

PART B OF THE SUPPORTING STATEMENT

1. QUESTIONNAIRE OBJECTIVES, KEY VARIABLES, AND OTHER PRELIMINARIES

1.1. Survey Objective

The objective of the survey is to collect information to inform decisions regarding how the nation's stormwater regulations should be strengthened and to support the technical and financial feasibility associated with such rulemaking. This Information Collection Request consists of several questionnaires to different target populations. EPA would send one questionnaire to NPDES permit authorities; two questionnaires (long or short) to owners or developers of new and redevelopment projects; two questionnaires to MS4s (one for those currently subject to EPA's Phase I or Phase II stormwater regulations and another for those that are not); and a questionnaire to transportation MS4s. The primary objectives of the NPDES and MS4 questionnaires are to gather information to characterize the current scope, components, and implementation of existing local and state stormwater programs and to estimate the burden and expenditures to comply with and enforce new requirements on stormwater discharges. The primary objectives of the owner/developer questionnaires are to: (1) characterize current building and real estate improvement projects including type, location, and size; (2) characterize the prevalence and type of stormwater controls implemented at new development and redevelopment sites to control long term stormwater discharges; and (3) characterize the operations and financial condition of owners and developers that could be subject to revised regulations.

1.2. Key Variables

The Owner/Developer Questionnaires request information on the following key variables:

- Project characteristics (e.g., type, location, duration, size, land cover, discharge location and method of stormwater conveyance)
- Long term stormwater best management practices and controls
- Stormwater permit and management requirements
- Firm, establishment, and project level financial information

The MS4 and Transportation Questionnaires request information on:

- MS4 characteristics (e.g., Phase I, traditional, State DOT)
- Stormwater program components (e.g., outreach, recordkeeping, training)
- Current MS4 stormwater permit and management requirements
- Long term stormwater best management practices and controls
- Budget and funding sources

The NPDES Permitting Authority Questionnaire requests information on:

- Current state stormwater program components
- Scope and extent of municipal stormwater program
- State stormwater municipal permit requirements
- State industrial and construction stormwater permits and requirements
- Budget

See Part A, Section 4(b) of this ICR for detailed information on the data items for the questionnaires.

1.3. Terminology

1.3.1. Statistical Definitions

In describing the statistical approach for the survey, EPA has used several statistical terms. They are:

- *Element or Sampling unit* is the individual reporting unit from which the questionnaire will collect information. For example, the MS4 questionnaire will be distributed to MS4s, and thus, each MS4 is an element or sampling unit.
- *Target population* consists of every element of interest for the questionnaire. EPA will use the questionnaire data to make statistical inferences about the target population. Statistical inferences include national estimates of certain characteristics, such as estimated number of municipalities with MS4s. Because of the importance of the target population in reporting outcomes, a principal task in the development of a sample survey design is establishing a clear, concise description of the target population for each questionnaire. The definition should identify every element of the target population so that all non-population elements can be excluded.
- *Sample* is the collective term for the sampling unit selected to receive the questionnaire. There are three types:
 - *Census* is a complete enumeration of all elements in the target population. Information about the target population are obtained directly from the reported values. In other words, statistical estimates are not necessary because information is provided for every element in the population.
 - *Statistical sample* is a statistically selected subset of the target population. In a statistical sample, the elements have a known probability of selection. Statistical inferences about the entire target population can be made from the sample data.
 - *Judgment sample* is an arbitrarily selected subset of the target population. In a judgment sample, the elements are selected based upon particular characteristics of interest. For example, EPA might select several MS4s because of unique and innovative stormwater control programs. Because the data are not statistically representative of the target population, they are not used to make statistical

inferences about the entire population. Instead, they are used to provide information about particular aspects of interest. For example, a judgment sample might be used as the basis for costs of innovative control programs. In its surveys, EPA typically receives a few unsolicited (voluntary) completed questionnaires and they are treated as judgment samples.

- *Sample frame* is a list or set of procedures for identifying all elements of a target population. Sample frames are essential to the quality of surveys because sample elements are drawn from them. Sample frames are typically created from one or more data sources. In addition to listing population elements, the resulting sample frame contains information that will be used to draw statistical samples. This information includes addresses and stratification variables used to draw the samples. Usually, each target population is associated with a single sample frame, although EPA is considering the statistically valid option of dual frames as explained later in the section.

1.3.2. Population-Related Definitions

In Part B, EPA has adopted the terminology used by the U.S. Bureau of the Census in describing where people live. Except as noted, the following definitions are quotes from the Census Bureau publication *Geographic Areas Reference Manual* (1994, retrieved April 7, 2010 from www.census.gov/geo/www/garm.html):

- *Counties*: For most States, counties are the primary administrative divisions. There are exceptions, however; Louisiana has parishes, while Alaska has boroughs for the organized portion of its territory. Because a large part of Alaska is not in any organized borough, the State and the Census Bureau cooperatively have subdivided the unorganized portion of Alaska into census areas for the purposes of presenting statistical data. Three States (Maryland, Missouri, and Nevada) each have one city that is governmentally independent of county organization; Virginia currently has 41 such cities. For both legal and statistical presentation purposes, these independent cities constitute primary administrative divisions of their States. Part of Yellowstone National Park in Montana is not within any county, therefore the Census Bureau treats it as the statistical equivalent of a county. The District of Columbia has no primary administrative divisions; the Census Bureau treats its entire area as the statistical equivalent of both a State and a county. (page 4-2)... Rhode Island's counties exist only for the purpose of judicial administration and have no associated governmental structure. In 1960, Connecticut abolished its county governments and transferred their functions to the State government; however, the State retained the former counties for election and judicial purposes. Nevertheless, in both States, the Census Bureau continues to report many types of data for these county-type entities, in part to retain data comparability with earlier censuses and the data sets of other government agencies. (page 4-12)
 - Part B of this ICR supporting statement uses the term "county" to refer only to those counties with primary administrative functions. Consequently, it has excluded counties in Rhode Island and Connecticut.

- *Minor civil divisions (MCD)*: Minor civil divisions are the primary subcounty governmental or administrative units; they have legal boundaries and names as well as governmental functions or administrative purposes specified by State law. The most familiar types of MCDs are towns and townships, but there are many others ... In some situations, the Census Bureau must complete the coverage of subcounty units by creating additional entities called unorganized territories (UTs) that it treats as being statistically equivalent to MCDs. The Census Bureau has established UTs in certain MCD States to account for the part or parts of a county that are not within any MCD or MCD equivalent... (page 8-1)
 - Part B of this ICR supporting statement uses the term “minor civil divisions” to refer only to those entities with governmental functions.

- *Incorporated places*: Incorporated places are established under the authorization of the governments in each of the 50 States. Requirements for incorporation vary widely among the States; some States have few specific criteria, while others have established population thresholds and occasionally other conditions (for example, minimum land area, population density, and distance from other existing incorporated places) that must be met for incorporation (see Table 9-1). The Census Bureau recognizes incorporated places in all States except Hawaii... Puerto Rico and several of the Outlying Areas under United States jurisdiction (Guam, the Northern Mariana Islands, and Palau) also have no incorporated places... Different States recognize a variety of entities as incorporated places. Usually, the designations city, town, village, and borough are most frequent... (page 9-2).
 - Part B of this ICR supporting statement uses the term “*local governments*” to refer to incorporated places and MCDs.

2. TARGET POPULATIONS, SAMPLE FRAMES, AND GENERAL SAMPLE DESIGNS

This section describes the target populations, sample frames, and general sample designs for each of the questionnaires that EPA is considering in this Information Collection Request (ICR). Section 3 provides more information about the different statistical approaches that may be used in developing the sample design for each questionnaire.

2.1. NPDES Authorities

For the NPDES Authorities Questionnaire, the *target population* is the permitting authority for each state and territory. EPA remains the permitting authority in a few states, territories and on most tribal lands, but most states are authorized to implement the NPDES Stormwater Program and administer their own stormwater permitting programs. The questionnaire will be sent to the stormwater coordinator for each state/territory.

EPA created its NPDES *sample frame* using contact lists collected from each EPA Region. The sample frame contains a mixture of state/territory contacts and EPA contacts, depending upon

which entity is responsible for the NPDES permits. For each contact, the sample frame identifies the contact name, contact title, mailing address, telephone number, and email address.

The *sample design* would be a census, because EPA intends to collect information about every state/territory. If any state/territory does not respond after several contacts, EPA staff will complete the questionnaire using publicly available information and best professional estimates.

2.2. Municipal Separate Stormwater Systems (MS4)

Municipal Separate Stormwater Systems (MS4) are conveyances or systems of conveyances that are:

- Owned by a state, city, town, village, or other public entity that discharges to waters of the U.S.;
- Designed or used to collect or convey stormwater (including storm drains, pipes, ditches);
- Not a combined sewer; and
- Not part of a Publicly Owned Treatment Works (sewage treatment plant).

MS4s may be classified in two categories:

- *Traditional* which are MS4s operated by local governments; and
- *Non-traditional* which are generally operated by non-local entities such as State transportation departments, universities, and hospitals.

Further, MS4s may be:

- *Federally regulated MS4s* if they are currently subject to EPA's Phase I or Phase II stormwater requirements for MS4s; and
- *Non-federally regulated MS4s* if they are not subject to EPA's Phase I and Phase II requirements. However, if they discharge into local waterbodies, operators must obtain a state or federal permit and develop a stormwater management program. In other words, EPA is using the term "federally regulated" to refer only to the Phase I and Phase II requirements.

The following sections describe EPA's plans for collecting data from traditional federally regulated MS4s (2.2.1), non-federally regulated MS4s (2.2.2), transportation MS4s (2.2.3), and other non-traditional MS4s (2.2.4). Within each of the MS4 categories, EPA plans to collect information from sampling units within all States, all Territories and some Tribes.

2.2.1. Traditional Federally Regulated MS4s

For the Federally Regulated MS4 Questionnaire, the *target population* is the governing body (e.g., county, borough, tribal) for the traditional MS4s that are currently subject to EPA's Phase I or Phase II stormwater requirements for MS4s. EPA created its Federally Regulated MS4 *sample frame* using contact lists collected from the stormwater coordinators from each state. For each MS4, the sample frame identifies the phase (I or II) and governing body for each MS4. EPA compared its contact information to Census Bureau lists of governing bodies: 8% are counties; 64% incorporated places; and 24% are MCDs. Of the remaining records, 1.1% are from U.S. territories (primarily Puerto Rico), and 2.6% (175) are other miscellaneous areas or districts as

shown in Table 2-1. Other information in the sample frame includes contact information, population, and land area. However, some MS4-relevant population and land information was not provided, and will be estimated using information from publicly available sources produced by the Census Bureau.

Table 2-1 Traditional MS4s: Other Areas and Districts

Army Housing Community	Property Owners Association
City (HI)	Public Utilities Commission
Community Development District	Sanitation & Flood Control District
Community Services District	Sanitation District
Conservation & Reclamation District	Sewer District
Drainage District	Sewer Services
Flood Control & Water Conservation District	Unincorporated
Flood Control District	Utility District
Improvement District	Valley
Irrigation District	Village (GA, MN, NV)
Levee Improvement District	Water Control District
Metropolitan Agency	Water Control/Improvement District
Metropolitan District	Water District
Metropolitan Stormwater Authority	Water Reclamation District
Municipal Utility District	Water Users Association
Municipality	Watershed District

EPA plans to select a statistical sample that includes MS4s of different population sizes, land areas, population densities (surrogate for impervious surfaces),¹ and locations throughout the United States. Because the number of local governments varies considerably from state to state, EPA intends to control the selection procedure so that states with the most MS4s do not dominate the sample. The statistical sample design may be a *cluster* design (e.g., state is a cluster) or a *stratified* random design. (Section 3 provides more information about the design approaches.)

EPA will consider whether the statistical sample should be supplemented by a judgment sample to obtain information from traditional MS4s with unique characteristics.

2.2.2. Non-Federally Regulated MS4s

To consider the potential impacts of expanding its municipal stormwater requirements to additional traditional MS4s, EPA developed a questionnaire for local governments that are not currently subject to Phase I or Phase II stormwater requirements for MS4s (i.e., “non-federally regulated MS4s”). The questionnaire is short with relatively few questions. The following paragraphs describe the selection approach that EPA is considering for the Non-Federally Regulated MS4 Questionnaire.

¹ To verify that population densities is a suitable surrogate for impervious surfaces, EPA intends to compare the population densities to impervious surface maps.

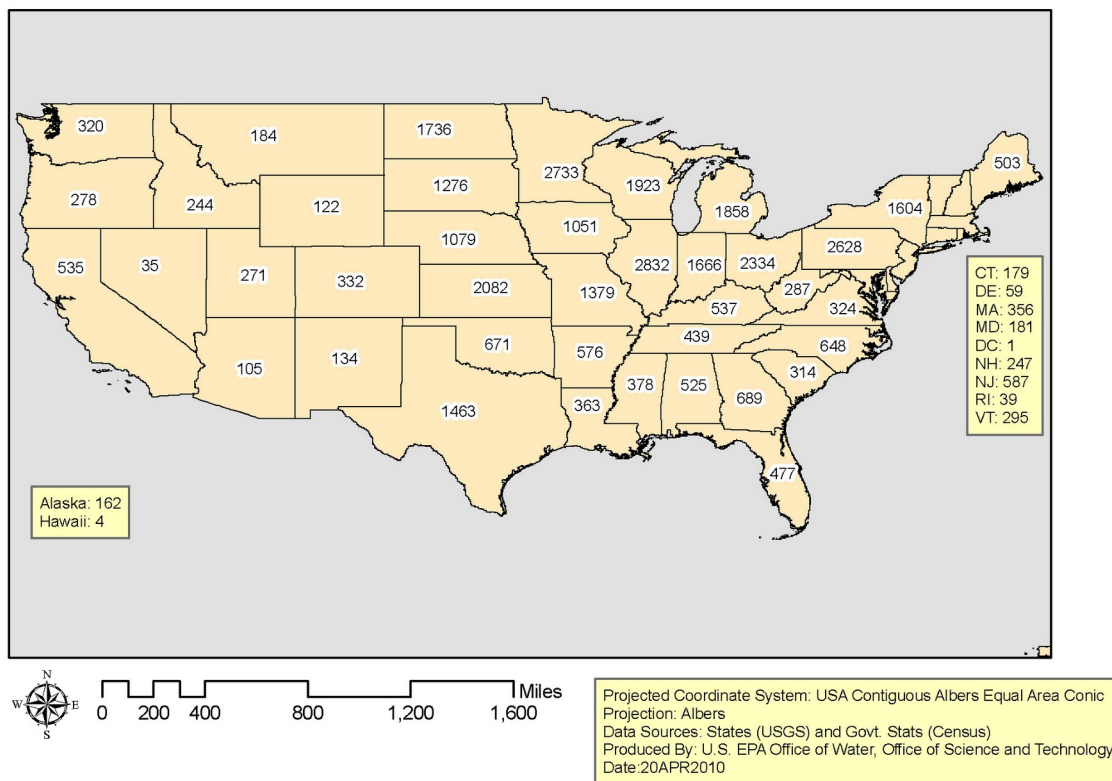
As it intends for the Federally Regulated MS4 Questionnaire, EPA would obtain information from local governments that can reply on the behalf of its community (see discussion of Census terms in Section B.1.3.2). For this questionnaire, the *target population* definition would be local governments that met one of the following definitions, or a close variation:

1. Areas that are not currently subject to Phase I or Phase II stormwater requirements. First, EPA would obtain a publicly available list of all local government entities, such as those maintained by the Census Bureau (see, for example, 2007 Governments Integrated Directory (GID) which is “The master list of all governments in the U.S.”² From this list, EPA would extract the county, municipality, and town/township information for its list of local governments. Figure 2-1 provides a map with the number of local governments in each state. After identifying the local governments, EPA would exclude local governments:
 - Listed in its federally regulated MS4 sample frame.
 - Within counties with a county-level federally regulated MS4, because, presumably, the county MS4 would apply to any sub-government (e.g., city) without its own MS4.The resulting, smaller, list would be the sample frame of non-federally regulated MS4 entities. If EPA used this target population definition, it would obtain information about stormwater practices and financial data from a broad spectrum of non-federally regulated MS4s, located both in urban and rural locations.

Figure 2-1: Number of Local Governments by State

² http://harvester.census.gov/gid/gid_07/options.html, retrieved April 7, 2010

**Total Number of Local Governments by State
(County, Municipality, Town/Township)
Source: 2007 Census of Governments**



2. Areas that, based upon the 2000 Census data, have the following characteristics:³
 - a. Are outside urbanized area populations (i.e., would not have been subject to Phase I or Phase II stormwater requirements in 2000)
 - b. Have a population greater than 10,000
 - c. Have a population density of 1,000 people or more per square mile

3. Areas that, based upon Census data, have been experiencing rapid growth in the last decade, and thus, are likely to meet the size requirements of federally regulated MS4s in the near future. To select incorporated areas experiencing such growth, EPA would probably use one of the following approaches or a similar approach:
 - a. A calculator tool such as MABLE developed by the Missouri Data Census Center (<http://mcdc2.missouri.edu/websas/geocorr2k.html>). Because this tool analyzes 2000 Census block and tract data, it would be possible to identify rapid growth in relatively small areas and link them with the appropriate governing body (e.g., unincorporated land near a city but governed by a county).

³ EPA previously summarized this information for each state and linked the information from <http://cfpub.epa.gov/npdes/stormwater/urbanmaps.cfm>. Using the Census sources directly, EPA has consolidated this information for all states.

- b. Estimated growth using yearly population estimates from the Census Bureau (see, for example, <http://www.census.gov/popest/datasets.html>).
- c. Coding provided by other government agencies. For example, the Economic Research Service (ERS) of the United States Department of Agriculture (USDA) categorizes counties using Urban Influence Codes. ERS describes the codes on its website (<http://ers.usda.gov/Briefing/Rurality/urbaninf/>) :

The 2003 Urban Influence Codes divide the 3,141 counties, county equivalents, and independent cities in the United States into 12 groups. Metro counties are divided into two groups by the size of the metro area—those in "large" areas with at least 1 million residents and those in "small" areas with fewer than 1 million residents. Nonmetro micropolitan counties are divided into three groups by their adjacency to metro areas—adjacent to a large metro area, adjacent to a small metro area, and not adjacent to a metro area. Nonmetro noncore counties are divided into seven groups by their adjacency to metro or micro areas and whether or not they have their "own town" of at least 2,500 residents.

With any of these target populations, EPA would select a *statistical sample* that would contain areas with a mixture of population sizes, population densities, land areas, rural-urban continuum codes,⁴ and geographic locations. To obtain a more balanced mixture of such characteristics, EPA might use information in the sample frame to stratify the target population. In addition, because the number of local governments can vary considerably from state to state (e.g., Nevada has 35; Indiana has 2832), EPA may use statistical approaches, such as *cluster sampling*, to ensure that states with few local governments are appropriately included in the statistical sample. In addition to the statistical sample, EPA might select a *judgment sample* to obtain specific information about unique communities, unique stormwater practices, and/or other unique characteristics.

2.2.3 Transportation MS4s

For the Transportation MS4 Questionnaire, the *target population* is the transportation department in each state/territory that is currently subject to Phase I or Phase II stormwater requirements for MS4s. EPA also is considering an expansion of the target population definition to include similar county transportation departments and highway projects.

EPA created its Transportation MS4 Questionnaire *sample frame* using contact lists collected from each EPA Region and information collected from state/territory websites. For each contact, the sample frame identifies the contact name, contact title, mailing address, telephone number, and email address. The *sample design* may be a census, because EPA intends to collect information about every state/territory. If any state/territory does not respond after several contacts, EPA staff will complete the questionnaire using publicly available information and best professional estimates.

⁴ The Economic Research Service of the United States Agriculture Department provides a 9-part codification that distinguishes metropolitan counties by size and nonmetropolitan counties by degree of urbanization and proximity to metro areas. (Description retrieved from <http://www.ers.usda.gov/data/RuralUrbanContinuumcodes/>, April 7, 2010.)

EPA also has a list of federally regulated MS4s that includes county transportation departments and highway projects. After evaluating the comments received on the second Federal Register Notice for the ICR, EPA will determine whether its transportation target population should be expanded to include these county transportation departments and highway projects. EPA also will consider whether the county DOTs should be limited to those specifically identified in EPA’s records with road projects, or expanded to all county DOTs. Depending on its intended use for the data, EPA then will consider whether it should select a statistical sample or a judgment sample. For either type of sample, EPA would include counties/projects from various locations throughout the United States.

2.2.4 Other Non-Traditional MS4s (no questionnaire planned)

Other than transportation MS4s discussed above, EPA has not included non-traditional MS4s in the survey effort. As shown in Table 2-2, non-traditional MS4s cover a range of different purposes. The summary was created from a list of non-traditional MS4s subject to Phase I or Phase II stormwater requirements identified by the stormwater coordinators from each state. For each non-traditional MS4, the list identifies the type of MS4 (e.g., hospital, university, federal) and contact information. Because the non-traditional MS4s vary in purpose, EPA is considering a data collection that will target only a few of the non-federal entities. The transportation MS4s will be addressed by the Transportation MS4 Questionnaire and federal entities through inter-agency consultations.

Table 2-2 Types of Non-Traditional MS4s

Type of Non-Traditional MS4	Number of MS4s Identified	Number of States With This Type
Agricultural Association	1	1
Airport	9	7
Conservancy District	7	2
Diking Improvement District	1	1
Exposition Site	1	1
Highway	130	40
Hospital	18	8
Metropark	7	3
Metropolitan Park District	1	1
Other	2	2
Park District	2	1
Public School District	93	5
Satellite Tracking Station	1	1
Stadium	4	3
State	151	23

Transit	8	3
University	271	36
Water and Sewer Authority	1	1

2.3 Owners and Developers of New and Redevelopment Projects

EPA has developed two versions of the Owner/Developer Questionnaire: short and long. The short and the long versions include the same basic questions, with additional, more detailed questions appearing on the long version. The two versions have the same target population, sample frame, and general sample design, and they are discussed in the following sections.

2.3.3 Target Population

The *target population* is the same for both questionnaires. The target population is owners and developers of new and redevelopment projects. Owner is defined as the firm, individual, or institution for which the project is being built. Developer is defined as a person, business, or partnership that controls project design and/or land development activities associated with a project. In some cases, the developer may also be the owner.

2.3.4 Sample Frame

EPA is considering a statistical approach that will select statistical samples independently from two sample frames, or a *dual frame*. Traditionally, questionnaire surveys are conducted by taking a sample from a single sampling frame that lists all known members of a target population. In some cases, it may be necessary or useful to sample from multiple sample frames that, as a whole, cover the target population. This is the case when it is either difficult to create a single sample frame or when there are several different organizations that provide information about different subsets of the population.

Specifically, for the Owner/Developer Questionnaire, EPA is evaluating the use of two data sources either to: 1) merge and create a single sample frame; or 2) select samples from each frame (dual frame approach). EPA's preference is to combine the information into a single sample frame (#1), but the available information may not allow linkages of the same entities in each database to be easily and efficiently determined, making the second approach more practical. For example, a company may be listed with slightly different mailing addresses in each data source, and thus, appear to be separate entities. The advantage of the dual frame approach is that linkage only needs to occur for the members that are selected from each sample frame. This is the case because only those members that are selected will be assigned a survey weight for purposes of data analysis. The next two sections describe the two data sources.

2.3.4.1 Dun and Bradstreet

“MarketPlace Pro,” formerly known as the Duns Market Identifiers (DMI) register, is maintained by Dun & Bradstreet (D&B) and covers the entire United States economy. DMI is a file

produced by D&B, Inc., that contains basic company data, executive names and titles, mailing and location addresses, corporate linkages, employment and sales data on over 10 million U.S. business establishment locations, including public, private, and government organizations. DMI is the only comprehensive publicly available database to provide coverage of business establishments. DMI is updated monthly and its coverage of the target population is relatively complete, but often contains out-of-date entries that can introduce inefficiencies in the sample design.

DMI provides the option of choosing alternative organizational levels. DMI defines a headquarters as a business establishment that has branches or divisions reporting to it, and is financially responsible for those branches or divisions. The headquarters record provides the total number of employees for the company, including the employees in the branches. Another corporate family linkage relationship provided by DMI is the subsidiary to parent linkage. A subsidiary is a corporation with more than 50 percent of its capital stock owned by another corporation and will have a different legal business name from its parent company.

Based on both primary and secondary NAICS codes (secondary NAICS codes for a specific entity may include up to 5 NAICS codes other than primary), EPA obtained approximately 740,000 records, screened for duplicates, from the D&B database. The following NAICS codes were used for selection:

- 236115 Single-family Builders, 236116 Multi-family Builders, 236117 Operative Builders, and any business with only a four digit NAICS code and listed as 2361 (i.e., all businesses in 2361 except 236118)
- 236210 Industrial Builders, 236220 (2362) Commercial and Institutional Builders
- 237210 (2372) Land Subdivision
- 237310 (2373) Highway, Street, and Bridge Construction
- 237990 (2379) Heavy and Civil Engineering Construction

2.3.4.2 Reed Connect

EPA also obtained detailed project and firm information from Reed Connect, a product of Reed Construction Data.⁵ Reed Connect provides information about nonresidual and multifamily residential projects. Projects tend to be relatively large, requiring subcontractor support. Project data reported by Reed that are relevant to this analysis include project categorization, project location, company contact information, company categorization, project value, and several building characteristics (e.g., site size and constructed square footage).

2.3.5 General Sample Design

As explained in Section 2.6.2, EPA is considering a dual frame approach to the *statistical sample* design for the owner/developer questionnaire. In this type of design, samples are statistically selected independently from each of two frames (i.e., Reed and D&B). Because the project-level information in the Reed database can better identify which firms meet the target population definition, EPA is considering selecting firms in the Reed database with a higher probability for the long questionnaire than those in the D&B database. It is possible that firms in the Reed

⁵ <http://www.reedconstructiondata.com/construction-project-leads/reed-connect/>

database will only receive a long questionnaire if selected, instead of a short questionnaire. Firms in the D&B database will receive either the short or the long questionnaire. EPA also will consider whether the statistical sample from either sample frame should be supplemented by a *judgment sample* to obtain information from owners/developers with unique characteristics.

Because some firms might have many projects, EPA is considering an approach to minimize the burden associated with reporting project-level information. Instead of reporting for an extended period or a specified number of projects, EPA is considering an alternative that will request each firm to report about its projects that were active on a specified date that EPA will randomly select for each questionnaire. In this manner, EPA would obtain statistically representative data about projects that could be extrapolated to the entire population.

3 CONSIDERATIONS IN SAMPLE SELECTION

This section provides more information about the general sample designs identified in the previous section. It describes the precision targets and statistical approaches to selecting the sample.

3.2 Precision

Precision is the sampling error (variability) associated with estimates calculated from the sample data and extrapolated to the larger target population. One measure of precision is the width of the confidence interval for the estimate. Confidence intervals provide a range of values for a particular estimate that would be likely if the study were repeated an infinite number of times (because, by statistical theory, our sample is only one of many possible samples that could have been selected). Thus, when using 95 percent confidence intervals, 95 percent of such intervals would include the true value, if we could take an infinite number of samples. The precision of the estimates depends on both the sample design and the sample size, that is, the number of elements in the sample.

Because EPA is developing a national rule, it is primarily concerned with the precision of the overall estimates. Consequently, in estimating the overall sample size, EPA intends to impose more stringent requirements for overall estimates than any subpopulation. First, EPA would assume that the sample (unadjusted for nonresponse) would be expected, with a certain confidence level (e.g., 90 or 95 percent), to yield sufficient data to estimate the value of an unknown proportion. EPA is considering a precision target of 90 percent which provides reasonable precision. If EPA were to use a more stringent precision target, such as 95 percent, it would need to collect more data which would increase the burden to the target populations.

Once it determines the overall precision target, EPA then allocates the sample among the different strata. EPA typically requires that each stratum meet a basic level of precision. For example, if the binomial distribution were used, a stratum sample might be selected so that it would be expected, with 90 percent confidence, to yield sufficient data to estimate the value of an unknown proportion to within ± 0.15 of its true value for the target population.

3.3 Statistical Approaches

This section describes the statistical approaches that EPA is considering for selecting samples for each questionnaire. Depending on the target population characteristics, it is possible that EPA may use a different sample design for each questionnaire. For each design, EPA may use the following approaches, either individually or in combination.

In any sample design, if a stratum has a sample size of less than 10, EPA intends to sample all elements within the stratum.

3.3.3 Binomial Distribution

The binomial distribution is often used as the basis of sample designs, and can be used to estimate precision. The binomial distribution applies to situations where there are only two outcomes (yes or no) to a dichotomous question such as “Were stormwater post construction controls that retain stormwater onsite implemented on this project?” The presence or absence of the attribute for a particular project is a dichotomous, or binary, variable. The binomial distribution models these data, based on the notion of obtaining national estimates of the percentage or proportion of projects in the target population (or a subset of the target population) that have a particular attribute. The binomial distribution also provides estimates of the variance that is used to calculate the confidence intervals. Because a proportion of 0.5 (or 50 percent) results in the largest possible variance for the binomial distribution, if EPA uses this approach, it would assume that the probability of one outcome would be 0.5 (e.g., stormwater post-construction controls that retain stormwater onsite were implemented by 50 percent of the respondents). In other words, if the population value is any value other than 50 percent, the survey estimate will be more precise – in statistical expectation – than it would be if the population value is 50 percent.

3.3.4 Stratified Sampling

For a stratified sample design, *stratification* is performed by selecting one or more characteristics of interest provided in the sample frame and dividing the members of the population into the strata based on those characteristics. *Stratified sampling* consists of selecting a sample from within each stratum, then combining them to constitute the total sample. There are several benefits that result from stratifying the population, including:

- Ensuring that the sample contains representatives from every stratum;
- Improving the precision of parameter estimates;
- Allowing important parameters to be estimated at the stratum level; and
- Allowing certain subpopulations of particular interest to be sampled at a greater rate than others.

Classifying the target population by all of the stratification variables leads to a number of cells that increases quickly by multiples of the strata categories. For a simple example, see Table 2-3 with six total cells for two gender and three age classifications.

Table 2-3 Simple Example of Two Stratification Variables Producing Six Cells

Male Infant	Male Child	Male Adult
Female Infant	Female Child	Female Adult

Cells with small population sizes are statistically inefficient, so EPA intends to classify each of the stratification variables into relatively few categories.

EPA is considering the following variables as stratification variables, but may consider others if appropriate:

- Location information for all questionnaires:
 - *Ecological regions*: EPA is considering the Level 1 ecological regions established for North America (www.epa.gov/wed/pages/ecoregions/na_eco.htm#Level%20I). EPA also considered evapotranspiration, precipitation, and other environmental factors, but considers the ecological regions to best categorize the United States for the purpose of the stormwater survey. To minimize the number of strata categories due to the number of regions, EPA intends to combine the smallest regions with another region in the same general location and/or climatic conditions. EPA also is considering a modification that would designate the Chesapeake Bay as a separate region to address Agency initiatives related to the Bay.
 - *Population density*: EPA is considering the use of population density as a surrogate for *impervious surfaces*. EPA intends to compare the results to recent research on the impact of impervious surfaces on watersheds.⁶
 - *Location Relative to Federally Regulated MS4s*: EPA is considering several approaches to identifying areas that are adjacent to federally regulated MS4s and/or growing to a size that may require the MS4 to comply with Phase I or Phase II stormwater requirements. EPA is considering the following approaches:
 - *Census Population Data*: The Census Bureau provides information (www.census.gov/popest/cities/cities.html) about annual population estimates for states, counties, MCDs, and incorporated places for the years 2000 through 2008. These estimates can be used to classify growth patterns.
 - *Rural-Urban Continuum Codes*: USDA/ERS has assigned one of nine Rural-Urban Continuum Codes to each U.S. county (or county equivalent). Metropolitan counties are assigned one of three codes according to the population of the associated metropolitan area in which they are located. Non-metropolitan counties are assigned one of six codes according to the size of their urban population and whether or not they are “adjacent” to a metropolitan area. (The term “adjacent” means that they are physically adjoining at least one metropolitan area, and at least two percent of their working population commutes to and from a central

⁶ For example, “Watersheds at Risk to Increased Impervious Surface Cover in the Conterminous United States,” David M. Theobald, et al, **Journal of Hydrologic Engineering**, April 2009, pages 362-368, retrieved from http://www.whrc.org/resources/published_literature/pdf/Theobaldetal.JHydroEng.09.pdf on April 22, 2010.

- county within the adjoining metropolitan area.)
- *Urban Influence Codes:* USDA/ERS also has assigned one of 12 Urban Influence Codes to counties (and county equivalents). Table 2-4 provides the definitions used for the codes. *Metropolitan* counties are assigned one of three codes. Among non-metropolitan counties, three codes are used to distinguish *micropolitan* counties by whether or not they are adjacent to a metropolitan area, and if so, if this metropolitan area has a population of at least one million. Seven additional codes are used to distinguish non-core counties by whether they are adjacent to a metropolitan area (and if so, whether this area has at least one million people), a micropolitan area, or neither..

Table 2-4 USDA/ERS Definitions Used for Urban Influence Codes

Metropolitan counties (1,089 counties)	Counties located within a metropolitan area. (<i>Metropolitan areas</i> consist of a “central” county having at least one <i>urbanized area</i> ¹ with population at least 50,000, and adjoining counties for which at least 25% of its population commutes to work in the central county, or at least 25% of its employment consists of workers from the central county.)
Micropolitan counties (675 counties)	Counties located within a micropolitan area, but not a metropolitan area. (<i>Micropolitan areas</i> consist of a central county having at least one <i>urban cluster</i> ² of at least 10,000 but less than 50,000, and adjoining counties for which at least 25% of its population commutes to work in the central county, or at least 25% of its employment consists of workers from the central county.)
Non-core counties (1,377 counties)	Counties not located within a metropolitan or micropolitan area.

¹ “Urbanized areas” are areas having an urban nucleus of at least 50,000 people and a core with population density of at least 1,000 people per square mile.

² “Urban clusters” are areas with populations between 2,500 and 50,000 people, and having the same population density requirements as urbanized areas.

- Owner/Developer Questionnaires:
 - *Industry Sector* would be determined by the primary business as identified by the primary NAICS code.
 - *Size* would be determined by business volume such as sales or revenue.
- MS4 Questionnaires:
 - *Population*
 - *Land area*
 - *Phase I or Phase II* (traditionally regulated MS4s only)

3.3.5 Cluster Sampling

A cluster sample is a probability sample in which each sampling unit is a collection, or cluster, of elements from the target population. For example, EPA might consider each state to be a cluster which is comprised of a collection of federally regulated MS4s. If EPA uses this design approach for any of its target population, it would send the questionnaire to a sample of elements within each cluster (e.g., a sample of federally regulated MS4s in each state).

3.3.6 Probability Proportional to Size

The probability proportional to size (PPS) sample design uses size as a factor in selecting the sample. In this design, the sample would include most of the largest elements and fewer of the smaller ones. Size can be defined in a number of different ways, such as population or revenue.

3.3.7 Spatial Sampling

In simplest terms, spatial context is the information required to locate a sample point on the landscape, for example, latitude and longitude. For our purposes, there might be benefits to selecting elements in a manner that would cover the entire country and be selected from different watersheds. Although any random sample could accomplish this in a sense, there might be advantages to placing some spatial constraints on the sample so that the spatial distribution of the sample closely matches the spatial distribution of the population. Although EPA could approximate the latitude and longitude for element, other approaches such as systematic sampling within ecoregions may be simpler and easier to implement.

3.3.8 Systematic Sampling

Instead of stratifying by ecoregions, EPA is considering systematic sampling to incorporate some spatial context into the sample designs. Systematic sampling involves selecting every k^{th} facility where k is determined by the selection rate. For example, for federally regulated MS4s within each ecoregion, EPA might sort the MS4 sample frame by state name, and then zip code within each state, and finally randomly sort the MS4 names within zip code. The next step would be to draw a systematic sample from the sorted list for each ecoregion. In this manner, the sample would be reasonably diverse from a geographical perspective.

4 SOURCES OF ERROR, INCLUDING NONRESPONSE

In developing the final sample design, EPA will consider the precision targets for data collected from the target populations. EPA also will consider potential error that could be associated with estimates calculated from the collected data, due to sources of error associated with sampling, such as response rates, as well as non-sampling sources of error, such as processing error. The following sections describe approaches that EPA may consider to minimize error in data collected using the final design.

4.2.3 Response Rates

In developing the sample design, EPA will consider both unit (questionnaire) and item (question) nonresponse. Response rates compare the number of completed questionnaires to the number that were distributed. EPA expects that the unit response rate would be 80 percent or better for this mandatory survey effort.

The survey would be conducted under the authority of the Clean Water Act. The cover letters and instructions would explain the legal authority, responsibility to respond, reasons for the questionnaire, and penalty for nonresponse. EPA would use reminder letters and/or telephone calls to remind respondents of the duty to respond under authority of the Clean Water Act. If possible, EPA would seek the endorsement of the major trade associations, which would be expected to increase the response rate from its members. EPA recognizes that some nonresponse is unavoidable, and in past survey efforts, EPA has waived the duty to respond in extreme and rare cases (e.g., natural disasters) which also might occur for this survey effort. To ensure that it receives enough completed questionnaires to meet its precision targets, EPA intends to adjust the sample sizes upward by 20 percent. In addition, EPA intends to further adjust the sample sizes upward to account for out-of-scope elements (including out-of-business). For example, a particular county may administer all of the stormwater programs in its county, and thus, all of the local governments (towns, etc.) located in that county would be considered “out-of-scope” for the questionnaire because they do not administer any stormwater programs.

Table 2-5 Sampled Populations: Out-of-Scope Assumptions

Questionnaire	Out-of-Scope Assumptions
Federally Regulated and Transportation MS4	0%
Non-Federally Regulated MS4	25%
Owner/Developer – Short	30%
Owner/Developer – Long	15%

Prior to distributing the detailed questionnaires (units), EPA would probably adjust the initial sample sizes to help ensure that the effective sample sizes (i.e., respondents) would be sufficient for precision requirements. For this reason, EPA would adjust the statistical sample size upwards to account for an estimated nonresponse rate of 20 percent. Nonresponse can result from a number of factors, including undeliverable addresses; out-of-business establishments for which nobody is available to respond; out-of-scope establishments that were incorrectly included in the sample frame; and refusals. EPA typically evaluates each of these components and adjusts its statistical estimates accordingly.

In addition to increasing the initial sample size, EPA would strive to improve the response rate by sending reminder letters and/or telephone calls. Furthermore, after receiving the responses, EPA intends to adjust the questionnaire weights for any nonresponse.

If the nonresponse rate is greater than 20 percent for any questionnaire, EPA will evaluate whether non-respondents appear to have different characteristics than respondents. EPA would examine these characteristics both for the entire industry and for subgroups in the analyses as recommended in the OMB guidance of January 20, 2006 (www.whitehouse.gov/omb/inforeg/pmc_survey_guidance_2006.pdf). For any differences, EPA intends to determine the major causes, and to incorporate appropriate adjustments for bias. (Bias is the difference between the expected value of an estimate and the true value of a parameter or quantity being estimated. If the data collection process generates estimates that are consistently (or on average) above or below the true value, the data collection process is biased.)

To minimize item nonresponse, EPA's subject matter experts have worked closely with MS4s, owners/developers, and states to develop questions that would be easy to understand with clearly defined and familiar terms; are formatted in a logical sequence; and would request data that are readily available. In this manner, EPA expects to minimize inaccurate or incomplete response of the questions that can occur due to misunderstanding or misinterpretation of questions and the unintentional skipping of questions by respondents. Additionally, EPA would operate an e-mail helpline and website to assist respondents with the questionnaires. If necessary, EPA would impute responses to key questions in our analyses.

4.2.4 Processing Errors

Processing errors can occur when questionnaire responses are coded, edited, and entered into the database. The design and implementation of the questionnaire database would employ a number of quality assurance techniques to reduce the frequency of such errors. These techniques may include the following:

- Investigate whether web surveys are practical for any of the questionnaires, which would minimize transcription errors from paper copies
- Double-entry keypunch verification on critical questions
- Computerized comparison of selected responses to detect inconsistencies and illogical responses
- Computerized analyses to screen for out-of-range and inconsistent numerical values
- Computerized analyses to detect missing numerical data and missing units

5 PRETESTS AND PILOT TESTS

EPA does not intend to pre-test the questionnaire. For more than 30 years, EPA's Engineering and Analysis Division has conducted surveys of numerous sectors to collect information to support regulation development activities in the effluent guidelines program. In past years, EPA has relied predominantly on active participation by trade groups in reviewing the questionnaires.

In EPA's experience, such collaboration generally tends to better reflect the sectors at large than pre-tests. For this reason, EPA considers additional review through the pre-test process to be unnecessary for this survey.

6 COLLECTION METHODS AND FOLLOWUP

Please See Part A, Section 5(b) of this ICR for this information.

7 DATA PREPARATION AND ANALYSIS

7.2.3 Data Processing Errors

As explained in Section 6 of Part A of this support statement, EPA may distribute the questionnaires in paper form, electronic PDF, through a letter with a link to the questionnaire for completion online, or some combination. Upon receipt of completed questionnaires, EPA would download the electronic or web responses directly to a database or prepare written responses for data entry. Concurrently, EPA and its contractors would review the questionnaires for completeness and accuracy. As necessary, EPA would perform follow-up calls to clarify inconsistencies in responses. Once the data are entered into a database, numerous manual and electronic QA activities would likely be performed and the results would be provided to engineering and economic staff for further resolution and documentation. This database would then be used to perform data analyses.

7.2.4 Analysis

The data collected through these questionnaires will provide EPA with information to characterize current building, transportation, and real estate improvement projects (i.e., new and redevelopment); long term stormwater controls and best management practices (BMPs) being installed at newly developed and redeveloped projects; state and local long term stormwater programs and requirements (including retrofit of existing development) and the areas covered by these requirements; the current capacity and expenditures by NPDES Permitting Authorities and local authorities to implement, enforce, and maintain long term stormwater programs and controls; and technical, financial, and environmental data needed to quantify the incremental pollutant removals, compliance costs, impacts, and benefits for various regulatory options that EPA might consider in this rulemaking. Ultimately, EPA would use the information to inform whether to expand its national stormwater program and how to best reduce long-term stormwater discharges from new and redevelopment and the built environment.

The objectives of the each questionnaire would be achieved by the statistically-designed sample survey because the resulting inferences and analyses would be as statistically unbiased and as precise as is practicable. EPA would apply sample weights derived from the statistical sample design and adjust for nonresponse to the data during statistical analysis. Weighting the data would allow inferences to be made about the entire target population, including those that did not

respond to the questionnaires. Another advantage is that weighted estimates would have smaller variances than unweighted estimates. EPA would use accepted statistical methods for survey statistics, such as those described in *Sampling Techniques* (Cochran, 1977) and *Survey Sampling* (Kish, 1965). EPA would use the data from the judgment sample separately in a qualitative manner.

See Part A, Section 2(b) of this Information Collection Request for a detailed discussion of the technical and economic analyses.

8 REFERENCES

Cochran, W.G. (1977). *Sampling Techniques*. New York: Wiley.

Dillman, D. (2000). *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley.

Israel, G. (1992) "Sampling Issues: Nonresponse," University of Florida, IFAS Extension Electronic Document. Available at: <http://edis.ifas.ufl.edu/PD006>.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.