SDWIS Data Quality Status

USEPA Region 9

March 2016

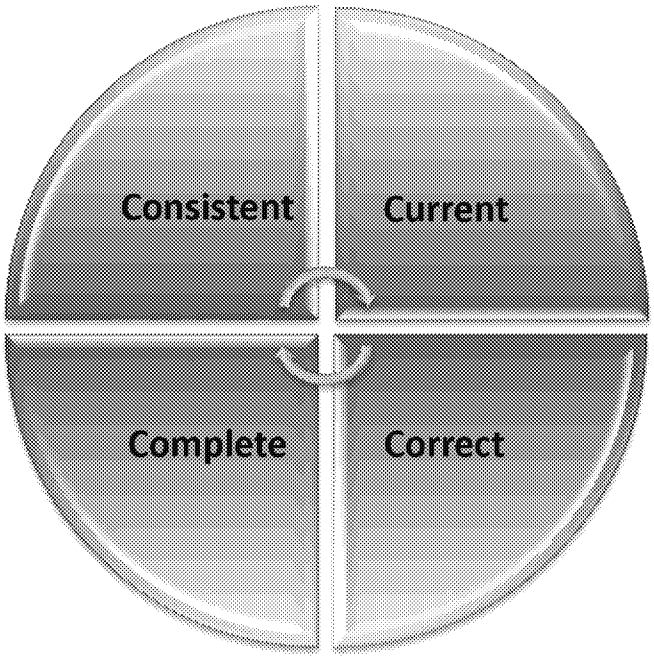

Table of Contents

[TOC \o "1-3" \h \z \u]

Introduction

Purpose of this Report

This purpose of this Safe Drinking Water System Information System (SDWIS) data quality status report is threefold:

- To present the current status of data reporting to the national drinking water database, SDWIS/Federal version (SDWIS/Fed), for each primacy agency in Region 9
- To recommend actions to improve the overall quality of data reported to SDWIS/Fed
- To inform end-of-year program review and upcoming year grant guidance discussions on agency-specific data management strengths and challenges

Data Quality

For the purposes of data reported to SDWIS/Fed, the term *data quality* encompasses the following four elements:

- Accuracy – the reported information accurately represents public water system status
- Completeness – all required data elements are reported for all regulated water systems
- Timeliness – violations are reported to SDWIS/Fed within a quarter of their determination
- Consistency – reported information reflects a consistent application of drinking water regulations

How to Use this Report

Charts, spreadsheets, and summary tables in this report present the current status of data reported to SDWIS/Fed. Each section includes *Next Steps* for addressing identified data quality issues.

Region 9 recognizes that no two primacy agencies are alike and that each faces unique data management challenges. The Region's goal is to collaborate with each agency to prioritize data management issues and to manage data quality improvement efforts.

ED_006532_00002047-00003

I. Database Management System Utilization

Database and Application Implementation

Region 9 primacy agencies use two database applications to manage their drinking water programs: SDWIS/State (developed by EPA Headquarters) and the Direct Implementation and Management Environment, DIME (developed by Region 9). (In addition, many of California's counties that serve as local primacy agencies use a local government automation system to oversee drinking water systems and record some of the Federally required data for migration to SDWIS/State.)

SDWIS/State and DIME provide automated compliance assistance functions which, when properly implemented and managed, reduce the level of effort and increase consistence in determining compliance with drinking water regulations.

The table below shows the current version of applications in use at Region 9 primacy agencies and the rules for which each version provides compliance assistance. It also shows the use of ancillary applications associated with SDWIS/State.

Region 9 Database Usage and Compliance Determination

January 2016

	Release Version						Compliance Determination Method												SDWIS/State Application Usage							
	Database[1]	Fed Rep	ODS Upload	Node Specification	Database Platform		TCR	Chems	Rads	Nitrate	LCR	SW TR	DBPR	GW R	P/N	CCR	Site Visits		XML Sampling	Lab to State	Migrate to State	Data Bridge	ESS	Drinking Water Watch	
American Samoa	DIME	3.4	Web	n/a	MS Access	AS												AS	Not current features in DIME[3]						AS
Arizona	S/S 3.33	3.4	Web	1.1	Oracle 10g	AZ												AZ						Not current features in DIME	AZ
California[2]	S/S 3.22	3.4	Web	1.1	SQL Server	CA												CA						Not current features in DIME	CA
CNMI	S/S 3.01	3.3	Node	2.0	SQL Server 2005	MP												MP						Not current features in DIME	MP
Guam	S/S 3.1	3.4	Node	2.0	SQL Server 2008	GU												GU						Not current features in DIME	GU
Hawaii	S/S 3.21	3.4	Node	2.0	SQL Server 2008	HI												HI							HI
Navajo Nation	DIME	3.4	Web	1.1	MS Access	NN												NN	Not current features in DIME[3]						NN
Nevada	S/S 3.31	3.4	Node	2.0	Oracle 11g	NV												NV						Not current features in DIME	NV
Region 09 Tribal	DIME	3.4	Web	n/a	Ms Access	09												09	Not current features in DIME[3]						09

Compliance Determination Key	
Automated determination using database functionality	
Partial-Manual, Partial-Automated, transition to full Automated	
Manual determination from data tracked in database	
Manual determ. from data tracked outside of database	
Violations not assigned or tracked in database	

Application Usage Key	
Use a Lot	
Use a Little	
Do Not Use	

Notes
[1] Database: DIME = Direct Implemenation Mangement Environment; S/S = SDWIS/State; PICME = Permits, Inspections, Compliance, Monitoring and Enforement system
[2] California District Offices and Los Angeles LPA only. Other LPAs are not SDWIS/State users.
[3] DIME will have an eLab feature added in the future that will use SDWIS Lab-to-State XML schemas to accept data from labs.

Double-click icon below to open file

[EMBED Excel.Sheet.12]

ED_006532_00002047-00004

Violation Reporting Limitations

New rules, notably the Revised Total Coliform Rule, require reporting of new violation types to SDWIS/Fed, which requires updating both SDWIS/State and DIME. Delaying implementation of new database releases may impede or prevent reporting to SDWIS/Fed data on violations of new rules to EPA that primacy agreements require.

In addition to SDWIS/State and DIME, agencies must also implement the most recent version of SDWIS/FedRep in order to assure complete reporting to EPA. FedRep extracts, validates, and packages data from SDWIS/State and DIME for uploading to SDWIS/Fed.

The following spreadsheet displays the current version of FedRep used by each Region 9 primacy agency and indicates the capability of FedRep to process violations for the most recent rules promulgated by EPA.

FedRep Version Reporting Capabilities for Newer Rules

(By Violation Type Codes)

FedRep	Primacy Agency									LCR	ST2 DBPR				LT2 SWTR								GWR											RTCR																		
Ver	09	AS	AZ	CA	HI	GU	MP	NN	NV	66	02	27	30	35	09	20	32	33	37	41	42	45	47	05	09	19	20	28	31	34	41	42	45	48	73	1A	2A	2B	2C	2D	3A	3B	3C	3D	4A	4B	4C	4D	4E	4F	5A	5B
3.5																																																					
3.4	X	X	X	X	X	X		X	X																																												
3.31																																																					
3.3							X																																														
3.2																																																					
3.2.2																																																					

Listed FedRep version reports these violation types

Listed FedRep version does not report these violation types

Double-click icon below to open file

[EMBED Excel.Sheet.12]

Transitioning to SDWIS/Prime

EPA is currently developing the Compliance Monitoring Data Portal (CMDP), the first functioning component of the next generation of SDWIS/State, to be called SDWIS/Primacy Agency (SDWIS/Prime). The CMDP and eventually SDWIS/Prime will reside on a remote server accessible to all primacy agencies via an internet browser (whereas SDWIS/State resides on a primacy agency's local server). Currently, the CMDP is in pilot testing with a planned release date of October 2016, with development of the remaining SDWIS/Prime functions to follow.

Because SDWIS/Prime is not fully developed and available for testing, agencies are concerned about functionality, accessibility, security, and other important aspects of the system, as the following table from 2014 shows. Nevertheless, EPA plans to stop supporting SDWIS/State one year after SDWIS/Prime is released, and in fact resource constraints as well as decreasing availability of expertise with SDWIS/State's older programming environment are already reducing EPA's ability to maintain SDWIS/State, so that will option will have limited viability.

Transitioning to SDWIS/Prime October 2014					
	Plan to Transition	If Not, Alternative	If Yes, Anticipated Dates (Month & Year)		Transition Concerns (list in priority order)
			Start Transition*	Complete Transition	
American Samoa	No	DIME			Concern about upload to SDWIS Prime transition from upload to SDWIS Fed, Fed Rep functionality. Interested in utilizing SDWIS Prime Web Portal for Laboratory electronic reporting (to DIME).
Arizona	YES		TBD	TBD	Funding to support modifications of current add-on applications (e.g., SWOCS) to functions with PRIME; rTCR functionality in PRIME.
California	Yes		TBD	TBD	CA depends heavily on Data Bridge/Migration-to-State, XML Sampling, and Lab-to-State to move LPA data into SDWIS; also depends on reporting capability and on the test environment for QA of LPA data before migrating it to Production. OGWDW has committed to providing a test instance of the database; CA wants to see that functionality before transitioning to Prime and hasn't seen any progress. CA plans to start using SDWIS for compliance determination and wants to see how Prime will handle moving sampling points. CA needs to normalize administrative contacts data in S/S before transition to Prime – currently identical contacts data are in TININDIV (Table/Inventory/Individual) records with different systems data, also in TINLGENT (Table/Inventory/Legal Entity)
CNMI	YES		May 31, 2015	September 30, 2015	Transition and maintenance costs, interfacing applications, ad hoc queries, system functionality and responsiveness, security, training, and transition timeframe.
Guam	YES		June 2015	June 2016	Ad hoc queries, Interfacing applications, system functionality & responsiveness, security, transition costs, maintenance costs, timing and availability of training.
Hawaii	YES		TBD	TBD	Interfacing applications, transition costs, system functionality & responsiveness, security.
Navajo Nation	No	DIME			Concern about upload to SDWIS Prime transition from upload to SDWIS Fed, Fed Rep functionality. Interested in utilizing SDWIS Prime Web Portal for Laboratory electronic reporting (to DIME).
Nevada	YES		TBD	TBD – not until we feel SDWIS PRIME is fully functional	Ad hoc queries (of live data), Interfacing applications, communication between add-on database with SDWIS PRIME, system functionality & responsiveness, security, transition costs, maintenance costs, timing and availability of training.
Region 09 Tribal	No	DIME			Concern about upload to SDWIS Prime transition from upload to SDWIS Fed, Fed Rep functionality. Interested in utilizing SDWIS Prime Web Portal for Laboratory electronic reporting (to DIME).

Double-click icon below to open file

[EMBED Excel.Sheet.12]

Next Steps

- Primacy agencies upgrade to SDWIS/State version 3.32 or the most current version of DIME and FedRep 3.51 for full reporting of violations for all rules including the Revised Total Coliform Rule.
- Primacy agencies incorporate the full functionality of SDWIS/State Compliance Decision Support (CDS) and DIME Compliance into their operating procedures in order to facilitate compliance determination, improve consistency among agencies, and document compliance decisions.
- Primacy agencies test and implement the Compliance Monitoring Data Portal for electronic reporting of laboratory sample analytical results to SDWIS/State for compliance monitoring as soon as practicable.
- Region 9 and primacy agencies stay informed on the status of SDWIS/Prime development and, where feasible, participate in development and testing in order to ensure that Prime will meet their programmatic requirements.

ED_006532_00002047-00006

II. Previous Program Review Recommendations

Program Review/Data Verification History

Program reviews (also referred to as *file reviews* or *data verifications*) were historically performed on-site by a team from The Cadmus Group, EPA Region 9 and, on occasion, EPA's Office of Ground Water and Drinking Water (OGWDW). Due to budget constraints, OGWDW discontinued funding for contractor support for on-site file reviews in 2013. Drinking water staff in EPA Regional Offices now conduct program reviews using regional variants of a national file review protocol. The following table displays program reviews for agencies in Region 9 since 2000.

Region 9 Primacy Agency Program Reviews Since 2000[1]									
Year	Amer. Samoa	Arizona	California	CNMI	Guam	Hawaii	Navajo Nation	Nevada	R9 Tribal
2000								8/21/00 (12/28/00)	
2001						5/7/01 (9/19/01)			
2002			2/25/02 [2] (9/19/02)						
2003									7/14/03 (7/2/04)
2004		6/28/04 (1/28/05)							
2005						2/14/05 (6/25/05)			
2006			09/25/06 (02/07/07)						
2007		2/5/07 (7/27/07)					10/22/07 (5/23/08)	8/13/07 (1/9/08)	
2008									11/17/08 (8/28/09)
2009			10/19/09 [3] (8/30/10)			7/06/09 (1/29/10)			
2010									
2011									
2012							3/5/12 (10/26/12)		
2013									
2014									
2015			9/21/2015 (4/2016)						

Notes

1. Dates in the table represent the first day of the site visit and, in parenthesis, the date of final report)
2. The 2002 California program review covered four offices: Lassen and Santa Barbara District Offices, and Shasta and Santa Barbara Counties.
3. The 2009 California review covered the following district offices: Monterey, Redding, Sacramento (HQ), Fresno, Bakersfield, San Bernardino, and Los Angeles.
4. The 2015 California review covered the Monterey district office and the Monterey County local primacy agency and focused on the groundwater and arsenic rules.

ED_006532_00002047-00007

Program Review Recommendations

The attached spreadsheet consolidates recommendations from the most recent program reviews performed in Arizona, California, Hawaii, Navajo Nation, Nevada, and EPA Region 9 tribal program.

Double-click icon below to open file

[EMBED Excel.Sheet.12]

Next Steps

- Primacy agencies and EPA PWSS Project Officers work together to prioritize and implement address recommendations that would significantly improve data quality.
- Primacy agencies and EPA PWSS Project Officers update the implementation status of recommendations in the attached spreadsheet during mid-year and end-of-year reviews and workplan negotiations.
- Region 9 conduct program reviews under a regional protocol using current Regional drinking water and enforcement staff.
- Region 9 consider updating past reviews to address implementation of "older" rules while focusing future program reviews on the "newer" rules (Stage 2 DBPR, Long Term 2 ESWTR, GWR, and rTCR) and other high priorities (arsenic, LCR).

ED_006532_00002047-00008

III. SDWIS/Fed Reporting Errors

FedRep and ODS Error Reporting Process

The SDWIS data flowchart, below, diagrams the processing of quarterly uploads of drinking water data from primacy agencies to SDWIS/Fed through the major components of the SDWIS system of databases and applications, including:

- SDWIS/State: the application most primacy agencies use to track public water system supervision activities
- SDWIS/FedRep: the application that validates data reported to EPA against established SDWIS business rules. The application generates summary and detailed data quality error reports for primacy agencies to use to correct their data prior to being processed into SDWIS/ODS.
- SDWIS Operational Data Store (SDWIS/ODS): the database component of SDWIS/Fed to which agencies upload data via EPA's Central Data Exchange
- SDWIS Federal Data Warehouse (SFDW): the repository of validated data for reporting and program management purposes

FedRep is installed at both the primacy agency and at EPA (see red arrows in diagram). To maximize data quality, primacy agencies can process data through FedRep iteratively until it is "clean" (i.e., has minimum errors) before sending it on to SDWIS/ODS. After upload to EPA, the data flows through EPA's version of FedRep which includes additional business rules to check the uploaded data and data residing in the ODS. Again, primacy agencies can use an iterative process to minimize reporting errors before the end of the reporting period (last day of calendar quarter).

[EMBED PowerPoint Slide 12]

SDWIS Data Flow

ED_006532_00002047-00009

Trends in SDWIS/ODS Error Report Counts

Primacy agencies use FedRep each quarter to extract federally required data from their database into three files for upload to SDWIS/ODS:

- Inventory (water systems, facilities, legal entities)
- Actions (violations, enforcement actions, site visits, event schedule activities)
- Samples (lead and copper 90th percentile sample results)

The extraction generates FedRep error reports on the data in the upload files, and the process of merging those files into SDWIS/ODS generates another set of ODS error reports on the data in the ODS after the merge. These errors are of two types:

- ***Rejection (ERR) errors*** – records rejected due to missing or invalid data that violates SDWIS business rules
- ***Data Quality (DQ) errors*** – records accepted although the data does not fully meet EPA reporting guidance

The data in the summary spreadsheets below show ODS errors for each agency for the last four years. Some error types for different agencies show decreasing trends, but many are volatile or fairly stable at high numbers. These trends suggest priorities in addition to the record rejection vs. completeness criteria.

Double-click icon below to open file, then click on a Primacy Agency tab

ODS **Inventory** Error Counts

[EMBED Excel.Sheet.12]

ODS **Actions** Error Counts

[EMBED Excel.Sheet.12]

ODS **Samples** Error Counts

[EMBED Excel.Sheet.12]

In addition to the ODS reports, FedRep error reports generated at primacy agencies on their SDWIS/State and DIME databases may include additional errors that do not appear in these SDWIS/ODS error reports. FedRep also generates log files of accepted and rejected records in each upload file that EPA does not receive.

Primacy agencies can request ODS extracts through EPA's Central Data Exchange (CDX) interface in order to analyze the data in the ODS after each upload. CDX also provides an interface to the SDWIS Federal Data Warehouse, the repository from which other applications draw SDWIS data, which allows primacy agencies to compare data in the source, ODS, and SFDW databases. Along with available documentation such as the data entry instructions (DEIs) for each rule, these resources can help primacy agencies determine how to correct data quality issues.

Next Steps

- Primacy agencies review data in the originating system, in ODS extracts, and in the SFDW to analyze and address data quality issues to reduce the number of FedRep and ODS errors from quarterly migration of data to EPA.
- Agencies prioritize rejection errors, focusing on violation and violation-enforcement association rejection errors, which affect the accuracy of enforcement action prioritization in the Enforcement Tracking Tool that imports its data from SDWIS/Fed.

ED_006532_00002047-00010

- After correcting data rejection errors, primacy agencies analyze and correct data quality errors, prioritizing active historical water systems and facilities and other errors that could complicate transition to SDWIS/Prime.
- Primacy agencies review the webinar training on the SDWIS data flow process and FedRep and ODS error reports, available on the ASDWA web site, and submit inquiries to Region 9 and the SDWIS Support helpline, as necessary to support data quality improvement efforts.

IV. Priority Data Management Issues

Some data quality issues have been proven persistent over time or deserve a higher priority of improvement efforts. The SDWIS/Fed Reporting System (SFRS) is available to all primacy agencies through CDX, and a version that omits sensitive data is available to the public as well. The SFRS provides a variety of report types, including inventory and violation reports, National Program Managers (NPM) measures, sanitary survey completeness, Lead and Copper Rule, among others. One notable report available from the SFRS is the Data Quality report, which provides measures of specific data quality characteristics. This section provides the current status of several of these issues.

Inventory and Location Measures

Locational data is increasingly important, notably for emergency planning and response and for drinking water source protection. The DWMAPs web site, for instance, uses SDWIS/Fed data to show approximate locations of drinking water sources. For similar reasons, an accurate record of which facilities are currently active and which are no longer active is important. The Inventory Issues workbook, attached below, provides Region 9 agencies' status on measures of facility and locational data quality. You can filter each sheet on the "State" column to display the data for a specific agency. The accompanying description document provides the applicable minimum reporting requirements.

Double-click icon below to open file

[EMBED Excel.Sheet.12][EMBED Word.Document.12 \s]

Sanitary Survey Completion

Sanitary surveys provide data for updating SDWIS inventory, in addition to identifying deficiencies in water system facilities and operations. Region 9 distributes quarterly summaries of the percent of required sanitary surveys completed to all primacy agencies, and the SDWIS Federal Reporting System also provides sanitary survey completion reports, both from SDWIS/Fed data. Completion rates from the July SFDW become the annual National Program Managers Measure SDW-01a for community water systems and SDW-01b for tribal community water systems.

SDWIS reporting guidance requires primacy agencies to report the outcomes of sanitary surveys as category evaluation codes for water system facilities -- source, pumps, treatment, storage, and distribution system -- as well as water system operational characteristics -- management and operations, operator compliance, and monitoring and reporting. While the completion measures do not account for category evaluations, the DQ00024 through DQ00032 errors on the ODS Actions error report in Section III show where category evaluations are missing.

Double-click icon below to open file

[EMBED Excel.Sheet.12]

Open-Ended Violations

EPA reporting requirements dictate that some violations must be reported *without* a Compliance Period End Date. These are called "open ended violations." In the current SDWIS/Fed Data Warehouse, Compliance Period End Dates of open-ended violations retain a null value when migrated from the ODS. In the Warehouse, a Compliance Status Code of "O" is generated if the violation has not been returned to compliance.

While open-ended violations for recent non-compliance events are expected, SDWIS/ODS contains a number of older violations that most likely have returned to compliance, but that still indicate non-compliance due to failure to associate returned to compliance enforcement actions with those violations. The attached table includes all violations with a Compliance Status Code of "O" currently in the SFDW for Region 9 primacy agencies, filtered to show only those violations whose Compliance Period Begin Date is more than five years ago.

Double-click icon below to open file

[EMBED Excel.Sheet.12]

Enforcement Targeting Tool ("ETT list") Data Quality Annotations

The Enforcement Targeting Tool ("ETT list") is the means for EPA and primacy agencies to implement EPA's Enforcement Response Policy for public water system supervision. Region 9 provides a version of the ETT list that facilitates communication between the Region and primacy agencies on the status of systems and the priority of enforcement response to violations of drinking water regulations. This "annotated ETT list" provides columns for indicating when data issues might affect enforcement prioritization.

Filtering the annotated ETT list on columns V and W can provide a list of PWSs noted as having data quality issues not due to data migration timing, which primacy agencies can then target for thorough data quality analysis. Most agencies have few or none such records while one has many; in either case it could be beneficial for each agency to review their ETT annotation SOP to ensure accurate recording of ETT data quality issues. Filtering on the "time lag" column W alone can provide a list of records, as in the NN example below, to compare with the next quarter's ETT list to ensure it reflects the expected results.

ED_006532_00002047-00012

Double-click icon below to open file

Agency	DQ issues noted	DQ-filtered ETT
AS EPA	1	[EMBED Excel.SheetMacroEnabled.12]
AZ DEQ	many	[EMBED Excel.SheetMacroEnabled.12]
CA WRCB	3	[EMBED Excel.SheetMacroEnabled.12]
CNMI BECQ	4	[EMBED Excel.SheetMacroEnabled.12]
GU EPA	0	
HI DOH	0	
NN EPA	0	[EMBED Excel.SheetMacroEnabled.12]
NV DEP	7	[EMBED Excel.SheetMacroEnabled.12]
R9 DI	3	[EMBED Excel.SheetMacroEnabled.12]

90th Percentile Lead Sampling Completion

Lead and Copper Rule reporting guidance requires primacy agencies to report 90th percentile lead levels for all medium and large water systems (population served greater than 3,300) and exceedances of the 90th percentile lead action level for small systems and all 90th percentile copper action level exceedances for all system. Agencies must conduct lead and copper tap sampling once every three years. The table attached below shows the completeness of reporting 90th percentile lead levels for medium and large community and non-transient PWSs for the 2013-2015 period.

Double-click icon below to display summary table

[EMBED Excel.Sheet.12]

Overall Data Quality Measure

Many of the measures included in this Region 9 *SDWIS Data Quality Status* report are available from the SDWIS Federal Reporting Services (SFRS) interface, including reports specifically for sanitary survey and lead sampling completion. The Data Quality report from SFRS provides some of the same measures -- e.g., of locational data quality and software version in use – that this document provides as well as some noteworthy measures not included here -- e.g., on timeliness and accuracy of violation reporting. The SFRS Data Quality report also provides an overall measure of data quality that is the basis of EPA's annual SDWIS data quality awards, among other uses. EPA is currently considering revising the set of measures that contribute to the composite score and seeking input on measures to include.

Double-click icon below to open file

Agency	Overall score	SFRS Data Quality Report
AS EPA	.971	[EMBED AcroExch.Document.DC]
AZ DEQ	.9834	[EMBED AcroExch.Document.DC]
CA WRCB	.9485	[EMBED AcroExch.Document.DC]
CNMI BECQ	.9484	[EMBED AcroExch.Document.DC]
GU EPA	NA	[EMBED AcroExch.Document.DC]
HI DOH	.9943	[EMBED AcroExch.Document.DC]
NN EPA	.9953	[EMBED AcroExch.Document.DC]
NV DEP	.9613	[EMBED AcroExch.Document.DC]
R9 DI	.9866	[EMBED AcroExch.Document.DC]

Next Steps

Primacy agencies should review the data in these spreadsheets and in the SDWIS Federal Reporting System Sanitary Survey, LCR, and Data Quality reports and take appropriate action as follows:

- Inventory and Location Details: input missing and incomplete inventory/locational data into SDWIS/State and DIME in order to report all required data
- Sanitary Survey completion: ensure that sanitary survey are reported as they occur, with required category evaluation codes, and record any missing category evaluation codes for previous sanitary surveys.
- Open-Ended Violations: review open-ended violations and, as appropriate, associate those violations with return-to-compliance enforcement codes (SOX/EOX).
- ETT list "Data Quality Issue" annotations: ensure that data quality issues are accurately annotated, including violation and contamination types. Review FedRep ODS data quality reports to correlate rejected violations, violation-enforcement action associations, and enforcement actions with ETT-listed PWSs. Review Data Entry Instructions for relevant rules as necessary to identify required data. Query SDWIS/State, ODS extract, and SFDW databases to identify missing data or confirm that all required data is present. Report uncorrected data quality issues documented through this process to the SDWIS Support helpline and EPA Region 9.
- 90th Percentile Lead Sample Results: review water systems that have not had 90th percentile lead levels reported to SDWIS/Fed to determine if any missing data results from non-compliance with sampling requirements, data entry issues, or problems with reporting the data to EPA, or if sampling and reporting is scheduled by the end of the sampling period.

ED_006532_00002047-00014