

2015 DWINSA

B.1 SURVEY OBJECTIVES, KEY VARIABLES AND OTHER PRELIMINARIES

2015 Drinking Water Infrastructure Needs Survey and Assessment (DWINSA), EPA ICR Number 2234.04, OMB Control Number 2040-02747.

B.1.a Survey Objectives

The primary objective of the 2015 DWINSA is to collect information from water systems on the infrastructure they need to continue to provide safe drinking water to consumers. These data are used to produce a national estimate as well as state-specific estimates of water systems' 20-year need. EPA has established policies to ensure that the overarching goals of the survey are met:

- Estimate the total national 20-year need
- Estimate the total 20-year need for each fully participating state
- Provide complete and accurate data to Congress
- Provide a tool to fairly distribute DWSRF capitalization funds to states
- Maintain the credibility of the DWINSA findings

EPA proposes to collect information on the cost of water systems' infrastructure needs. If cost data are not available from systems, EPA proposes to collect information that will enable the Agency to model costs. In the data collection instrument, the respondent will identify needs on a project-by-project basis and list the "type(s) of need" that the project will meet. The "types of need" include raw water source, transmission, source water treatment, storage, distribution, pumping stations and other needs.

EPA will use the information from the DWINSA to estimate capital investment requirements of drinking water systems. The information will be used to allot DWSRF monies among states and as part of an allotment formula for the DWSRF Tribal Set-Aside (TSA) Program.

For the 2015 DWINSA, EPA is proposing to use a modified panel approach to select survey respondents. The modified panel approach will involve dropping a random selection of 25 percent of the systems serving 3,301 to 100,000 people that participated in the 2011 DWINSA and then drawing a random sample to replace those systems in the survey for the 2015 DWINSA. This will be done for each state and by strata. By primarily using information from the 2011 DWINSA, this approach would reduce the amount of time required to prepare and review the responses from systems resurveyed in 2015.

For the new systems selected for the 2015 Assessment, EPA will use the same methodology as used in previous DWINSA. The sampling design is discussed in detail below.

B.1.b Key Variables

Several key variables are available from the Safe Drinking Water Information System (SDWIS). To ensure accuracy, the 2015 DWINSA will verify these data by asking respondents to confirm existing information (pre-populated on the data collection instrument) or correct it. These variables include population served, total design capacity, number of service connections, primary source of supply, ownership type (private or public) and whether the system purchases water from or sells water to another water system.

Information on capital needs will be collected from respondents on a project-by-project basis. For each project, respondents will be asked to provide the following types of information: type of need; reason for

need; documentation of need and cost (if necessary); if the project is a new project or to replace, rehabilitate or expand existing infrastructure; if the project is needed now to protect public health or if it is needed over the next 20 years to continue to provide safe drinking water; the federal regulation or state requirement (if the project is to meet a current regulation or state requirement); design capacity of source, storage and treatment projects; length and diameter of pipe projects; diameter for projects such as water meters; cost of the project (if available); and date of the cost estimate. For most of these variables, respondents will choose the appropriate “documentation,” “type of need,” “reason for need” or “regulation or requirement” from EPA’s Lists of Codes.

The principal variable of interest is total projected capital needed for each water system in the 2015 DWINSAs for the time period of January 1, 2015, through December 31, 2034. The total capital need for all systems in each state (to be derived from the statistical sample of systems) is the key variable that decision-makers at EPA use to allocate funds to states based on need.

The method of data collection has been designed to minimize burden on respondents while ensuring that information is collected in a consistent manner. Collecting information on a project-by-project basis, for example, will be particularly helpful in reducing burden since most respondents develop Capital Improvement Plans (CIPs) on a project-by-project basis.

Information on type of need will be used to disaggregate total capital needs for EPA’s Report to Congress. Information on the reason for need will be used to verify the public health benefit of the need. Information on the date of the cost estimate will be used to provide a consistent basis for cost estimates across systems. Information on a regulation or requirement will be used to determine the reported project costs related to federal regulations or state requirements.

If a system cannot provide cost estimates, additional data are necessary so that the Agency can impute costs. Each of these variables is described in greater detail later in this document.

B.1.c Statistical Approach

The 2015 DWINSAs are being designed to achieve a desired level of precision for state-level estimates of total capital needs for systems serving more than 3,300 persons. EPA proposes a modified panel approach that includes a census of large systems and a survey of a statistical sample of medium-sized systems to estimate total capital needs. This statistical approach minimizes burden while achieving the desired level of precision.

The 2015 DWINSAs design divides CWSs serving populations of more than 3,300 into two groups: CWSs serving populations of more than 100,000 and systems serving populations of 3,301 to 100,000. EPA proposes to sample with certainty systems serving more than 100,000 persons. These systems have the largest capital needs and they have the staff to respond efficiently to the 2015 DWINSAs. EPA proposes to use a random sample of systems serving 3,301 to 100,000 persons. This methodology can reduce burden and still achieve the DWINSAs data quality objectives.

To further reduce burden, EPA proposes using a modified panel approach for the 2015 DWINSAs. Rather than select a completely new sample of systems in 2015, EPA will reassess the needs of most of the systems that participated in the 2011 Assessment. EPA will replace 25 percent of the sample of systems serving 3,301 to 100,000 people. By state and stratum, EPA will randomly select 25 percent of the sample to drop and will then randomly select replacement systems from the sampling frame. By primarily using information from the 2011 DWINSAs, this approach would reduce the amount of time required for systems to prepare and states to review the responses from systems resurveyed in 2015. This approach

will maintain EPA's sampling targets for each stratum and ensure that EPA continues to meet its precision targets for each state. Additionally, by using information from the 2011 review, the modified panel approach would reduce the amount of time required for EPA's contractor to review each system's response. By replacing 25 percent of the sample, EPA will reduce a potential source of bias introduced by the panel. (When a completely new sample was selected for each assessment, the sampling error was a random component that changed from survey to survey. With the panel approach, this error becomes systematic.) By refreshing 25 percent of the sample, the approach alleviates this potential source of bias and helps to ensure that the 2015 sample represents the need as it exists in 2015.

To meet the state-level precision targets, EPA will use the same strata as in the 2011 DWINSA. As previously mentioned, EPA will adjust the sample size to accommodate changes in the sample frame. These changes may address new systems, systems that are no longer active and systems that have "migrated" between strata (become smaller or larger, or changed source). If EPA determines that there have been substantial changes in the size of the sample frame since the 2011 DWINSA, EPA will adjust the sample size as needed to ensure that the precision targets are met for each state. As in the 2011 DWINSA, EPA will first determine the total sample size for each state to meet the target level of precision. EPA will then allocate the sample to strata in order to maximize the efficiency of the design.

EPA is designing and conducting the 2015 DWINSA with the assistance of a contractor:

Contractor

The Cadmus Group, Inc.
100 5th Avenue, Suite 100
Waltham, MA 02451
(617) 673-7000

Contractor Roles

- Oversight of data collection instrument design
- Oversight of statistical sample design
- Training
- Data collection instrument package distribution: logistics
- Technical support for respondents and states
- Review of all survey responses for conformance to survey policies and project allowability criteria
- Model development
- Data processing
- Statistical sample design

B.1.d Feasibility

The 2015 DWINSA data collection instrument has been designed with the capabilities of the typical respondent in mind. To fully assess feasibility, the Agency undertook the following steps. EPA convened a workgroup (see Section A.5.b) to comment on the proposed data collection and its feasibility. The data collection instrument to be used for the 2015 DWINSA is generally the same form as used for the past two DWINSAs. For the 2007 DWINSA, EPA conducted a pre-test in which EPA's contractor met with individual CWS operators and discussed the proposed survey. System operators were asked to comment on all proposed data elements and the feasibility of collecting information by a mail survey. The Agency recognizes that most systems serving fewer than 50,000 persons and some that serve 50,000 or more may not have cost data or documentation of costs for some projects. In those cases, the 2015 DWINSA data collection instrument requests other readily available information that EPA can use to model costs. EPA will emphasize to respondents that they are not expected to develop cost estimates for the purposes of the 2015 DWINSA. In addition, EPA (or states) will provide systems with technical assistance for completing the data collection instrument.

EPA has developed cost models for most of the infrastructure needs included in the 2015 DWINSA based on the size and capacity of a project. These cost models were originally developed during the 1995 DWINSA, have been updated during subsequent assessments, including the 2011 DWINSA, and will be used again for the 2015 DWINSA. New cost models may be developed for weaker cost models, influential cost models and new technology.

The time frame for the 2015 DWINSA is acceptable to the users of data within the Office of Ground Water and Drinking Water (OGWDW) and sufficient to complete a report to Congress by its anticipated due date of early 2017. The schedule also is acceptable to other users of the data.

B.2 SURVEY DESIGN

This section contains a detailed description of the statistical survey design and modified panel approach including a description of the sampling frame, sample identification, precision requirements and data collection instrument.

The sample design for the 2015 DWINSA is stratified random sampling within each state. In cases where the state is not participating in the data collection for systems serving 3,301 to 100,000 persons, EPA will not provide state-specific results, as the data collection for these states does not meet the DWINSA data quality objectives. EPA will include an overall national result for the systems serving 3,301 to 100,000 persons, using the average need by strata of the systems in states that are participating in the full 2015 DWINSA. For states that are fully participating in data collection for systems serving 3,301-100,000 persons, the 2015 DWINSA will use a modified panel approach for sampling these systems within each state. This approach is described in more detail in Section B.2.b.

Stratification increases the precision of estimates compared with a simple random sample of the target population of systems. In stratified samples, the target population is divided into non-overlapping groups, known as strata, from which separate samples are drawn. The goal of stratified sampling is to choose sample sizes within each stratum in a manner designed to obtain maximum precision in the overall estimate for the population. Stratification variables for this study include: population size (populations of: 3,301 to 10,000; 10,001 to 25,000; 25,001 to 50,000; 50,001 to 100,000 and more than 100,000) and primary sources of supply (surface and ground). Systems serving more than 100,000 persons are selected with certainty. For the 2011 DWINSA, the size of each state's sample of systems serving populations of 3,301 to 100,000 was set to meet the 2011 DWINSA's data quality objectives. For the 2015 DWINSA, the survey will rely on a modified panel approach in which 75 percent of the 2011 DWINSA respondents serving populations of 3,301 to 100,000 are resampled, and 25 percent are put back into the frame and a new 25 percent are drawn.

EPA's precision target for the 2015 DWINSA is to be 95 percent confident that the true need for each state lies within an interval of plus or minus 10 percent of the estimated need. These precision targets are identical to the targets for the 2011 Assessment. The 2011 sample, modified as described above, will meet the assessment's precision target. The sample sizes will be adjusted to account for changes in the inventory of systems, if necessary, to ensure the 2015 sample meets the precision targets.

B.2.a Target Population and Coverage

The target population for the 2015 DWINSA is the number of CWSs in the nation. A CWS is a public water system (PWS) that serves at least 15 service connections used by year-round residents or regularly serves at least 25 year-round residents (40 CFR 141.2). The 2015 DWINSA is designed to produce estimates of the capital need of systems serving more than 3,300 persons for each participating state. In

partially surveyed states, EPA will be able to provide state-specific results for systems serving 100,000 or more persons. EPA will include an overall national result for the systems serving 3,301 to 100,000 persons using the average need by strata of the systems in participating states and the total number of systems by strata in the partially surveyed state. The 2015 DWINSA is designed to produce estimates of the capital need of systems serving 3,300 and fewer persons by state and for the nation as a whole using the results of the 2007 DWINSA.

B.2.b Sample Design

This section describes the sample design. It includes a description of the sampling frame, target sample size, stratification variables and sampling method. The sampling design employed is a stratified random sample of CWSs. The strata employed in the design are discussed in Section B.2.b.iii. Neyman allocation is used to efficiently allocate the sample of water systems among the strata.

B.2.b.i Sampling Frame

The sampling frame is developed from SDWIS. SDWIS is a centralized database for information on PWSs, including their compliance with monitoring requirements, maximum contaminant levels (MCLs) and other requirements of the SDWA. The following information will be extracted from SDWIS for the statistical survey and verified by participating states:

- Name of system.
- Contact person.
- Address of system.
- Population served.
- Total design capacity.
- Number of connections.
- Primary source (surface water or ground water).
- PWS identification number (PWSID).
- Ownership type.
- Whether the water system purchases/sells water from/to another water system.

From these data, EPA will develop the frame from which EPA will calculate summary statistics (e.g., number of systems per state in pre-defined strata) for use in calculating sample size. For the modified panel approach, the 2015 sampling design will use the 2011 DWINSA sample and make targeted modifications to account for changes in the inventory of systems between 2011 and 2015. Systems that have closed since the 2011 DWINSA will be removed from the 2015 survey. The needs of merged systems will remain in the survey as the need of the combined system. New systems added to the inventory (i.e., systems that served 3,300 or fewer people in 2011 that now serve more than 3,300 people, or newly created systems) will be sampled to ensure the sample is representative of systems in 2015. New large systems (those serving more than 100,000 persons) will also be added to the census.

Justification for the Use of SDWIS

The following criteria are often used in assessing a proposed sampling frame:

- It fully covers the target population.
- It contains no duplication.
- It contains no foreign elements (i.e., elements that are not members of the population).
- It contains information for identifying and contacting the units selected in the sample.
- It contains other information that will improve the efficiency of the sample design.

The units of observation for this survey are CWSs, a subset of PWSs. SDWIS is the ideal choice for a sample frame because of its inclusive coverage of all units of observation for the 2015 DWINSA. In addition, SDWIS has two other advantages: it contains information that will facilitate contacting the respondents and it contains other information that is useful in stratifying the sample, thereby improving the efficiency of the sample design.

In previous surveys where SDWIS was used as a sample frame, there have been criticisms of its utility. Since 1989, EPA has conducted audits of the quality of SDWIS data. As a result, EPA is aware of the problems with SDWIS. The audits, however, show that errors in classification of systems by strata proposed for the 2015 DWINSA are rare. The audits show that systems are misclassified by population or source in less than one percent of all cases.

To mitigate any potential problems with the sample frame, the 2015 DWINSA design anticipates substantial state involvement in the 2015 DWINSA process. For example, states will be checking the sample frame of systems that will be used to determine the final sample. In EPA's experience, states often have in-house data systems with very accurate data. Even if these data are not transmitted to SDWIS, they are available and can be used by states to check the sample frame.

B.2.b.ii Sample Size

Exhibit B-2-1 at the end of this subsection shows the preliminary sample sizes for the 2015 DWINSA. For the modified panel approach, sample sizes for the 2015 DWINSA will be the same as for the 2011 DWINSA except for changes to accommodate: (1) the addition of large systems serving more than 100,000 people since the 2011 survey; (2) any partially surveyed states from the 2011 survey that fully participate in the 2015 survey; and (3) changes in the inventory of medium systems, including systems changing size or source categories or the creation of new systems. As shown on Exhibit B-2-1, the sampling design will be implemented to achieve state-level precision targets for CWSs serving more than 3,300 persons. Precision targets are discussed in Section B.2.c.

The task of determining the sample size for each stratum requires two steps. The first step determines the sample size for each state that achieves the precision targets for that state. The second step allocates the sample among the relevant strata in the state. The strata are described in section B.2.b.iii.

The first step calculates the total number of systems required at the state level to meet the precision requirements. The sample size is given by:

$$n_{0g} = \frac{\left(\sum_{h=1}^H N_{gh} S_{gh} \right)^2}{V_g}$$

Where: n_{0g} = the sample size for state g (prior to the finite population correction)

N_{gh} = the total number of systems in the g^{th} state in the h^{th} stratum (taken from SDWIS)

S_{gh} = the standard deviation of the variable of interest in the g^{th} state for the h^{th} stratum (estimated using data from the data from previous assessments)

H = the number of strata defined in the sample design for the g^{th} state

V_g = the desired sampling variance for the total system (those serving more than 3,300 persons) capital needs estimate for state g.

The desired error in the sample is expressed as a relative error. In the above equation, $V_g = (d/Z_\alpha * \hat{Y}_g)^2$. \hat{Y}_g is an estimate of the total capital needs for a given state. \hat{Y}_g is computed for each state by calculating the mean total capital needs for stratum h (from the prior DWINSAs) and multiplying this mean by the actual number of systems in each stratum for that state (N_{gh}). Summing across strata provides an estimate of \hat{Y}_g . d is the half-width of the desired confidence interval (0.10 for the Assessment). Z_α is the value of a standard normal distribution for a confidence level of $1 - \alpha$, (1.96 for the Assessment).

Because the number of water systems is known and finite, the following population correction is applied:

$$n_g = \frac{n_{0g}}{1 + \frac{1}{V_g} \sum_{h=1}^H N_{gh} s_{gh}^2}$$

The second step allocates the total sample to each sampling stratum. EPA will randomly draw this number of samples from each of these strata. The Neyman allocation is used to determine the sample size for each stratum:¹

$$n_{gh} = n_g \left(\frac{N_{gh} s_{gh}}{\sum_{h=1}^H N_{gh} s_{gh}} \right)$$

(Because systems serving populations more than 100,000 are to be sampled with certainty, H is the number of strata of systems serving 100,000 or fewer persons.)

In order to implement these sample size and sample allocation equations, EPA needs estimates for V_g , N_{gh} , s_{gh} and mean total capital needs by stratum. Information on mean total capital needs by stratum and s_{gh} were estimated using data from the prior DWINSAs.

Exhibit B-2-1 State Sample Sizes

State	Estimated Total Number Of Systems Serving More Than 3,300 Persons	Estimated Sample Size For Systems Serving More Than 3,300 Persons
Alaska *	17	1
Alabama	348	131
Arkansas	181	83
American Samoa	1	1

¹ J. Neyman, "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, Vol. 97 (1934), pp. 558-606; as cited in William G. Cochran, *Sampling Techniques* (New York: John Wiley & Sons), 1977.

State	Estimated Total Number Of Systems Serving More Than 3,300 Persons	Estimated Sample Size For Systems Serving More Than 3,300 Persons
Arizona	130	39
California	688	170
Colorado	169	61
Connecticut	57	36
District of Columbia	1	1
Delaware *	30	3
Florida	387	128
Georgia	237	66
Guam	3	3
Hawaii *	30	2
Iowa	138	53
Idaho *	51	1
Illinois	460	98
Indiana	214	82
Kansas	117	65
Kentucky	259	141
Louisiana	231	65
Massachusetts	253	73
Maryland	59	26
Maine	35	25
Michigan	303	63
Minnesota	181	91
Missouri	216	121
Northern Mariana Islands	2	2
Mississippi	202	105
Montana *	35	1
North Carolina	280	80
North Dakota *	32	0
Nebraska *	43	2
New Hampshire *	39	1
New Jersey	242	61
New Mexico *	60	1
Nevada	35	15

State	Estimated Total Number Of Systems Serving More Than 3,300 Persons	Estimated Sample Size For Systems Serving More Than 3,300 Persons
New York	362	50
Ohio	320	90
Oklahoma	165	86
Oregon	116	54
Pennsylvania	349	83
Puerto Rico	118	58
Rhode Island *	28	3
South Carolina *	159	9
South Dakota *	45	2
Tennessee	288	158
Texas	987	144
Utah	109	50
Virginia	160	56
Virgin Islands	2	2
Vermont *	34	0
Washington	213	58
Wisconsin	181	58
West Virginia *	110	1
Wyoming *	27	0
Total	9,539	2,859

*Fifteen states are expected to opt out of participating in the statistical portion of the survey (i.e., collecting data from systems serving 3,301 to 100,000 persons). However, those states that have systems that serve more than 100,000 people will participate in the census portion of the survey (i.e., collecting data from systems serving more than 100,000 persons). For those 15 states, the number in the “Estimated Sample Size for Systems Serving More Than 3,300 Persons” represents the total number of systems in the state that serve more than 100,000 persons.

B.2.b.iii Stratification Variables

The objective of stratification is to increase the efficiency of the sampling design (thereby reducing the number of systems to be sampled for a given level of precision). Stratified sampling may produce a gain in precision in the estimates of the characteristics of the target population as compared to simple random sampling. In stratified sampling, the target population (i.e., CWSs) is divided into non-overlapping strata that are internally homogeneous, in that the measurements vary little from one unit to another (i.e., the within-stratum variance is minimized). If the within-stratum variance is relatively small, then a precise estimate of the variable of interest can be obtained with relatively small samples. Each of the strata estimates can be combined to obtain a precise estimate for the overall target population.

EPA’s drinking water programs have historically evaluated CWSs based on (1) the number of persons served and (2) the primary water source (ground water and surface water).² Using total capital need information obtained from prior DWINSAs, EPA evaluated several classification schemes. This analysis showed that the stratification scheme used in prior assessments (10 strata based on size and source) would be appropriate for the 2015 DWINSA. For some states, EPA may combine the 10,001 to 25,000 and 25,001 to 50,000 size categories within each source category, resulting in 8 rather than 10 strata. EPA will combine these two size categories only if the sample using 8 strata is more efficient than the sample using 10 strata. The proposed strata for systems serving more than 3,300 persons are as follows:

Size of Population Served	Source	Sample Methodologies
3,301 – 10,000	Ground	Panel approach with 25 percent refresh using a random sample.
3,301 – 10,000	Surface	
10,001 – 25,000	Ground	Panel approach with 25 percent refresh using a random sample. In some states the number of strata will be reduced based on analysis of optimal stratum boundaries. Specifically, in some states systems serving between 10,001 and 50,000 will be in one size group rather than two.
10,001 – 25,000	Surface	
25,001 – 50,000	Ground	
25,001 – 50,000	Surface	
50,001 – 100,000	Ground	Panel approach with 25 percent refresh using a random sample
50,001 – 100,000	Surface	
More than 100,000	Ground	Sampled with certainty
More than 100,000	Surface	

B.2.b.iv Sampling Method

As indicated above, all CWSs serving populations of more than 100,000 will be sampled with certainty.

For systems serving 3,301 to 100,000 persons, all CWSs will be allocated to eight strata, based on the population served and primary water source. The sample size for each stratum in each state will be determined by the sampling strategy outlined above. As previously described, the modified panel approach that will be used for the 2015 sample will begin with the 2011 DWINSA sample and make targeted modifications to account for changes in the sample between 2011 and 2015. EPA will then “refresh” the sample by randomly replacing 25 percent of the 2011 sample of systems serving 3,301 to 100,000 people with systems that were not included in the 2011 Assessment. To refresh the sample, 25 percent of systems serving 3,301 to 100,000 persons will be dropped from the survey and returned to the pool of systems that were not selected in 2011. The “refresh” will then randomly select systems from among the pool of systems that were not in the 2011 sample, the systems that were randomly dropped and new systems so that the number of systems in the sample reaches the same size as the 2011 sample.

The sampling method for the 25 percent refresh will be similar to the approach used in the 2011 survey. An equal probability random sample will be drawn from each stratum. Anticipating a level of non-response, EPA will over-sample the refresh systems to achieve the desired number of completed data collection instruments. Since the expected response rate for systems serving 3,301 to 100,000 persons is approximately 90 percent, EPA has increased the sample by approximately 10 percent. However, as

² For the purposes of the 2015 DWINSA, purchased surface water systems are included with ground water systems. This design yields lower within-stratum variance and has been used since the 1999 DWINSA.

discussed below, the DWINSA has consistently achieved a higher response rate than estimated. Therefore EPA has included the full sample size estimate in the burden estimate of this ICR.

B.2.c Precision Requirements

B.2.c.i Precision Targets

The sampling design for the 2015 DWINSA will be implemented at the state level. EPA's goal is to be 95 percent confident that the margin of error, when estimating the total capital needs facing these systems in each state, will be plus or minus 10 percent of the total need for these systems. For example, if the total need for these systems in a state is estimated to be \$2 billion, EPA will be 95 percent confident that the actual total need is between \$1.8 billion and \$2.2 billion.

B.2.c.ii Nonsampling Error

EPA has developed an assessment approach that will employ several quality assurance techniques to maximize response rates, response accuracy and processing accuracy to minimize non-sampling error.

Particular emphasis will be placed on maximizing response rates. Standard methods that have proved effective in other surveys involving states and water systems will be used, including the following:

- EPA and the states will coordinate in the production of a cover letter for the 2015 DWINSA. EPA's opinion (shared by state drinking water administrators and trade associations) is that surveys on state letterhead will be better received than surveys on EPA letterhead. Therefore, states can use state-level cover letters signed by a senior state official instead of the EPA letter.
- The states will place a telephone call to each participating system to ensure that they understand the survey process and their role.
- The data collection instrument design, content and format were reviewed by states that participated in the 1995, 1999, 2003, 2007 and 2011 DWINSAs.
- Questions being asked are those that owners or operators of systems should know. EPA does not ask questions that require monitoring, research or calculations on the part of the respondent.
- The data collection instrument design is limited to a cover page of system information and characteristics and one project table, with three optional tables to record general information about the system's infrastructure inventory. By limiting the information requested, EPA believes that the average water system respondent can complete the data collection instrument in approximately 5.53 hours. Exhibit A-6-6 shows the breakdown of the total burden hours for CWSs by system size in the 2015 DWINSA.
- Respondents will be encouraged to call state personnel who will be trained to answer questions. In addition, EPA will provide technical assistance to states and water systems.
- The electronic format of the survey will make returning the data collection instrument convenient.

Standard methods to reduce other sources of non-sampling error also will be used:

- EPA expects complete coverage of the target population using SDWIS, supplemented by state review of all systems.

- Data will be 100 percent independently keyed and verified.
- The data collection instrument is pre-coded to improve accuracy by eliminating unnecessary processing steps.

Supplementing these standard methods, EPA proposes several unique steps to eliminate non-sampling error which have been developed in concert with organizations representing the states and water systems. These organizations believe that the 2015 DWINSA is important and that a high level of participation by all water systems is essential to its success. Because of the substantial commitment being made by states and water systems to the 2015 DWINSA, EPA believes that response rates will be higher than most surveys of similar respondents. To ensure success, states and organizations representing water systems are taking the following steps.

- **Participation of the states.** Because the 2015 DWINSA will be used to allocate DWSRF funds to states, each entity has a strong interest in achieving a high response rate. EPA believes that their participation will be a key factor in guaranteeing high response rates and low item non-response. Personnel who work with water systems every day are in a strong position to encourage systems to complete the 2015 DWINSA form. States have committed to assisting EPA in achieving a high response rate by participating in follow-up activities. EPA will provide technical assistance to any system that has questions about the 2015 DWINSA.
- **Participation of Organizations Representing Water Systems.** EPA anticipates public support of organizations representing water systems. The prior assessments were supported by groups such as the American Water Works Association (AWWA), the National Association of Water Companies (NAWC) and the Association of Metropolitan Water Agencies (AMWA).

This support by the organizations representing the respondents for the 2015 DWINSA can be helpful in many ways to minimize non-sampling errors. For example,

- In past DWINSAs, national water associations sent letters to each system in their membership, stressing the importance of surveying drinking water infrastructure needs. These letters, along with the letter from the states, helped convince water systems to respond. EPA will seek similar support from these associations for the 2015 DWINSA effort to encourage systems to complete the data collection instrument.
- In the past DWINSAs, the largest association representing water systems serving populations greater than 3,300, AWWA, provided support through its national organization. To improve the response rate, AWWA enlisted the support of its state affiliates to conduct telephone follow-up calls to encourage response. AWWA assisted in past DWINSAs to help achieve high response rates. EPA will seek similar support from AWWA in support of the 2015 DWINSA.
- **Communications Strategy.** EPA has developed a comprehensive communications strategy that will inform likely respondents of the need for their participation. This strategy includes articles in magazines, newsletters and bulletins of all major organizations that represent (or communicate with) water systems. This includes publications of all of the organizations mentioned above, plus the state and local affiliates of these organizations. The strategy is designed to develop widespread peer-group support for participation in the 2015 DWINSA.

B.2.d Data Collection Instrument Design

Questions about system characteristics (name, population served, number of connections and other customary business information) will be pre-populated on all data collection instruments. The respondent needs only to enter accurate information if any pre-populated information is not correct.

The 2015 DWINSA is based on a matrix project table that requests a list of capital water system infrastructure projects that the system plans for the period 2015 through 2034. For each project listed, the water system is asked to provide:

- Type of need.
- Reason for need.
- Documentation of need.
- If the project is for new infrastructure or to replace, rehabilitate or expand existing infrastructure.
- If the project is needed now to protect public health or if it is needed over the next 20 years to continue to provide safe drinking water.
- The federal regulation or state requirement if the project is needed to meet a current federal regulation or state requirement.
- Design capacity of source, storage and treatment projects.
- Length and diameter of pipe projects.
- Diameter for projects such as meters.
- Cost of the project (if available).
- Date of the cost estimate (if necessary).
- Documentation of cost (if necessary).

For most of these variables, respondents will choose the appropriate “documentation,” “type of need,” “reason for need,” or “regulation or requirement” from EPA’s “Lists of Codes” (Appendix B). The data collection instrument has been designed to be concise, to avoid jargon and to avoid ambiguous words or instructions. Terms and formats have been standardized to the extent possible. There is no intentional bias in the ordering of the items.

B.3 PRE-TESTS AND PILOT TEST

B.3.a Pre-tests

For the 2007 DWINSA the data collection instrument and some policies were modified substantially. EPA conducted two pre-tests of the data collection instrument for the 2007 DWINSA. These pre-tests were conducted by EPA’s contractor, The Cadmus Group, Inc. The pre-tests gathered feedback on the effectiveness of the data collection instrument; highlighted imprecise, ambiguous or redundant questions; and indicated where further inquiry was needed. A pre-test was held in both Maine (four participants) and Montana (three participants). These states were chosen because they were both partially surveyed states and therefore most of their systems did not participate in the 2007 DWINSA. Also, the contractor

conducting the pre-tests has offices in both these states and by conducting the pre-test in these states was able to reduce costs. The names of the seven systems were provided to EPA by the 2007 DWINSA state contacts. Based on the comments received, EPA made modifications to the data collection instrument. Since EPA's pre-tests of the 2007 DWINSA data collection instrument were so extensive, and because few changes have been made to the data collection instrument since the 2007 DWINSA, EPA believes that a pre-test is not needed for the 2015 DWINSA.

The data collection instrument was further modified for the 2011 DWINSA by the addition of questions and codes to gather information on projects with "green" and climate readiness attributes. Consequently, EPA conducted a limited peer review focused on these new questions. EPA did not conduct a pre-test of the 2011 DWINSA data collection instrument. Based on the limited number of states that submitted projects with "green" or climate readiness attributes indicated in the 2011 effort, EPA concluded these attributes were likely underreported. For the 2015 DWINSA, the "green" and climate readiness questions will be removed from the data collection instrument. Instead, EPA will explore streamlined approaches that might enable the Agency, during the Survey review process, to identify and flag projects that are likely to have "green," climate readiness or climate resilience attributes.

B.3.b Pilot Test

To eliminate unnecessary burden on states and water systems, it has been decided that no pilot test for the 2015 DWINSA will be conducted. A pilot test was conducted for the 1995 DWINSA and consisted of 60 CWSs from New York and Texas.

One change for the 2015 DWINSA is that EPA will deliver all data collection instruments electronically rather than mailing hard copies, as was done for the 1995, 1999, 2003, 2007 and 2011 DWINSAs. Several states received and used electronic data collection instruments in the 2011 DWINSA and states attending the May 2014 workgroup meeting expressed no concerns about using the electronic format. EPA therefore believes this approach has been well tested and has proven to be successful; therefore, it is not necessary to repeat this testing step.

B.4 COLLECTION METHODS AND FOLLOW-UP

B.4.a Collection Method

The proposed collection method is an electronic survey. The data collection instrument including the Lists of Codes will be sent to the states via e-mail. State drinking water agencies will provide the data collection instrument (with the project table prepopulated for systems that participated in 2011 and blank for those that did not) and other necessary documents to the systems in the sample. They will follow-up if the data collection instrument has not been returned in 30 days. For a complete description of the follow-up procedures proposed to increase the response rate, see section B.2.c.ii.

B.4.b Survey Response and Follow-up

The target response rate (defined as the ratio of responses to eligible respondents) for the 2015 DWINSA is 90 percent. EPA realizes that this is an ambitious target, but EPA believes that there are special circumstances that warrant such a target. Also, overall response rates of 94, 97, 96, 93 and 97 percent were achieved in the 1995, 1999, 2003, 2007 and 2011 surveys, respectively. In the first five surveys, EPA conducted the following activities to achieve that high response rate.

- ***Seek Support from the Respondent Population.*** This is a national survey of infrastructure needs for drinking water systems. EPA will work to bring to the attention of water systems, as well as

all national organizations representing these systems, the importance of the DWINSA results. As with the previous five surveys, all national organizations will be contacted by EPA to seek their endorsement of the DWINSA and to communicate to their members the importance of a high response rate to their members. As discussed in Section B.2.c, in past surveys, organizations have provided access to their newsletters and magazines to publicize and endorse participation in the DWINSA. For the 2015 Survey, EPA will seek similar efforts by these organizations.

- ***Follow-up by States and Respondent Peer Groups.*** Since a majority of states have indicated their willingness to participate in follow-up activities, EPA has requested that state personnel, most of whom are personally familiar with the respondents, conduct follow-up procedures including the use of reminder letters and telephone calls to systems that have not responded with the needed information or documentation. If the follow-up fails after three attempts (one reminder letter plus two telephone follow-up calls), EPA is planning to shift to a second approach of peer-group follow-up by members of a trade association, such as AWWA.

B.5 ANALYZING AND REPORTING SURVEY RESULTS

B.5.a Data Preparation

State personnel will check all cost data and documentation to ensure that it is consistent with state and national standards. States will then send the completed and reviewed data collection instruments to EPA for a second round of review by EPA contractor staff.

Once data have been checked, the contractor will key and verify the data. Senior data entry staff will be used for the verification process to improve quality control. Editing will include automated logic and range checks and checks for missing data. Missing cost data will be modeled, using other information provided by the respondents on the data collection instrument. When modeling is insufficient, missing data will be imputed using standard methods such as cell means and regression. The sample of water systems will be weighted so that stratum estimates can be summed to prepare state-level estimates for the 2015 DWINSA.

B.5.b Analysis

EPA will prepare a report that tabulates the results of the 2015 DWINSA and explains the precision of the estimates of total capital needs. Examples of statistics that will be produced include:

- Total capital needs by state and by types of need.
- Total capital needs by domains within the total population, e.g., systems serving populations greater than 100,000.
- Standard errors calculated for key statistics.

The analysis will be similar to that of previous DWINSAs.

B.5.c Reporting Results

The 2015 DWINSA results will be made available to EPA and the public through:

- A printed report that is submitted to Congress on drinking water infrastructure needs. This report will be made available to all participants in the 2015 DWINSA and the public through EPA's Safe Drinking Water website.
- Desktop computer access to state data on the DWINSA Web site without modeled project costs (each state can access only its own data).
- Desktop computer access to the entire data system (EPA only).

A report providing the cost models used to develop costs for the 2015 DWINSA will be made available to EPA and the public through EPA's Safe Drinking Water Web site.

