

## **PART B OF THE SUPPORTING STATEMENT**

### **B.1. SURVEY OBJECTIVES, KEY VARIABLES, AND OTHER PRELIMINARIES**

#### **B.1.a. Survey Objectives**

The overall purpose of conducting this survey is to supply the Office of Ground Water and Drinking Water (OGWDW) with accurate, up-to-date information on the water systems it regulates so that it may comply with statutory requirements and executive orders, support the administration's recommendations for reauthorizing the Safe Drinking Water Act (SDWA), and accomplish its many other program responsibilities as described in Part A, Sections A.2.a, A.2.b, A.3.d, and 5.c.

In accomplishing this purpose, the survey will achieve the following objectives:

- Help to determine the cost of any new federal drinking water regulations.
- Help predict the ability of water systems of various sizes to absorb additional regulatory costs imposed by EPA.
- Determine the extent of need by water systems for technical and financial assistance in maintaining existing operations and in meeting current Federal requirements.
- Determine the collective impacts of all water regulations implemented since the last survey in 2000.
- Help guide implementation of significant new administration initiatives.
- Assist the agency in crafting regulations that are sensitive to small system needs and capabilities.
- Guide states and other assistance providers in targeting needy water systems and developing technical assistance programs.
- Help determine the operating and financial factors associated with successful regulatory compliance, and identify factors associated with poor compliance records.
- Help to verify the accuracy of information contained in the Safe Drinking Water Information System (SDWIS).

#### **B.1.b. Key Variables**

To satisfy the objectives of the survey the following kinds of essential information for the water systems will be collected:

- Size of population served.
- Type of ownership.
- Type, number and geographic distribution of water source.

- Volume of water pumped and delivered.
- Type and extent of distribution system.
- Types of water treatment processes employed.
- Revenues (by source).
- Expenditures (by source).
- Rate structure.
- Sources of capital.

**B.1.c. Statistical Approach**

The survey will be of a scientific probability sample of community water systems (CWSs). The inferences to be made to the entire target population will satisfy the survey objectives and the information needs of EPA’s programs. A sample survey is being conducted instead of a census survey because the latter would be prohibitively expensive, and unnecessarily burdensome to CWSs.

The survey is being designed and conducted with the assistance of a contractor:

Contractor	Contractor Roles
The Cadmus Group, Inc. 57 Water Street Watertown, MA 02472	<ul style="list-style-type: none"> <li>• Technical oversight for all contractor activities.</li> <li>• Questionnaire design and testing.</li> <li>• Oversight of field operations, including site visits and administration of questionnaire.</li> <li>• Statistical sample design.</li> <li>• Data processing.</li> <li>• Training, mailing, and logistics.</li> <li>• Preparation of tables and final statistical report and survey documentation.</li> <li>• Data collection and data processing.</li> </ul>

**B.1.d. Feasibility**

For medium, large, and very large systems, the survey questionnaires have been designed with the capabilities of the typical respondent in mind. A major portion of the survey questionnaires are refinements of the 2000 survey. This survey provided guidance in developing questions that respondents would be able to answer. Generally, the respondent should be able to obtain the requested information from readily accessible records and reports kept by the system.

To reduce the burden on small systems, EPA contractor engineers will visit the water systems that serve 3,300 or fewer people. The contractor engineers will spend approximately an hour with the system owner or operator, requesting information that will be helpful in estimating system infrastructure needs. The engineer will then conduct a physical inspection of the system to confirm information provided by the owner or operator.

Reliance on site visits to small CWSs was strongly recommended by the EPA workgroup to avoid problems faced in past surveys of small CWSs:

- *Total non-response.* Since many systems have not clearly identified responsible parties, and since responsible parties often are reluctant to respond to data collection instruments, it is difficult to use a mail or telephone survey to obtain the necessary information.
- *Item non-response.* System owners and operators often do not know the details of the operating or financial characteristics of their systems. Because EPA contractor engineers will conduct site visits to gather data, item non-response should be eliminated.
- *Reliability.* State drinking water regulators are suspicious of information provided directly from owners or operators of small CWSs. Unlike larger systems, small CWSs usually do not have professional, certified operators. Instead, one is likely to meet trailer park owners, volunteers from homeowners associations, and others who are not water supply professionals.

Finally, employing site visits will substantially reduce the burden on small CWSs. Total burden on the systems, on average, will be about an hour. Instead of completing a data collection instrument, the system owner or operator can answer questions asked by the visiting engineer. The approach was discussed with knowledgeable state drinking water regulators, as well as representatives of small CWSs, and all parties agreed that it was the best approach to achieve the desired results of the survey.

Sufficient funds are available in the existing contract to complete the survey.

The time frame for the survey is acceptable to users of the data within the OGWDW.

This page intentionally left blank.

## **B.2 SURVEY DESIGN**

### **B.2.a. Target Population and Coverage**

The target population is CWSs in operation in 2007. A CWS is a public water system 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).

### **B.2.b. Sample Design**

The CWSS will be based on a nationally representative sample of CWSs. The survey will use a stratified random sample design to ensure the sample is representative. The sample will be stratified by several characteristics of water systems to increase the efficiency of estimates based on the sample. To improve the quality of the data collected, the survey will be administered on-site for small systems; to limit the travel costs involved in visiting each small system in the sample, they will be selected in geographic clusters in a two-stage design.

#### ***B.2.b.1. Sample Frame***

The sampling frame is developed from the SDWIS. SDWIS is a centralized database of information on public water systems, including their compliance with monitoring requirements, maximum contaminant levels (MCLs), and other requirements of the SDWA Amendments of 1996. The following information will be extracted from SDWIS for the statistical survey:

- Name of system.
- Address of system.
- Population served.
- Total design capacity.
- Number of connections.
- Primary source (surface water or ground water).
- Public water system identification number (PWSID).
- Ownership type.
- Consecutive system (i.e., does system purchase or sell water).

From these data, we will develop a list frame from which we will (1) calculate summary statistics for use in calculating sample size, and (2) randomly choose systems within the design strata which will take part in the survey.

#### *Justification For The Use of SDWIS*

SDWIS is the appropriate sampling frame because:

- 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.

SDWIS is the ideal choice for a sample frame because of its inclusive coverage of all units of observation for this survey. 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.

The use of SDWIS as a sample frame in previous surveys was subject to some criticism. 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 this survey are rare. Audits show that systems are misclassified by population or source in fewer than one percent of all cases.

The EPA is taking several steps to prepare SDWIS for use as a sample frame for the 2007 DWINSA. Problematic data will be identified and the states will be asked to review the data and provide any necessary changes. These data will serve as the sample frame for the 2007 CWSS.

### ***B.2.b.2. Sample Size***

The domains of the population of interest for the OGWDW are based on two major characteristics of the systems.

- **The primary source of water.** Systems that rely primarily on ground water are distinguished from surface water systems.
- **The size of the population served by the system.** Eight size categories will be used: systems that serve less than 100 people; systems that serve 101 to 500 people; systems that serve from 501 to 3,300 people; systems that serve from 3,301 to 10,000 people; systems that serve from 10,001 to 50,000 people; systems that serve 50,001 to 100,000 people; systems that serve from 100,001 to 500,000 people; and systems serving more than 500,000 people.

The two water sources and the eight system sizes produce 16 strata.

The regulatory impact models require reasonably precise parameter estimates from each of these domains. The sample size in each domain should be large enough to provide a sufficient number of completed questionnaires to obtain estimates with reasonable precision. Table B-1 shows the number of systems in the sample frame and the minimum sample size required to obtain an estimate for a 50 percent statistic with an error not exceeding  $\pm 10$  percentage points (except for a 1 in 20 chance) in each domain. We will take a census of systems serving more than 100,000 people.

We used a 50 percent statistic because the standard error is largest when the population percentage is 50 percent. The error will be smaller for other population percentages.

### ***B.2.b.3. Stratification Variables***

The sample is stratified to achieve two goals. First, stratifying the data allows us to draw inferences about specific population domains. For example, EPA may wish to draw conclusions about systems serving populations of less than 10,000 or 3,300. We can ensure that estimates of the sub-populations will meet the required levels of precision by drawing the necessary number of observations for each stratum.

The second goal achieved by stratifying the data is that we can increase the efficiency of our estimates by grouping systems into relatively homogeneous strata. The strata were chosen to

minimize the differences among systems within strata, and to maximize the differences among strata. Based on the results of previous surveys, we assume there are important differences in the way systems are operated and in their finances across the strata selected. The operating characteristics and treatment requirements of ground water systems tend to be different from surface water systems. The operating and financial characteristics of large systems tend to be more complex than small systems. System management, and the resources available to it, also may vary by system size.

**Table B-1. Frame and Sample Sizes by Strata**

Source of Water	Population Served	Frame Size	Sample Size
Ground	100 or less	12,503	96
	101-500	14,827	103
	501 - 3,300	12,063	158
	3,301 - 10,000	3,672	94
	10,001-50,000	2,330	93
	50,001 - 100,000	352	76
	100,001 - 500,000	188	188
	More than 500,000	23	23
Surface	100 or less	268	71
	101-500	573	83
	501 - 3,300	1,164	89
	3,301 - 10,000	953	88
	10,001-50,000	1,043	88
	50,001 - 100,000	298	73
	100,001 - 500,000	300	300
	More than 500,000	69	69
All		50,626	1,692

The size of the population served and the primary source of water are the explicit stratification variables. The sample frame will be divided by eight population sizes: less than 100; 100 – 500, 501 – 3,300, 3,301 – 10,000, and 10,000 – 50,000, 50,001 – 100,000, and 100,001 – 500,000, and more than 500,000. Systems will be further divided by two source-water categories: ground water and surface water.

***B.2.b.4. Sampling Method***

Systems serving populations greater than 100,000 will be selected with certainty. For systems serving populations of 100,000 or fewer, the sampling method will be an equal probability systematic sample within each explicitly defined stratum. There are 12 such strata, given by the intersection of the six population size classes whose systems are not being selected with certainty and the two water sources. Within each stratum, the data will be sorted by EPA Region and size

of the population served. This ensures the geographical dispersion among the sample systems and increases the probability that a range of population sizes within each explicit stratum is sampled. The total sample size – including the strata that will be sampled with certainty – will be 1,703.

The survey statisticians will prepare detailed specifications to direct the sampling, and to document the process. These specifications will ensure the sample is drawn in conformity with the sample design and in a statistically valid manner. Standard statistical software will be used to draw the sample.

#### ***B.2.b.5. Multi-Stage Sampling***

In order to achieve the required precision, reduce the burden to small systems, and to keep costs down, a two-stage cluster sample will be used for systems serving fewer than 3,300 people. The use of a two-stage sample design will result in slightly reduced precision for the stratum-level estimates.

##### *First-Stage Sample*

All small CWSs will be assigned to a county (or county equivalent in jurisdictions that do not have counties). Data on all small CWSs will be sorted by county so that EPA can determine the number of systems, by strata, in each county. If a particular county does not contain the required number of systems (a minimum of 6 systems), it is grouped with an adjacent county; the combined county group is referred to as a county-cluster. The first-stage sample will be approximately 120 counties, selected with probability proportional to size, where size is a composite measure of the number of small systems in each county. This method ensures that counties with more CWSs serving 3,300 or fewer people have a greater probability of being selected.<sup>1</sup>

States will be given a SDWIS list of small CWSs in the county (or counties) selected in the first-stage sample for their jurisdictions, and EPA will ask states to verify that the systems on the list are active CWSs with populations of 3,300 and fewer and assigned to the appropriate county. If the number of systems in a county is large (e.g., 100 or more), EPA will select a sub-sample of the systems in that county to reduce the burden on the state. This review by the states will produce a clean sample frame for the second-stage sample.

##### *Second-Stage Sample*

In the second stage, a stratified random sample of five systems is drawn from each of the counties or county-clusters selected in the first stage of sampling.

#### **B.2.c. Precision Requirements**

##### ***B.2.c.1. Precision Targets***

To satisfy EPA's decision-making needs, the sample is designed to provide estimates of percentages of error not exceeding 10 percentage points (expect for a 1 in 20 chance) within each domain for an estimated proportion. The domains are shown in Table B-1. For example, suppose

---

<sup>1</sup> This method is based on Folsom, R.E., F.J. Potter., and S.R. Williams, "Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains," *American Statistical Association 1987 Proceedings of the Section on Survey Research Methods*, August, 1987, pp. 792-796.



50 percent of the systems in a domain report that they boost chlorine residuals in their distribution system. EPA could be 95 percent confident that between 40 percent and 60 percent of the systems within this domain boost chlorine residuals.

Systems serving populations greater than 100,000 will be sampled with certainty; therefore, there will be no sampling error in these domains.

### ***B.2.c.2. Nonsampling Error***

EPA will use several quality assurance techniques to maximize response rates, response accuracy, and processing accuracy to minimize nonsampling error. A pre-test will supplement the experience of EPA and its contractors in formulating a strategy to reduce non-sampling error.

The data collection approach proposed by EPA should reduce both overall non-response rates and item non-response. It also should improve both the accuracy and consistency of the data reported by the systems. Small systems will be visited by EPA contractor engineers, which will ensure that the necessary data are collected and are consistent across systems. Technical assistance will be made available to medium, large, and very large systems to help them fill out the questionnaire and to ensure accurate and consistent responses to the survey's questions. Each system will be contacted by contractor staff; also, systems will be provided a dedicated toll-free telephone number they can call for assistance.

In addition, the following steps will improve the quality of the data:

- A brief cover letter will be sent to the respondents that will explain the purpose of the survey and the information to be collected. It will be on official EPA letterhead, and will be signed by a senior EPA official.
- The data collection instrument design, content, and format will be thoroughly reviewed and tested before the survey begins. The approach will be pre-tested to ensure that the necessary data can be collected.
- Items being asked are those that owners or operators of systems should know. We do not ask questions that require monitoring, research, or calculations on the part of the respondents.
- Standardized software will be used for sample selection.
- We expect complete coverage of the target population using SDWIS, as updated by EPA.
- Data will be 100 percent independently keyed and verified.

The Agency will also will develop an electronic reporting form that systems can use to complete the questionnaire. It is anticipated that this form will reduce the burden for systems completing the form and for Agency to perform data entry and improve the accuracy of the data. The electronic versions of the questionnaire will include data validation checks to confirm responses as they are provided. Systems also will have the option of filling out a web-based questionnaire. The web-based version also will include data validation checks to confirm responses as they are provided and will allow respondents to pause in answering the questions and return to complete the questionnaire at a later time .

#### **B.2.d. Questionnaire Design**

The questions contained in the surveys are designed to obtain information on a variety of water system related technical and financial data. Many of the questions seek to obtain qualitative information, which would then be used to create descriptive statistics. Section A.4.b.1. provides the justification for each question in the survey.

In order to limit the burden on respondents and to improve the quality of the data collected, small CWSs will not be required to fill out a questionnaire. As discussed in section B.4, EPA contractor engineers will visit water systems serving 3,300 or fewer people, and will fill out the questionnaire based on the information they collect. Contractor personnel will provide assistance to systems serving populations over 3,300. If the information requested in a question is available in schematics or other documentation of the system, the system may simply attach the documentation to the question rather than fill out the question.

To ensure quality of the data collected by the questionnaire, it was subject to several rounds of close review and comment by EPA, Cadmus, and independent reviewers. (See section A.3.c in Part A for the list of reviewers.) Special attention will be paid to the presentation and layout of the questions, response categories, response recording blocks, and instructions to clarify the questionnaire for respondents.

A copy of the survey instrument is in Appendix C.

## **B.3. PRE-TESTS AND PILOT TESTS**

### **B.3.a. Pre-tests**

The survey instrument was sent to nine water systems. Three systems participated in a focus group discussion of the questionnaire. A fourth system provided comments separately. The other five systems did not comment on the questionnaire; OGWDW is following up with these systems to get any comments they may have on the questionnaire. The questionnaires and instructions were revised in accordance with the results of the pre-test.

### **B.3.b. Pilot Study**

Following OMB approval the, a pilot study of up to 50 respondents will be conducted for both the site visits to small systems and the multi-step data collection for medium and large systems.

The purpose will be to fine-tune any troublesome questions, increase clarity, reduce respondent burden, and test survey processing systems before the surveys are fielded. EPA plans to use the pilot responses for any water systems also selected in the main survey.

This page intentionally left blank.

## **B.4. COLLECTION METHODS AND FOLLOW-UP**

### **B.4.a. Collection Methods**

Site visits will be conducted of small systems — those serving 3,300 or fewer systems. EPA contractor engineers will physically inspect the system and interview system owners and operators. They will then fill out the questionnaire based on the information they collect.

Systems serving populations of over 3,300 will be asked to fill out the questionnaire electronically or by hand. Each potential respondent will be contacted to provide technical assistance and to respond to questions. Systems will be asked to mail in site plans and other diagrams along with the completed questionnaire.

Several measures will be in place to assure the quality of the data collection. The final version of the questionnaires will incorporate lessons learned from the pre-test and pilot test. Complete and detailed question-by-question specifications will be prepared for every questionnaire item, to unambiguously document for interviewers and analysts the meaning, purpose, and context of all questions. Interviewers will receive formal training for interviewing techniques and for specific CWSS topics. Staff with detailed substantive knowledge of water systems will conduct follow up phone calls to address technical issues that may arise. Supervisory staff will monitor the interviews. Special data collection operations (such as tracing) will be used to improve response rates, sample coverage, and locating the correct sampled CWS.

### **B.4.b. Survey Response and Follow-Up**

The response rate is the ratio of responses to eligible respondents. The target response rate for the survey will be 75 percent for systems serving more than 3,300 people and 95 percent for systems serving 3,300 and fewer, for an overall response rate of 81%. These anticipated response rates are based on actual response rates from the 2000 CWSS, factoring in a slight increase resulting from the small systems site visits being conducted in conjunction with the 2007 DWINSA, which historically achieves a higher response rate, and the use for the first time of a web-based questionnaire to facilitate responses.

This page intentionally left blank.

## **B.5. ANALYZING AND REPORTING SURVEY RESULTS**

### **B.5.a. Data Preparation**

The contractor will key and independently verify the data. Senior data entry operators will be used for the verification to assure quality control. Editing will consist of automated logic and range checks, and checks for missing data. Missing data will be imputed by using standard methods such as cell means, regression, and hot deck to supply values for the missing items.

### **B.5.b. Analysis**

The statistical procedures used to analyze the data will be left up to the individual data users. The contractor will prepare a report tabulating the results of the survey and indicating the precision of the resulting national estimates. Examples of statistics that will be produced are:

- Frequency distributions of all discrete variables in the questionnaire.
- Counts of customers being served by water systems in each domain (e.g., the number of active residential connections served by ancillary systems, with fewer than 500 customers, receiving primarily surface water).
- Counts for each domain of interest of water systems with certain characteristics (e.g., the number of treatment technologies for each of the population categories and water sources).
- Mean and median rates by customer category, revenues, expenses, capital expenditures, and other financial data for each domain of interest.

The fully weighted data will be provided on a data file accessible to the EPA program offices for customized analyses.

### **B.5.c. Reporting Results**

The survey results will be made available to the Agency and the public through the following means:

- A printed report of key statistical tables. These survey results will be distributed to all interested offices at EPA. Additional copies will be made available to the general public through the National Technical Information Service (NTIS);
- An electronic copy of the report will be posted to the OGWDW web site; and
- Mainframe access (Agency only).

A report containing the questionnaire, sampling plan, weights, variances, and formulas and response rates will be prepared and distributed with the data. Record layouts, codes and complete file documentation will be developed for Agency mainframe data file users.