Act (SDWA) Amendments of 1996 requires primacy agencies to assist small water systems through Capacity Development.
SS_EL_OTHER	Text	Other -- value that can be set in addition to the eight elements in EPA/State Joint Guidance on Sanitary Surveys. Default is Not Evaluated.
COMMENT_TEXT	Text	Additional information that the Inspector wishes to record about the site visit.

Exhibit C.10: Description of tblSixYrSanSurvDef (sanitary survey deficiencies table)

Field Name	Data Type	Description
SixYrSanSurvDef_ID	Number	Unique identifier for each sanitary survey deficiency record.

Field Name	Data Type	Description
SixYrWS_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWs table.
SixYrSaniSur_ID	Number	Identifier that relates each record to the unique record in the tblSixYrSaniSur table.
TINDEFCY_IS_NUMBER	Number	Identifier for each sanitary survey deficiency that must be combined with TINSVDFA_TINVISIT_ST_CODE when used.
TINSVDFA_TINVISIT_ST_CODE	Text	State in which the site visit occurred using the states' two letter abbreviation.
VISIT_DATE	Date/Time	A value that represents the calendar date on which a visit was made to a PWS.
REASON_CD	Text	Code that represents the reason for which a Site Visit was made to a public water system. SNSV = Sanitary Survey
SEVERITY	Text	The type of deficiency: SIG = Significant; REC = Recommendation made; MIN = Minor
SANITARY_SRVEY_CAT	Text	Categorizes the deficiency into one of the ten category evaluation summaries during the Site Visit: the eight sanitary survey elements identified by the EPA/State Joint Guidance on Sanitary Surveys (i.e., "DS," "FW," "MR," "OC," "PU," "SM," "SO," and "TR"), plus the two elements required by the Public Health Security and Bioterrorism Preparedness and Response Act of 2002, and the SDWA Amendments of 1996 (i.e., "SE" and "FI"). "Other" or "OT" is a catch-all category. "Unknown" is included to enable the migration and storage of historical deficiencies as well as new ones which have not yet been classified. DS = Distribution System; FI = Financial; FW = Finished Water Storage; MR = Monitoring and Requirements (M&R)/Data Verification; OC = Operator Compliance with State Requirements; OT = Other; PU = Pump/Pumping Facility & Control; SE = Security; SM = System Management & Operation; SO = Source; TR = Treatment; UK = Unknown
DETERMINATION_DATE	Date/Time	The actual date the deficiency was determined if different from the VISIT_DATE.
DESCRIPTION_CV	Text	Four- character alphabetic code representing descriptions of the deficiency that may be controlled by the System Administrator. Values are stored in the Permitted Values table in the System Administration component.
TINDFTYP_DESCRIPTION_TXT	Text	Free text description of the Deficiency type.
RESOLVED_DATE	Date/Time	The date the deficiency was resolved.
COMMENTS	Text	A field where CDS Compliance Report processes can record any additional information that may be useful when a user is determining what action to take relative to the candidate violation.
TINDEFCY_DESCRIPTION_TXT	Text	Free text description of the Deficiency

Exhibit C.11: Description of tblSixYrSSCorAct (Sanitary survey corrective actions table)

Field Name	Data Type	Description
SixYrSSCorAct_ID	Number	Unique identifier for each corrective action record.
SixYrWS_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWs table.
TINVISIT_IS_NUMBER	Number	Identifier for each site visit record that must be combined with TINVISIT_ST_CODE when used.
TINVISIT_ST_CODE	Text	State in which the site visit occurred using the states' two letter abbreviation.
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	State in which the compliance schedule is relevant using the states' two letter abbreviation.

Field Name	Data Type	Description
TYPE_CODE_CV	Text	Activity type code. Permitted values are established by primacy agencies.
EFFECTIVE_DATE	Date/Time	A value that represents the calendar date on which a variance, exemption, or other event became, or will become, effective.
STATUS_CODE	Text	(F)inal, (P)roposed, (S)uperceded. This value will be used to populate the Status of the Compliance Schedule that is associated to the Site Visit.
STATUS_DATE	Date/Time	The date of the last status code update.
CLOSED_DATE	Date/Time	Date the compliance schedule was closed.
DESCRIP_TXT	Text	Narrative information about the activity type.
DESCRIPTION	Text	A description of the measure type.
VISIT_DATE	Date/Time	The date on which the Site Visit was made to the water system.
REASON_CD	Text	Code that represents the reason for which a Site Visit was made to a public water system.

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

Field Name	Data Type	Description
SixYrWsfPlt_ID	Number	Unique identifier for each treatment plant water system facility record.
SixYrWsf_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWsf table.
ST_ASGN_IDENT_CD	Text	A state-assigned value which identifies the treatment plant water system facility.
TYPE_CODE	Text	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	Text	(Unfiltered (UF), Conventional Filtration (CF), Direct Filtration (DF), Diatomaceous Earth (DE), Other(OT), and other permitted values that the System Administrator may add)
DESCRIPTION	Text	A description of the filter.
DISINFECT_CONCENTN	Text	Disinfectant Concentration in mg/L
CONTACT_TIME_STAT	Text	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
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 in order 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.

Field Name	Data Type	Description
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.
FBR_CORCTV_ACT_DT	Date/Time	The date that the water system achieved the corrective action required for the filter backwash.
D_INITIAL_USERID	Text	The User ID of the person who created this record.
FBR_COMMENTS	Text	A memo field into which a user may enter comments about the Filter Backwash Recycled 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 tblTreatProcess (Treatments associated to treatment plants table)

Field Name	Data Type	Description
TreatProcess_ID	Number	Unique identifier for each treatment record.
SixYrWsf_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWsf 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.

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

Field Name	Data Type	Description
wsfFlows_ID	Number	Unique identifier for each water system facility flow record.
SixYrWsf_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWsf table.
TINWSFF_IS_NUMBER	Number	Identifier for each water system facility flow entry that is unique when combined with SixYrWsf_ID.
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.
TINWSF0IS_NUMBER	Number	Identifier for each supplying water system facility that is unique when combined with TINWSF0ST_CODE.
TINWSF0ST_CODE	Text	State in which the supplying facility is located using the states' two letter abbreviation.

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

Field Name	Data Type	Description
WsfInd_ID	Number	Unique identifier for each water system facility indicator record.
SixYrWsf_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWsf table.
TINWSFIN_IS_NUMBER	Number	Identifier for each water system facility indicator that is unique when combined with SixYrWsf_ID.
INDICATOR_NAME	Text	The water system facility indicator name.
DESCRIPTION	Text	The description of the water system facility indicator name.
INDICATOR_VALUE_CD	Text	The value of the indicator established by the primacy agency.
INDICATOR_DATE	Date/Time	The date associated with the indicator.

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

Field Name	Data Type	Description
WsInd_ID	Number	Unique identifier for each water system indicator record.
SixYrWS_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWs table.
TINWSIN_IS_NUMBER	Number	Identifier for each water system indicator that is unique when combined with SixYrWS_ID.
INDICATOR_NAME	Text	The water system indicator name.
DESCRIPTION	Text	The description of the water system indicator name.
INDICATOR_VALUE_CD	Text	The value of the indicator established by the primacy agency.
INDICATOR_DATE	Date/Time	The date associated with the indicator.

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

Field Name	Data Type	Description
WsPurch_ID	Number	Unique identifier for each water system buyer and seller record.
SixYrWS_ID	Number	Identifier that relates each record to the unique record in the tblSixYrWs table.
TINWSYS0IS_NUMBER	Number	Identifier for each supplying water system that is unique when combined with TINWSYS0ST_CODE.
TINWSYS0ST_CODE	Text	State in which the supplying water system is located using the states' two letter abbreviation.
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.
TINWSF_ST_CODE	Text	State in which the facility is located using the states' two letter abbreviation.
TINWSF0IS_NUMBER	Number	Identifier for each supplying water system facility record that must be combined with TINWSF0ST_CODE when used.
TINWSF0ST_CODE	Text	State in which the supplying facility is located using the states' two letter abbreviation.

Exhibit C.18: Description of lkp_SixYrSar_Transaction_QAFlag (Transaction QA Flag – Lookup Table)

Field Name	Data Type	Description
uid	Number	QA Flag ID (number 1 through 17) to identify the reason the record was flagged.
QA_FLAG	Text	Text describing the QA flag. 1: Duplicate 2: Transient (i.e., transient system collected contaminant result for which it wasn't required) 3: Non-compliance result (i.e., record identified as not being for compliance) 4: Non-routine result (i.e., sample type code is something other than routine, confirmation, repeat, or maximum residence (MR; appropriate for DBPs only)) 5: GT 4XMCL (i.e., detected concentration is greater than 4 times the contaminant's MCL or MRDL (for disinfectants))

Field Name	Data Type	Description
		6: GT 10XMCL (i.e., detected concentration is greater than 10 times the contaminant's MCL or MRDL (for disinfectants)) 7: LT MDL (i.e., detected concentration is less than the contaminant's Minimum Detection Limit) 8: LT 1/10MDL (i.e., detected concentration is less than one-tenth (1/10) the contaminant's Minimum Detection Limit) 9: UNITS (i.e., detected concentration is expressed in an erroneous unit of measure) 10: Purchased Water Systems (i.e., purchased water system collected contaminant result for which it wasn't required) 11: Outside Date Range (i.e., sample was collected prior to 1/1/2006 or after 12/31/2011) 12: Non-Public Water System (i.e., sample was collected by a non-public water system) 13: Missing Inventory Data (i.e., system doesn't have any associated inventory data in tblSixYrWs table) 14: Convert (for CA nitrate data; detected concentrations were converted to Nitrate (as N)) 15: Raw (raw water results) 16: Formerly Purchased (i.e., results from systems that were originally thought to be 100 percent purchased but were later determined not to be) 17: Rad-NTNC (i.e., non-transient system collected radionuclide data for which it wasn't required)
Active	Yes/No	Indicates whether the action is active or not.

Exhibit C.19: Description of lkp_SixYrSar_Transaction_Action (Transaction Action – Lookup Table)

Field Name	Data Type	Description
uid	Number	Action ID (number 1 through 4) to identify the action necessary for each flagged record.
Action	Text	Text describing how the QA issue will be resolved. 1: No change; 2: Change; 3: Exclude; 4: On hold
Active	Yes/No	Indicates whether the action is active or not.

Exhibit C.20: Description of tblSixYrSar_Transaction (Transaction Table)

Field Name	Data Type	Description
TransactionID	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.)
SixYrSar_ID	Number	Unique identifier for each sample analytical result (enables linking to tblSixYrSar).
TSASAR_IS_NUMBER	Number	Identifier for each sample analytical result that is unique when combined with TSASAR_ST_CODE.
TSASAR_ST_CODE	Text	State from which the data came using the states' two letter abbreviation.
QA_FLAG_ID	Number	A coded value (1 through 17) that identifies the reason that the record was flagged. See "lkp_SixYrSar_Transaction_QAFlag" for a definition of the 17 codes.
Action_ID	Number	A coded value (1 through 4) that identifies the reason that the record was flagged. See "lkp_SixYrSar_Transaction_Action" for a definition of the 4 codes.
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.

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

Appendix D: Guide to the QA/QC of the Fluoride SYR3 ICR Dataset

The SYR3 ICR dataset for fluoride underwent a separate QA/QC review than the rest of the chemical phase and radionuclide contaminants to identify and exclude fluoride samples from fluoridated water systems from the SYR3 occurrence analyses. An overview the fluoride QA/QC review is included in this appendix.

The original fluoride dataset used in support of the occurrence analysis was originally maintained in the following four tables:

- Water System Table (tblSixYrWs) – provides system information such as PWSID, source water type, system type, and population.
- Water System Facility Table (tblSixYrWsf) - contains facility-level information such as facility ID and facility type.
- Sample Point Table (tblSixYrSpt) - contains sampling information such as sample point type and source type.
- Sample Analytical Result Table (tblSixYrSar) – contains monitoring records such as sample date, sample type code, analyte, concentration, and reporting level.

Each table contains its own unique identifier (e.g., water system ID, water system facility ID, etc.) and the monitoring data table (tblSixYrSar) contains references to the unique identifiers of each of the other tables so that monitoring results can be matched with sample-point, facility, and system information.

In cases where the VALUE field in the tblSixYrSar table was incomplete, it was populated using the following logic:

> For all non-detections (i.e., [DETECT] = 0), [VALUE] was set equal to [DETECTN_LIMIT_NUM] if [CONCENTRATION_MSR] = 0 or null. If [CONCENTRATION_MSR] > 0 and [DETECT] = 0, then [VALUE] = [CONCENTRATION_MSR].
>
> For all detections (i.e., [DETECT] = 1), [VALUE] = [CONCENTRATION_MSR].

In the fluoride dataset, VALUE was only populated for the detections. All of the non-detections' values and units were blank. Therefore, EPA implemented the procedures outlined above to generate VALUE field entries for non-detections. EPA also standardized the reporting units for fluoride (e.g., converting micrograms to milligrams).

Cleaning Procedure

The following steps provide details on the 10 queries used in the fluoride QA/QC review process:

Query 1: Create Fluoride_Orig table by combining relevant fields from the four original data tables, then append to a blank Fluoride table with standard column headings (standard column

headings are found in the table "z Occurrence-Fields-Blank;" open this table and save it as "Fluoride" to create a blank table for Query 001b to function). Fields to include are:

tblSixYrWs	STATE_CODE, NUMBER0, NAME, ADJUSTED_TOTAL_POPULATION, D_PWS_FED_TYPE_CD
tblSixYrSpt	SOURCE_TYPE_CODE, TSASMPPT_TYPE_CODE
tblSixYrWsf	ST_ASGN_IDENT_CD, TINWSF_NAME, TYPE_CODE
tblSixYrSar	SixYrSar_ID, TSASAMPL_IS_NUMBER, LAB_ASGND_ID_NUM, COLLLECTION_END_DT, TSASAMPL_TYPE_CODE, DETECT, VALUE, UNIT, Analyte_ID

Query 2: Update concentration values and units for non-detections, following the procedure mentioned above. Update the blank "Value" column the non-detect values converted to mg/L. Replace blank and zero values with the mean non-detections values for the same systems, if available. For blank and zero values without within-system values, update using the state specific MRL values.

Query 3: The water system table (tblSixYrWs) classifies the water system type into the following four categories:

C = Community water system
NC = Non-community water system
NTNC = Non-transient non-community water system
NP = Non-public water system

Tag all systems classified as "NP," "NC," or with a blank system type as a PWSTYPE exclusion. These system types were consistent when compared to SDWIS/FED classifications.

Query 4: Identify low and high outliers to be excluded from dataset. Consistent with past occurrence analysis:

- Low outliers for detects and non-detects are values below the lowest water MDL. The lowest MDL for fluoride is 0.002 µg/L
- High outliers for detects are any value 10x greater than the current MCL
- High outliers for non-detects are any value greater than the current MCL

Query 5: Perform cleaning procedure to identify duplicates consistent with past occurrence analysis. Identify additional duplicates flagged in the original dataset (tblSixYrSar_Transaction table).

Query 6: Update size category using the following thresholds:

1: <=100
2: 101 - 500
3: 501 - 1,000
4: 1,001 - 3,300
5: 3,301 - 10,000
6: 10,001 - 50,000
7: 50,001 - 100,000
8: 100,001 - 1,000,000
9: >1,000,000

Query 7: Exclude applicable flagged data from the original database. The following lists the types of samples that were flagged in a table named "tblSixYrSar_Transaction," which is in the original data:

1: Duplicate
2: Transient (i.e., transient system collected contaminant result for which it wasn't required)
3: Non-compliance result (i.e., record identified as not being for compliance)
4: Non-routine result (i.e., sample type code is something other than routine, confirmation, or maximum residence (MR; appropriate for DBPs only))
5: GT 4XMCL (i.e., detected concentration is greater than 4 times the contaminant's MCL)
6: GT 10XMCL (i.e., detected concentration is greater than 10 times the contaminant's MCL)
7: LT MDL (i.e., detected concentration is less than the contaminant's Minimum Detection Limit)
8: LT 1/10MDL (i.e., detected concentration is less than one-tenth (1/10) the contaminant's Minimum Detection Limit)
9: UNITS (i.e., detected concentration is expressed in an erroneous unit of measure)
10: Purchased Water Systems (i.e., purchased water system collected contaminant result for which it wasn't required)
11. Outside Date Range (i.e., sample was collected prior to 1/1/2006 or after 12/31/2011)
12: Non-Public Water System (i.e., sample was collected by a non-public water system)
13: Missing Inventory Data (i.e., system doesn't have any associated inventory data in tblSixYrWs table)
14: Convert (for CA nitrate data; detected concentrations were converted to Nitrate (as N))

Of these categories, "duplicates" was used previously. The remaining categories whose flagged samples should be excluded from the occurrence dataset are: transients, non-routine, non-compliance, nonpublic, date outlier, and missing inventory data.

Additionally, tag all purchased water systems for exclusion. These systems have source water (SRCWATER) values classified as "GWP," "SWP," or "GUP."

Query 8: Identify original water samples. The sample point table (tblSixYrSpt) contains two different columns where original water samples are potentially identified:

TSASMPPT_TYPE_CODE- 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
WS = Water System Facility Point

SOURCE_TYPE_CODE The type of water source
FN = Finished, treated
RW = Raw, untreated
x = unknown

Comparing data from these columns show a lot of inconsistency in characterization of source water samples in the database. The state SDWIS data located in the system facility table (tblSixYrWsf) is a more dependable starting point for identifying source water data. The following macros identify and tag source water data:

Query 8a: Set all source water status to 0.

Query 8b: Convert blank SDWIS type codes to "BL."

The SDWIS dataset (fourth quarter 2010 SDWIS/Fed freeze) contains the following two tables needed to identify source water samples that are 'upstream' of samples taken at a treatment or distribution entry point:

dbo_FacilityFlow – Table that shows relationship between facility flows

dbo_WaterSystemFacility – Table of all system facilities, needed to link stated assigned facility identifiers available in SYR3 ICR3 with facility numbers in the facility flow table

Queries in the fourth quarter 2010 SDWIS/Fed freeze create a table called "FacilityFlow," which must be exported into the ICR dataset (in this example the Fluoride_6YR3 dataset) before running the next query (008c).

Query 8c: EPA identified excluded source water samples using the Facility Flow table exported from a fourth quarter 2010 SDWIS/FED freeze. This table identifies source water facilities as

those with the most commonly occurring source water identifiers in Type_Code [i.e., "IN" (intake) or "RS"(reservoir) or "SP" (spring) or "WL" (well)]. Using the facility FacIDFrom column in dbo_FacilityFlow, we identify all source water facilities that occur in the fluoride dataset. Using the FacIDTo column, we identify treated water facilities as the most commonly occurring facility types associated with treated water [i.e., Type_Code of "TP" (treatment plant) or "DS" (distribution system) or "CW" (clear well)]. The treated water samples tags come from dbo_WaterSystemFacility so that they are tagged regardless of whether the treatment facility appears in the fluoride dataset. The resulting table (FacilityFlow) identifies possible source water samples for exclusion as those tagged as a source water facility that flows to a treated water facility.

Query 8d: Create system table with counts of total samples and counts of source water samples. After creating the table called raw_water_table, open the table and create a new column called 'All raw' and set the data type to yes/no. The next query needs this field to run properly.

Query 8e-f: Exclude all source water samples for systems that also provide downstream treated water samples.

As an example, below are three sampling points for PWS 041200003 in the fluoride dataset.

PWSID	WATER TYPE	SAMPLETYPE	STATE_ID	STATE_ASSIGNED_NAME	TYPE_ CODE
041200003	FN	EP	201	TREATMENT PLANT #1	TP
041200003	RW	RW	101	WEST WELL #1	WL
041200003	RW	RW	104	NEW EAST WELL #4	WL

Based on the FacilityFlow information below, we can tag both wells as source water facilities that occur upstream of a treated water facility (TRUE values in FromSourceWater field). Because the treated water facility is in the fluoride database, we tag the samples for the wells as source water samples to exclude.

FacStateID-From	FacIDFrom	FromSourceWatrer	FacIDTo	ToTreatedWater
104	10810	TRUE	10996	TRUE
101	10969	TRUE	10996	TRUE
201	10996	FALSE	10953	TRUE

Query 9: Create entry point IDs. Following the naming procedures used in past occurrence analysis, create entry point IDs for all non-excluded data-points.

Query 10: Create Fluoride_Final table using all non-excluded data.

Appendix E: User Guide to Downloading and Using SYR3 and Related Data from EPA's Website

This appendix includes a copy of the user guide to downloading and using the SYR3 and related data from EPA's website. This document is also posted online with the data.

Note: Reference citations in this User Guide differ from those in "The Data Management and Quality Assurance/Quality Control Process for the Third Six-Year Review Information Collection Rule Dataset."

User Guide to Downloading and Using SYR3 Data from EPA's Website

To support the national contaminant occurrence and exposure assessments performed under the Six-Year Review process, EPA analyzes compliance monitoring data 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 to help the Agency determine if there may be a meaningful opportunity to improve public health protection. EPA conducted a voluntary data request from the states and primacy agencies to obtain the compliance monitoring data necessary to analyze national contaminant occurrence in support of the third Six-Year Review (SYR3). This data request was conducted through the Information Collection Request (ICR) process. EPA requested that states and primacy agencies submit their SDWA compliance monitoring data collected between January 2006 and December 2011. For more information on the process undertaken to request the voluntary submission of compliance monitoring data by the states, see the third Six-Year Review ICR renewal (75 FR 6023, USEPA, 2010).

Through extensive data management efforts, quality assurance evaluations, and communications and consultations with state data management staff, EPA established a single contaminant occurrence dataset that consists of compliance monitoring data from 54 out of 67 states/primacy agencies (46 states plus Washington, D.C. and the tribal data). This dataset is referred to as the National Compliance Monitoring ICR Dataset for the third Six-Year Review (or "SYR3 ICR Dataset"). The 54 states/primacy agencies that provided data for the SYR3 ICR Dataset comprise 95 percent of all PWSs and 92 percent of the total population served by PWSs nationally, and are geographically representative of PWSs nationwide. The SYR3 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.

The SYR3 ICR Dataset is the largest, most comprehensive set of drinking water compliance monitoring data ever compiled and analyzed by EPA to inform decision making. EPA conducted a quality control evaluation of these data submitted by states and other primacy agencies, and assembled these data into a database. The database is more than twice the size of the one collected to support of the Second Six-Year Review (SYR2) with more than 47 million records from approximately 167,000 public water systems, serving approximately 290 million people nationally. The dataset includes the results of all compliance monitoring data (all sample analytical detections and non-detections) from January 2006 to December 2011 for regulated chemical phase contaminants, radionuclides, disinfectants and disinfection byproducts (D/DBPs), DBP precursors, microbial contaminants, disinfectant residuals and treatment information. Note that only the data that passed the QA/QC process are posted online.

Additional reference material is available to assist with the assessment of the SYR3 data.

- EPA's Six-Year Review website
- The Data Management and Quality Assurance/Quality Control Process for the Third Six-Year Review Information Collection Rule Dataset (USEPA, 2016a)

The data are posted online in several zip files. Each zip file includes text files for multiple contaminants/parameters. The number of records and contaminants/parameters included in each file vary. The remainder of this document is organized as follows:

- Section 1 describes the data being posted for phase chemicals, radionuclides and disinfection byproducts.
- Section 2 describes the data being posted for disinfection byproduct precursors.

- Section 3 describes the data being posted for microbial contaminants and associated disinfectant residuals.
- Section 4 describes data being posted for additional parameters.
- Section 5 describes the treatment data being posted.
- Section 6 describes the data quality considerations of the SYR3 ICR data.
- Section 7 describes supplemental data sources being posted.

Section 1: Phase Chemicals, Radionuclides and Disinfection Byproducts

Exhibit 1 contains a list of the data elements, column names and a brief description of the data for each data element included in each of the SYR3 ICR text files for the individual phase chemicals, radionuclides and disinfection byproducts.

Exhibit 1: Six-Year 3 Data Field Names and Definitions

Data Element	Column Name	Description
Contaminant Identification Code	Analyte ID	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.
State Code	State Code	2- digit state code. Note that the state code "IM" refers to non-community water system data from the State of Illinois.
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 Population Served	Retail population served by a system.
Adjusted Total Population-served	Adjusted Total Population Served	Total population served by a system, adjusted to reduce double-counting of population served by purchasing water systems.
Source Water Type	Source Water Type	Type of water at the source. Source water type can be: • Ground water (GW); • Surface water (SW); • Purchased Surface Water (SWP); • Purchased Ground Water (GWP); • Ground Water Under Direct Influence of Surface Water (GU); or • Purchased Ground Water Under Direct Influence of Surface Water (GUP).
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; • 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

Data Element	Column Name	Description
		• 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; or • RT = Routine.
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 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).

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

Summary of SYR3 Phase Chemicals, Radionuclides and Disinfection Byproduct Data

Exhibit 2 provides a count of states, total number of sample records and systems for each phase chemical, radionuclide and disinfection byproduct whose data is posted online. The user may want to compare their counts of records downloaded for each contaminant of interest to this table 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.

Exhibit 2: Six-Year Review 3 Data Summary for Contaminants/Parameters

Contaminant	Analyte ID	Number of States with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Phase Chemicals					
1,1,1-Trichloroethane	2981	50	374,181	55,735	SYR3_PhaseChem_1
1,1,2-Trichloroethane	2985	50	371,877	55,733	SYR3_PhaseChem_1
1,1-Dichloroethylene	2977	50	379,522	55,728	SYR3_PhaseChem_1
1,2,4-Trichlorobenzene	2378	50	369,032	55,725	SYR3_PhaseChem_1
1,2-Dibromo-3-chloropropane (DBCP)	2931	50	188,597	37,226	SYR3_PhaseChem_1
2,3,7,8-TCDD (Dioxin)	2063	30	20,244	3,216	SYR3_PhaseChem_1
2,4,5-Trichlorophenoxypropionic Acid (Silvex)	2110	50	126,887	36,897	SYR3_PhaseChem_1
2,4-Dichlorophenoxyacetic acid (2,4-D)	2105	50	131,047	37,690	SYR3_PhaseChem_1
Alachlor	2051	50	153,083	42,955	SYR3_PhaseChem_1
Antimony	1074	49	164,961	50,532	SYR3_PhaseChem_1
Arsenic	1005	50	297,354	54,845	SYR3_PhaseChem_1
Asbestos	1094	39	12,084	5,785	SYR3_PhaseChem_1
Atrazine	2050	50	162,134	44,310	SYR3_PhaseChem_1
Barium	1010	49	165,387	50,711	SYR3_PhaseChem_2
Benzo(a)pyrene	2306	50	131,437	34,341	SYR3_PhaseChem_2
Beryllium	1075	49	164,392	50,195	SYR3_PhaseChem_2
Cadmium	1015	49	165,247	50,583	SYR3_PhaseChem_2
Carbofuran	2046	50	122,110	34,614	SYR3_PhaseChem_2
Chlordane	2959	49	128,870	35,685	SYR3_PhaseChem_2
Chromium (Total)	1020	49	167,251	50,597	SYR3_PhaseChem_2
cis-1,2-Dichloroethylene	2380	50	376,300	55,734	SYR3_PhaseChem_2

Contaminant	Analyte ID	Number of States with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Cyanide	1024	49	119,659	36,907	SYR3_PhaseChem_2
Dalapon	2031	49	146,702	36,005	SYR3_PhaseChem_2
Di(2-ethylhexyl)adipate (DEHA)	2035	50	133,169	34,628	SYR3_PhaseChem_2
Di(2-ethylhexyl)phthalate (DEHP)	2039	49	133,523	33,923	SYR3_PhaseChem_2
Dinoseb	2041	50	126,014	36,701	SYR3_PhaseChem_2
Diquat	2032	46	69,829	17,906	SYR3_PhaseChem_2
Endothall	2033	45	61,972	15,538	SYR3_PhaseChem_3
Endrin	2005	50	136,623	38,453	SYR3_PhaseChem_3
Ethylbenzene	2992	50	372,709	55,754	SYR3_PhaseChem_3
Ethylene Dibromide (EDB)	2946	49	184,784	37,499	SYR3_PhaseChem_3
Fluoride	1025	49	256,237	47,227	SYR3_PhaseChem_3
Glyphosate	2034	45	70,016	18,502	SYR3_PhaseChem_3
Heptachlor	2065	50	137,286	38,691	SYR3_PhaseChem_3
Heptachlor Epoxide	2067	50	137,081	38,625	SYR3_PhaseChem_3
Hexachlorobenzene	2274	50	137,816	38,498	SYR3_PhaseChem_3
Hexachlorocyclopentadiene	2042	50	140,004	38,743	SYR3_PhaseChem_3
Lindane (gamma-Hexachlorocyclohexane)	2010	50	139,076	39,260	SYR3_PhaseChem_3
Mercury (Inorganic)	1035	49	164,558	50,552	SYR3_PhaseChem_3
Methoxychlor	2015	50	139,744	39,187	SYR3_PhaseChem_3
Monochlorobenzene	2989	50	371,311	55,676	SYR3_PhaseChem_3
Nitrate (as N)	1040	49	1,157,522	132,176	SYR3_PhaseChem_3
Nitrite (as N)	1041	49	445,544	85,742	SYR3_PhaseChem_3
o-Dichlorobenzene	2968	50	370,929	55,732	SYR3_PhaseChem_4
Oxamyl (Vydate)	2036	50	121,508	34,518	SYR3_PhaseChem_4
p-Dichlorobenzene	2969	50	371,276	55,739	SYR3_PhaseChem_4
Pentachlorophenol	2326	50	140,486	40,322	SYR3_PhaseChem_4
Picloram	2040	50	128,401	37,445	SYR3_PhaseChem_4
Polychlorinated biphenyls (PCBs)	2383	44	86,405	21,571	SYR3_PhaseChem_4
Selenium	1045	49	165,672	50,568	SYR3_PhaseChem_4
Simazine	2037	50	156,862	43,240	SYR3_PhaseChem_4
Styrene	2996	50	370,368	55,731	SYR3_PhaseChem_4
Thallium	1085	49	164,156	50,522	SYR3_PhaseChem_4
Toluene	2991	50	373,021	55,748	SYR3_PhaseChem_4
Toxaphene	2020	49	127,187	37,043	SYR3_PhaseChem_4
trans-1,2-Dichloroethylene	2979	50	371,580	55,633	SYR3_PhaseChem_4
Xylenes (Total)	2955	50	323,477	51,074	SYR3_PhaseChem_4
Radionuclides					
Alpha Particles	4000	47	60,803	13,309	SYR3_Rads
Beta Particles	4100	41	43,278	11,531	SYR3_Rads
Combined Radium-226 & -228	4010	42	73,018	15,805	SYR3_Rads

Contaminant	Analyte ID	Number of States with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Uranium	4006	49	86,208	12,155	SYR3_Rads
Disinfection Byproducts					
Total Trihalomethanes	2950	46	532,002	36,691	SYR3_THM
Bromoform	2942	42	433,636	34,788	SYR3_THM
Chloroform	2941	42	434,624	34,839	SYR3_THM
Bromodichloromethane	2943	42	433,663	34,815	SYR3_THM
Dibromochloromethane	2944	42	433,141	34,735	SYR3_THM
Total Haloacetic Acids	2456	45	475,592	33,518	SYR3_HAA
Monochloroacetic acid	2450	36	283,260	25,202	SYR3_HAA
Dichloroacetic acid	2451	36	282,778	25,221	SYR3_HAA
Trichloroacetic acid	2452	36	282,732	25,213	SYR3_HAA
Monobromoacetic acid	2453	36	282,799	25,196	SYR3_HAA
Dibromoacetic acid	2454	36	282,986	25,210	SYR3_HAA
Bromate	1011	29	8,884	222	SYR3_Bromate_Chlorite
Chlorite	1009	28	25,989	220	SYR3_Bromate_Chlorite

Section 2: Disinfection Byproduct Precursors

Data for three disinfection byproduct precursors are being posted online: total organic carbon (TOC), alkalinity and pH. In addition to the "full" datasets for TOC and alkalinity, a "paired" TOC dataset was created that included, for each treatment plant, the average monthly concentrations of TOC and alkalinity in source (raw) water paired with the corresponding average finished water concentration of TOC. The "paired" TOC dataset was used to evaluate the percent removal of TOC using the SYR3 data; see Chapter 7 and Appendix C in USEPA (2016d) for more details on the "paired" TOC dataset.

Exhibit 3 contains the list of data elements, column names, and a brief description of the data for each data element included in the "paired" TOC dataset. For a list of data elements included in the "full" TOC, alkalinity and pH datasets, refer to Exhibit 1.

Exhibit 3: SYR3 "Paired" TOC Dataset Field Names and Definitions

Data Element	Column Name	Description
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.
Sample Collection Date (Month)	Month	Month (1 through 12).
Sample Collection Date (Year)	Year	Year (2006 through 2011).
Retail Population-served	Retail 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 Code	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.

User Guide to Downloading and Using SYR3 Data from EPA's Website

Summary of SYR3 Disinfection Byproduct Precursor Data

Exhibit 4 provides a count of states, total number of sample records and systems for TOC, alkalinity and pH.

Exhibit 4: Six-Year Review 3 Data Summary for TOC, Alkalinity and pH

Contaminant	Analyte ID	Number of States with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Disinfection Byproduct Precursors - Full Datasets					
Total Organic Carbon	2920	32	232,567	2,836	SYR3_Precursors
Alkalinity	1927	38	201,682	15,059	SYR3_Precursors
pH	1925	40	208,203	25,509	SYR3_Precursors
Disinfection Byproduct Precursors - Reduced Dataset					
Paired TOC-alkalinity dataset[1]	N/A	22	65,771	1,208	SYR3_PairedTOC-Alkalinity

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

Section 3: Microbials and Associated Disinfectant Residuals

Summary of SYR3 Microbial and Residual Data

Data for three microbial contaminants (total coliforms, *E. coli*, and fecal coliform) and associated disinfectant residual data are being posted online. A "full" dataset includes all data for total coliforms (TC), *E. coli* (EC), and fecal coliform (FC) and associated disinfectant residual data (when available) that have passed the initial QA process. A "reduced" dataset includes a subset of the data for disinfecting systems with disinfectant residual. These data were used to support the analyses in USEPA (2016c). Only the data with paired chlorine residual concentrations (free and/or total chlorine) were included in the analysis; thus, these TC-EC-FC data represent only a subset of all total coliform results submitted via the SYR3 ICR. See Appendix A in USEPA (2016c) for details on the QA/QC documentation for both the full and the reduced microbial datasets.

For a list of data elements included in the full TC, EC, and FC datasets, refer to Exhibit 1. For a list of data elements included in the Reduced Dataset for Analysis of Disinfecting Systems with Disinfectant Residuals, refer to Exhibit 5.

Exhibit 5: SYR3 Reduced Dataset for Analysis of Disinfecting Systems with Disinfectant Residuals - Field Names and Definitions

Data Element	Column Name	Description
Contaminant Identification Code	Analyte ID	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.
State Code	State Code	2- digit state code. Note that the state code "IM" refers to non-community water system data from the State of Illinois.
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 Population Served	Retail population served by a system.
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: DS = Distribution System.
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;

Data Element	Column Name	Description
		• FC = First Customer; • FN = Finished Water Source; • MD = Midpoint in the Distribution System; • MR = Point of Maximum Residence; • RW = Raw Water Source; • SR = Source Water Point; 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: • RP = Repeat; or • RT = Routine.
Six Year ID	Six Year ID	Unique identifier for each analytical result.
Sample Collection Date	Sample Collection Date	Date the sample was collected, including month, day, and year.
Presence Indicator Code	Presence Indicator Code	Indication of whether results of an analysis were positive or negative for TC, EC and FC. • P = Presence • A = Absence.
Residual Field Free Chlorine	Residual Field Free Chlorine mg/L	Amount of free chlorine residual (in mg/L) found in the water after disinfectant has been applied. These concentrations were measured in the field at the same time and location as coliform samples (TC-EC-FC samples).
Residual Field Total 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).

Exhibit 6 provides a count of states, total number of sample records and systems for total coliform, *E. coli*, fecal coliform, and their associated free and total chlorine residual concentrations for both the full and reduced datasets.

Exhibit 6: Six-Year Review 3 Data Summary for Microbials and Associated Disinfectant Residuals

Contaminant	Analyte ID	Number of States with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Microbials and Residuals – Full Datasets					
Total coliform	3100	46	9,766,686	113,548	SYR3_TC-DR-06-08; SYR3_TC-DR-09-11
E. coli	3014	44	1,804,329	55,509	SYR3_EC-FC-DR
Fecal coliform	3013	39	264,090	17,821	SYR3_EC-FC-DR
Microbials and Residuals - Reduced Dataset					
Total coliform	3100	41	4,750,432	36,753	SYR3_Microbes_DR

Contaminant	Analyte ID	Number of States with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
E. coli	3014	35	889,570	18,896	SYR3_Microbes_DR
Fecal coliform	3013	25	64,304	2,986	SYR3_Microbes_DR
Field free chlorine residual[1]	N/A	--	4,007,235	33,054	SYR3_Microbes_DR
Field total chlorine residual[1]	N/A	--	2,521,771	17,757	SYR3_Microbes_DR

[1] Measured in the field at the same time and location as coliform samples were collected.

Summary of Reduced Dataset for Analysis of Undisinfected Ground Water Systems

Data for total coliforms, *E. coli*, and fecal coliform paired with system disinfection status are also posted online. To simplify statistical modeling of the TC, EC, and FC data for that analysis, the data for each system and month were reduced to a small number of summary counts: (a) the total number of routine samples assayed, (b) the number of routine samples testing positive for TC, (c) the total number of TC positive routine samples tested for EC and (d) the number of routine samples testing positive for EC. Rather than include a record for each sample assayed, the reduced dataset includes, for each water system and month, counts of the routine and repeat samples for TC, EC and FC. (See Exhibit 7.) In the final "reduced" dataset, there are data for a total of 80,692 water systems from 39 states/entities. (The zip file containing these data is "SYR3_Microbes_GW.") See Appendix D in USEPA (2016c) for details on the steps used to produce this reduced dataset.

A subset of these data were used to represent "undisinfected" ground water systems. In this analysis, "undisinfected" ground water systems referred to those that do not practice disinfection or have very low disinfectant residuals (i.e., less than 0.1 mg/L). These data were used to support additional analyses in USEPA (2016c). See Appendix F in USEPA (2016c) for details on the analysis of undisinfected ground water systems.

Exhibit 7: SYR3 Reduced Dataset for Analysis of Undisinfected Ground Water Systems - Field Names and Definitions

Data Element	Column Name	Description
Public Water System Identification Number (PWSID)	PWSID	Public water system identification number (PWSID).
Sample Collection Date (Month)	Month	Month (1 through 12).
Sample Collection Date (Year)	Year	Year (2006 through 2011).
Retail Population-served	Retail 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 (GW-SW)	Primary water source for the water system. GW = Ground Water (also includes Purchased GW)

Data Element	Column Name	Description
		SW = Surface Water (also includes Purchased SW; Ground water Under Direct Influence of SW; and Purchased Ground Water Under Direct Influence of SW)
Disinfection Status	Disinfecting?	An indication if the system disinfects its water (Y = Yes; blank = No). All systems with a source water type = "SW" were assumed to be disinfecting. Note: An explanation of the determination of the ground water systems' disinfection status is included on pages 2 and 3 of this document.
Count Routine TC samples	TC Samples (routine)	The count of routine total coliform (TC) samples.
Count Routine TC+ samples	TC+ Samples (routine)	The count of routine TC positive samples.
Count Routine EC samples	EC Samples (routine)	The count of routine *E. coli* (EC) samples.
Count Routine EC+ samples	EC+ Samples (routine)	The count of routine EC positive samples.
Count Routine FC samples	FC Samples (routine)	The count of routine fecal coliform (FC) samples.
Count Routine FC+ samples	FC+ Samples (routine)	The count of routine FC positive samples.
Count Repeat TC samples	TC Samples (repeat)	The count of repeat TC samples.
Count Repeat TC+ samples	TC+ Samples (repeat)	The count of repeat TC positive samples.
Count Repeat EC samples	EC Samples (repeat)	The count of repeat EC samples.
Count Repeat EC+ samples	EC+ Samples (repeat)	The count of repeat EC positive samples.
Count Repeat FC samples	FC Samples (repeat)	The count of repeat FC samples.
Count Repeat FC+ samples	FC+ Samples (repeat)	The count of repeat FC positive samples.

Section 4: Additional Parameters

Data for 11 additional parameters have been provided; however, these parameters did not undergo the same quality assurance evaluations as the parameters that were analyzed as part of the SYR3 process. For more information on the quality assurance evaluations performed for these parameters, see USEPA (2016a). Exhibit 8 provides a count of states, total number of sample records and systems for the additional parameters whose data are being posted online. For a list of data elements included in the data posted online for these additional parameters, refer to Exhibit 1.

Exhibit 8: Six-Year Review 3 Data Summary for Additional Parameters

Parameter	Analyte ID	Number of States with Data	Total Number of Sample Records	Total Number of Systems	Zip Filename
Additional Parameters[1]					
Heterotrophic bacteria	3001	18	48,908	797	SYR3_AdditionalAnalytes
Enterococci	3002	2	9	3	SYR3_AdditionalAnalytes
Giardia lamblia	3008	5	426	42	SYR3_AdditionalAnalytes
Chlorine[2]	0999	11	1,505,286	3,673	SYR3_AdditionalAnalytes
Chloramine[2]	1006	5	58,012	474	SYR3_AdditionalAnalytes
Chlorine dioxide	1008	10	7,181	22	SYR3_AdditionalAnalytes
Residual chlorine[2]	1012	3	70,582	1,081	SYR3_AdditionalAnalytes
Free residual chlorine data[2]	1013	1	5,852	741	SYR3_AdditionalAnalytes
SUVA	2923	2	2,447	34	SYR3_AdditionalAnalytes
UV-254	2922	2	2,010	31	SYR3_AdditionalAnalytes
DOC	2919	4	16,669	163	SYR3_AdditionalAnalytes

[1] Coliphage was requested in the SYR3 ICR, however, no coliphage records passed the quality assurance evaluation.
[2] Reported independently of the coliform sample results.

Section 5: Treatment Data

Exhibits 9 and 10 provide a comprehensive summary of the data elements included in the treatment information within the SYR3 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 analytes that were analyzed as part of the SYR3 process.

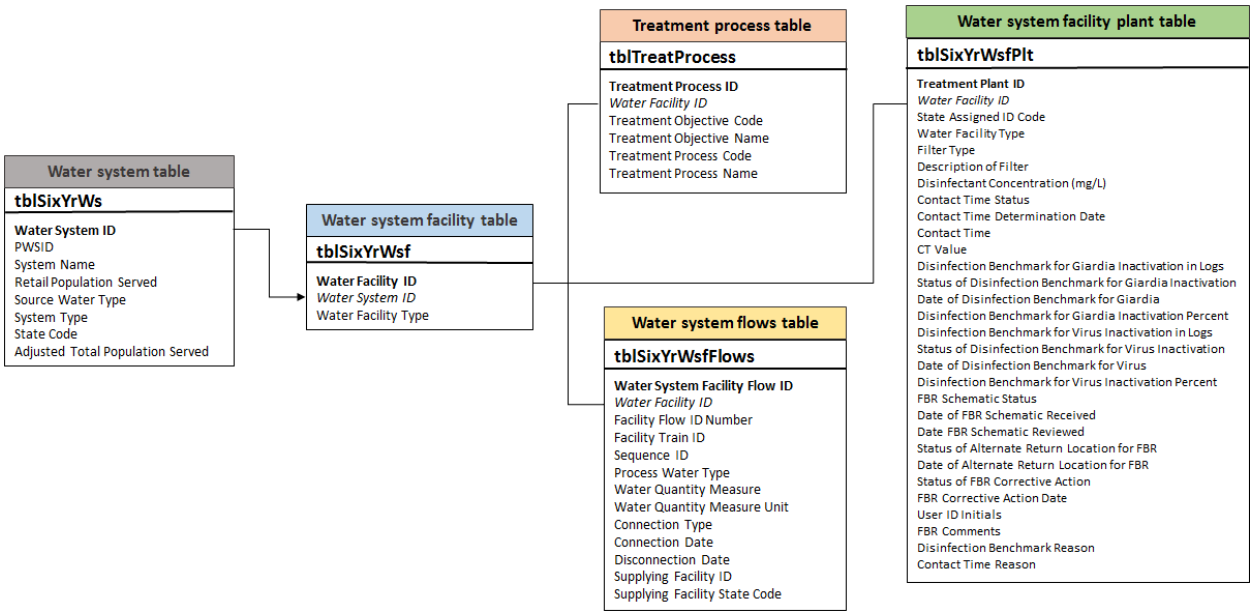
Exhibit 9 identifies the data elements used in the treatment information tables and a description of each data element. However, a majority of these data elements are not populated. Exhibit 10 represents the database relationships between tables in the SYR3 ICR treatment database. This diagram shows 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 (connection) between two or more related tables. Relationships between key fields in different tables are illustrated by the lines connecting the tables.

Exhibit 9: Treatment Data Dictionary (Filename: SYR3_Treatment)

Data Element	Description
	Water system facility plant table (tblSixYrWsfPlt)
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 tblSixYrWsf 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.

Data Element	Description
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.
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 (tblTreatProcess)	
Treatment Process ID	Unique identifier for each treatment record.
Water Facility ID	Identifier that relates each record to the unique record in the tblSixYrWsf 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 (tblSixYrWsfFlows)	
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 tblSixYrWsf table.
Facility Flow ID Number	Identifier for each water system facility flow entry that is unique when combined with SixYrWsf_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 Units	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	State in which the supplying facility is located using the states' two letter abbreviation.

Website Exhibit 10: Treatment Data Diagram



User Guide to Downloading and Using SYR3 Data from EPA's Website

Section 6: SYR3 Data Considerations

The SYR3 ICR data is of reasonable quality and is representative and appropriate for use to support national, scientifically-defensible findings. Data has undergone appropriate quality assurance evaluation and enough states provided compliance monitoring to be representative for national-scale analyses. EPA used the data in analytical activities informing decisions for Six Year 3. The data include sufficient information for users to be able to reproduce the SYR3 analyses.

There are a few limitations of the final SYR3 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 1 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 SYR3 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 SYR3 ICR dataset. Furthermore, there were four states and some other tribes/territories whose data are missing entirely from the analysis.

A thorough QA/QC process was undertaken to evaluate these SYR3 ICR data used for analyses. However, it is possible that data entry errors may still exist in the final SYR3 ICR Dataset. The QA/QC review focused only on the data elements essential for analysis.

For a complete discussion of the SYR3 ICR dataset, including a description of the quality assurance/quality control review, refer to USEPA (2016a) and USEPA (2016b). For more detailed information on the microbial contaminants' occurrence analysis, refer to USEPA (2016c). For more detailed information on the occurrence analysis of contaminants regulated under the Stage 1 and Stage 2 Disinfectant and Disinfection Byproducts Rules, refer to USEPA (2016d).

Instructions on Importing SYR3 Datasets to Excel

These text files are tab delimited and have no text qualifier. Field names are included in the first row of each file. A basic understanding of Microsoft Excel is necessary to effectively use these instructions. Using Microsoft Excel 2013 or a newer version is recommended due to the size of the dataset(s). Note, however, that the complete SYR3 ICR Dataset is too large to be imported into Excel. The data are available for download for each parameter and should be imported into a data management system that supports large datasets for analysis.

Part One: Downloading and Importing Data (Note that instructions may vary depending on the version and software used to import data.)

1. Begin by reviewing the SYR3 ICR Dataset Summary (Exhibit 2) and in particular note the table of Data Field Names and Definitions (Exhibit 1).

2. Access the SYR3 ICR data by going to the Six-Year Review homepage. Click on the link for "Six-Year Review 3."

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 unzip or extract the zip file contents by clicking **Open with** and using Win Zip or Microsoft Compression.

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 unzipped or extracted in Step 4, and click **Import** on the text file that of interest.

9. The Text Import Wizard – Step 1 of 3 will appear. The default settings will be displayed and should have **Delimited** selected as the **Original data type**. Select the checkmark box next to **My data has headers.** Click **Next>**.

10. The Text Import Wizard – Step 2 of 3 will appear. The default settings will be displayed and should have **Tab** selected as the **Delimiter** while **Treat consecutive delimiters as one** should be **unselected**. Select **Text qualifier** as **{none}** from the dropdown menu. Click **Next>.**

11. The Text Import Wizard – Step 3 of 3 will appear. The default settings will be displayed and will specify each column data format a**s General**. Click **Finish**. See #18 for further details about formatting.

12. The **Import Data** prompt will appear. Click **OK**. This import may take several minutes.

13. Save the Excel spreadsheet file.

Part Two: Filtering and Formatting Data in Excel

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

15. 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."

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

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

18. 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:

a. **Text** for: 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.
b. **Number** for: Analyte ID, Retail Population Served, Adjusted Total Population Served, Water Facility ID, Sampling Point ID, Six Year ID, Sample ID, Detection Limit Value, Detect, Value, Residual Field Free Chlorine mg/L, Residual Field Total Chlorine mg/L.

Section 7: Supplemental Data Sources

Several supplemental data sources were used to support the national contaminant occurrence and exposure assessments performed under the Six-Year Review process. These supplemental data sources are described below.

Disinfection Byproducts (DBP) Information Collection Rule (ICR) (Filename: DBPICR_Aux1)

The DBP ICR "Aux 1" database houses monitoring data from large public water systems (PWSs serving a population greater than or equal to 100,000) from the 18-month period of July 1997 to December 1998. A total of 296 water systems reported data; included in the database are monitoring results for microbials and DBPs, plant treatment, source water characteristics and disinfectant type information. This database was previously used in the development of the Stage 2 D/DBPR. Refer to McGuire et al (2002) for additional information.

For the SYR3 review, this database was used for several purposes, including the following: to investigate changes in disinfection practices; to evaluate changes in DBP precursor occurrence and removal; and to evaluate chlorate occurrence and co-occurrence of chlorate and chlorite. Refer to USEPA (2016d) and USEPA (2016f) for additional information.

Within the "Aux 1" version of the database, there are 31 relational tables within the database, plus several other tables providing additional information such as descriptions of each table, data element, attribute, etc.

The DBP ICR (Aux 1) database is posted online in Microsoft Access. The data documentation file is posted alongside the data. This documentation explains to the user all of the various data elements and tables included in the database.

EPA ICR Treatment Study Database (TSD) (Filename: ICR_TSD)

The ICR TSD was constructed to manage the treatment study data submitted by the systems required to conduct DBP precursor removal studies under the 1996 ICR. Results from 99 treatment studies (63 granular activated carbon (GAC) and 36 membrane studies), are reported in this database. This database was previously used in the development of the Stage 2 D/DBPR. Refer to McGuire et al. (2002) for additional information.

For the SYR3 review, this database was further used to evaluate the reduction of brominated DBP formation by GAC. Refer to USEPA (2016d) for additional information.

The TSD is posted online in Microsoft Access. There is a data documentation file (entitled "TS Database User's Guide") posted alongside the data to provide an explanation to the user all of the various data elements and tables included in the database.

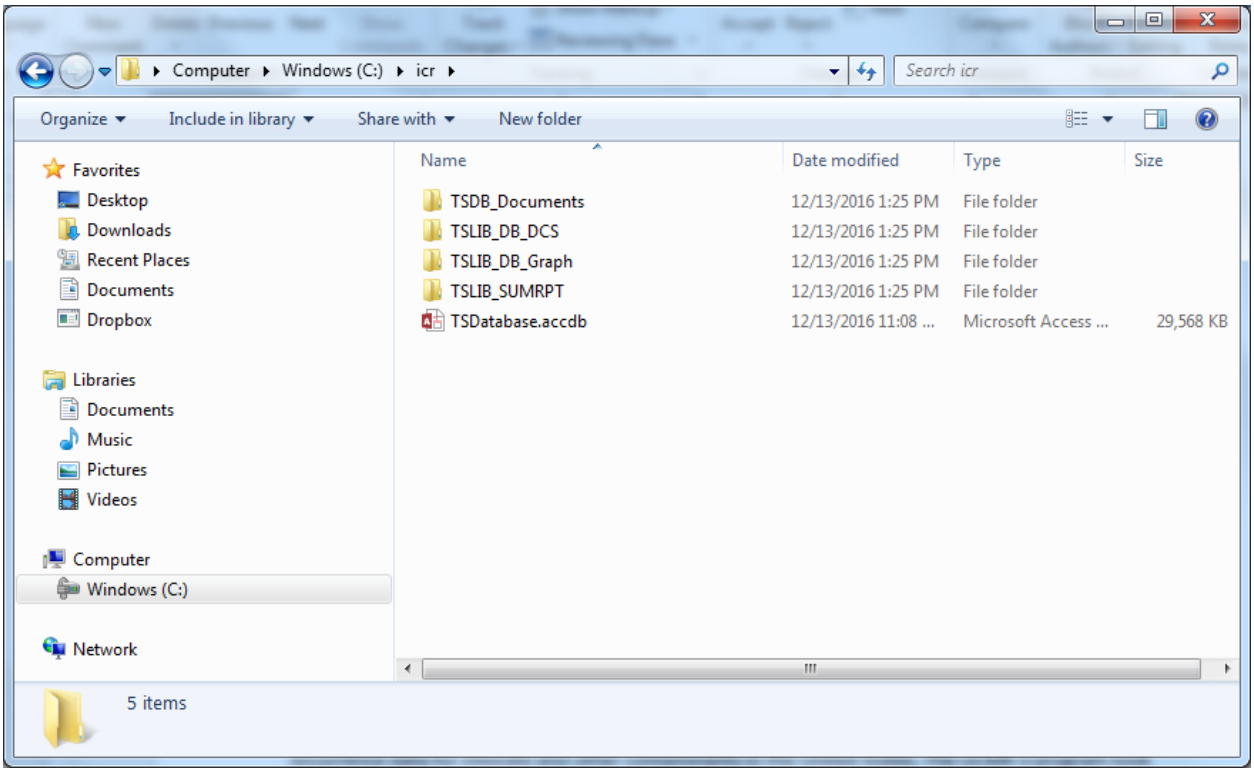
TSD files posted online:

1. ***TSDatabase.accdb*** (the TSD Access database file) – 28 MB
2. ***TSDB_Documents***: Includes pdf documents that users access from the database's "Documentation" section
    a. *BenchPilotManual*: ICR Manual for Bench- and Pilot-Scale Treatment Studies

   b. *DataSprdShtMnl*: ICR TS Data collection spreadsheets User's Guide
   c. *GAC Base Analysis Doc*: Base Analysis Document: GAC Studies
   d. *Membrane Base Analysis Doc*: Base Analysis Document: Membrane Studies
   e. *TS Database User's Guide*: Treatment Study Database User's Guide
3. ***TSLIB_DB_DCS***: Excel Data Collection Spreadsheets for all samples
4. ***TSLIB_DB_Graph***: PDF Graphical Summary Files for all samples
5. ***TSLIB_SumRpt***: PDF Summary Reports for all samples

Structure of TSD Files for Posted Online:

1. Download and save ***ICR_TSD*** to your local hard drive in "C:\".
2. Extract the files from ***ICR_TSD*** and rename the destination folder as "C:\icr". See screenshot below for an example of the structure and location of files once the data have been extracted and saved locally.



Second Unregulated Contaminant Monitoring Rule (UCMR 2) Data

Data are available for nitrosamine occurrence in finished drinking water in public water systems (PWSs) from the nationally representative monitoring completed under the Second Unregulated Contaminant Monitoring Rule (UCMR 2). UCMR 2 monitoring included monitoring for all six nitrosamines discussed in the SYR3 nitrosamine support document (USEPA, 2016e): N-nitrosodi-n butylamine (NDBA), N nitrosodiethylamine (NDEA), N nitrosodimethylamine (NDMA), N nitrosodi-n propylamine (NDPA), N nitrosomethylethylamine (NMEA) and N nitrosopyrrolidine (NPYR).

UCMR 2 monitoring, conducted between January 2008 and December 2010, provided data about nitrosamine occurrence; these data are available from the agency's website (https://www.epa.gov/dwucmr/occurrence-data-unregulated-contaminant-monitoring-rule#2).

Third Unregulated Contaminant Monitoring Rule (UCMR 3) – July 2016 version (Filename: UCMR3_July2016)

The data available for chlorate occurrence in finished drinking water in PWSs are from the nationally representative monitoring completed under the third round of the Unregulated Contaminant Monitoring Rule (UCMR 3). The UCMR 3 monitoring provides nationally representative contaminant occurrence data for chlorate and other contaminants in the United States. The UCMR 3 program took place from 2012 to 2015.

The UCMR 3 occurrence analyses presented in SYR3 chlorate support document (USEPA, 2016f) are based on data collected through May 2016 and released in July 2016 (USEPA, 2016g). EPA expects a relatively small amount of data reporting to continue after July 2016. The UCMR 3 dataset will not be considered "final" until early 2017. EPA does not anticipate that there will be any substantial difference between findings based on the July 2016 dataset and findings based on the final dataset.

Safe Drinking Water Information System (SDWIS) Information

The Safe Drinking Water Information System (SDWIS) contains information about public water systems and their violations of EPA's drinking water regulations, as reported to EPA by the states. Several versions of SDWIS datasets were used to support the national contaminant occurrence and exposure assessments performed for SYR3. This section provides the applicable SDWIS dataset file names on EPA's occurrence data webpage, and describes how these data were used for SYR3.

Note that the varying activity issues in the SDWIS datasets described below could cause confusion about the understanding of the data being presented. For example, there are active and inactive systems, non-public systems, systems that have merged with other systems and potential future systems included in the SDWIS datasets. The inactive, non-public and potential future systems were not used in the occurrence analyses but are included in the data posted online. There are also systems that have been inactive for many years.

SDWIS 2011 Pivot Tables (Filename: SDWIS2011_Pivot)

SDWIS inventory data were used to assess representativeness of SYR3 ICR data on both state and national levels. This is discussed further in chapter 6 of the D/DBPR support document (USEPA, 2016d). Note: the data within this file represents data ending in FY 2013. The file does contain information from 2010 to 2013; however, only the 2011 data were used for this analysis.

SDWIS Violation Data (Filename: SDWISViolations_2006-2011)

SDWIS violation data were used to assess violation rates and representativeness of populations. EPA conducted this assessment for the IOCs, SOCs, VOCs, and radionuclides.

2011 SDWIS/FED Freeze (Filename: SDWIS2011_Freeze)

A SDWIS/FED freeze from December 2011 was used to populate missing inventory information (e.g., source water type or population served) for some of the non-SDWIS states. This version of SDWIS was also used to evaluate the completeness of the data submitted for SYR3.

Note that Safe Drinking Water Information System (SDWIS) Quarterly Freeze is a copy of the data contained in SDWIS as of a specific year and quarter and includes all information available in the system at that time.

2010 SDWIS/FED Freeze (Filename: SDWIS2010_Freeze)

A SDWIS/FED freeze from December 2010 was used to identify the system type and for the national extrapolation of small system occurrence data for chlorate. Refer to the SYR3 chlorate support document (USEPA, 2016f) for additional information.

SDWIS Buyers-Sellers (Filename: SDWISBuyers_Sellers)

A list of buyer-wholesaler relationships from a fourth quarter 2010 SDWIS/FED freeze was used to adjust the population values of the wholesale systems to include the population of the systems that they sell water to (the purchased water systems). Refer to "The Analysis of Regulated Contaminant Occurrence Data from Public Water Systems in Support of the Third Six-Year Review of National Primary Drinking Water Regulations: Chemical Phase Rules and Radionuclides Rules" (USEPA, 2016b) for additional information.

2005 SDWIS Freeze (Filename: SDWIS2005_Freeze)

A 2005 SDWIS freeze was used in the occurrence analyses of nitrosamines to categorize PWSs by their source water type and by the size of the population served. Refer to the SYR3 nitrosamine support document (USEPA, 2016e) for additional information.

LT2 Round 1 Monitoring Data

In support of its LT2 analyses, EPA used data from the Data Collection and Tracking System (DCTS) pull from April 2012, which contained 44,944 records representing all system sizes. EPA posted the original and "cleaned-up" datasets on the EPA website at: https://www.epa.gov/dwsixyearreview/long-term-2-enhanced-surface-water-treatment-lt2-rule-round-1-source-water. Refer to the LT2 support document (USEPA, 2016h) for additional information.

References

McGuire, M.J., J.L. McLain, and A. Obolensky (eds.). 2002. *Information Collection Rule Data Analysis*. Denver, CO: American Water Works Research Foundation Research Foundation and American Water Works Association, 600 p. Available online at http://www.waterrf.org/PublicReportLibrary/90947.pdf.

United States Environmental Protection Agency (USEPA). 2010. Agency Information Collection Activities; Submission to OMB for Review and Approval; Contaminant Occurrence Data in Support of EPA's Third Six Year Review of National Primary Drinking Water Regulations (Renewal). Notice: February 5, 2010, Volume 75, Number 24, Page 6023-6024.

USEPA. 2016a. The Data Management and Quality Assurance/Quality Control Process for the Third Six-Year Review Information Collection Rule Dataset. EPA-810-R-16-015. December 2016.

USEPA. 2016b. The Analysis of Regulated Contaminant Occurrence Data from Public Water Systems in Support of the Third Six-Year Review of National Primary Drinking Water Regulations: Chemical Phase Rules and Radionuclides Rules. EPA 810-R-16-014. December 2016.

USEPA. 2016c. Six- Year Review 3 Technical Support Document for Microbial Contaminant Regulations. EPA-810-R16-010. December 2016.

USEPA. 2016d. Six- Year Review 3 Technical Support Document for Disinfectants/Disinfection Byproducts Rules. EPA-810-R-16-012. December 2016.

USEPA. 2016e. Six- Year Review 3 Technical Support Document for Nitrosamines. EPA-810-R-16-009. December 2016.

USEPA. 2016f. Six- Year Review 3 Technical Support Document for Chlorate. EPA-810-R-16-013. December 2016.

USEPA. 2016g. Third Unregulated Contaminant Monitoring Rule Dataset. July, 2016 version.

USEPA. 2016h. Six- Year Review 3 Technical Support Document for Long- Term 2 Enhanced Surface Water Treatment Rule. EPA-810-R-16-011.