

---

# Supporting Statement for United States Energy and Employment Report Data Collection

## **Part B: Collections of Information Employing Statistical Methods**

---

**OMB No. 1910-5179**

---

August 2021



# Table of Contents

---

Part B: Collections of Information Employing Statistical Methods.....	i
B.1. Respondent Universe.....	1
B.2. Statistical Methods.....	2
B.3. Maximizing Response Rates.....	3
B.4. Test Procedures and Form Consultations.....	6
B.5. Statistical Consultations.....	6

---

## B.1. Respondent Universe

**Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used.**

Geographic coverage includes the 50 States and the District of Columbia. Private establishments and government units are included, but establishments must report at least one permanent employee to be included. Data are to be collected for establishments in 150 detailed industries identified to be of specific interest for the Energy Jobs Report (EJR) Survey. The industries are defined using the 6-digit detail of the North American Industry Classification System (NAICS; includes 1,193 6-digit industries). Attached is a table summarizing the survey frame the industry sector level (Attachment B).

The sampling frame is a representative sample of employers based on establishment totals from the Quarterly Census of Employment and Wages (QCEW) Longitudinal Database (LDB) maintained by the Bureau of Labor Statistics, stratified by employment size categories developed by the Census Bureau County Business Patterns data set. The actual contact information and business names are drawn from a private dataset, DataAxleUSA, because the QCEW is confidential. About 1,680,000 establishments with employment of 15.5 million are in the 150 in-scope industries. Due to clustering by size, a total of 1,021,005 establishments are included in the sampling frame.

For the purposes of EJR sample allocation, we aggregate 150 detailed industries into 7 groups of industries or “allocation” NAICS (ANAICS). For most in-scope industries, the ANAICS is the 2-digit NAICS and includes all in-scope NAICS-defined industries within the 2 digit. Within some 2-digit industries, ANAICS splits out specific 5- and 6-digit NAICS industries where we anticipate having a higher incidence of energy activity. ANAICS 2- and 3-digit coding is the same as for NAICS, though restricted to EJR-eligible industries.

Industry sectors are also defined for use in allocation. Industry sectors are 2-digit ANAICS with two exceptions. The manufacturing sector combines three 2-digit codes. The trade sector combines retail trade and wholesale trade.

About 10,000 in-scope “Known Universe” establishments were pre-identified as having “energy” activity. A database of likely energy establishments will be developed internally by Contractor by collecting industry association databases, approved utility contractor lists, and other public and private sources. By comparing the information obtained through these sources and comparing the NAICS codes of these establishments on the QCEW, Known Universe establishments will be matched to the QCEW/DataAxleUSA dataset and a “known” indicator will be used to assist in oversampling “known” establishments.

Contractor will contact about 35,000 establishments per year. The total survey completion targets will be based on a sample selected using the QCEW/DataAxleUSA frame for the second quarter of 2016. Quotas will be established for each NAICS code or ANAICS code by size. Although the precision is not specified, the goal is to publish industry sector data in every State and to obtain useful data at the national and state level for each ANAICS and identified technology area. If possible within budget constraints, some additional information by 6-digit NAICS is also desired. EJR establishments are not separated but are included in columns for the private, Federal government, State government, and local government sectors.

Stratification – The EJR will be stratified by 6-digit NAICS or ANAICS and size class (1-9, 10-19, 20-49, 50-99, and 100+ employees) and systematic samples selected in the noncertainty strata. Known

---

establishments can be of any ownership, are processed separately, and are excluded from the other portions of the frame. Federal government stratification is State by industry sector. State government stratification is State by industry sector. For private establishments (excluding the Known Universe) three levels of stratification are examined during sample allocation: 1) State x industry sector, 2) national ANAICS, and 3) national 6-digit NAICS.

## B.2. Statistical Methods

### **Describe the procedures for the collection of information including:**

EJR panels will have a probability-based sample aimed at satisfying data needs at both the State and industry sector level and the national ANAICS level. The basic sampling unit is an establishment. Response quotas will be established based on the representation of total establishments by 6-digit NAICS, times the proportion of establishments in each size category as identified in the most recent available data from Census Bureau County Business Patterns.

Restricted to in-scope industries, establishment on the QCEW frame are separated into 4 mutually exclusive parts that are separately sampled. Approximate sample counts refer to a sample selected from the QCEW frame for quarter 2 of 2010.

- Known Universe; census, with up to six attempts; stratification industry by size class (can have any ownership code)
- Federal Government; sample 50; stratification state by industry sector
- State Government; sample 50; stratification state by industry sector
- Private; sample 15,000; complex stratification using state, industry

Known Sampling - All establishments in the Known Universe will be contacted up to six times. The responses will be treated separately, and the overall employment from the Known Universe sample will be deduplicated from the appropriate panel of ANAICS, based on the Known Universe respondent NAICS code.

Private Establishments and Government (excluding Known Universe) -The allocation has 4 basic steps.

- Establishments by State - relying on the most recent data available from QCEW, Contractor will determine the proportion of establishments in each selected NAICS, as a percentage of the total establishments in all selected NAICS.
- NAICS Establishments by Size- relying on the most recent data available in the Census Bureau's County Business Patterns, Contractor will determine the proportion of establishments within each size category in each 6-digit NAICS. The total NAICS quota will then be allocated by the size proportions to develop the percentage of total state-level sample.
- Deduplicate Known Universe Establishments from Sampling Universe - verifying by name, NAICS, contact name, address, phone, and other identifying information, Known Universe establishments will be removed from the private, state, and federal government sampling universes.
- Establish Quotas - State-level quotas are established by multiplying the total number of proposed survey completions per state by the percentage established in "Establishments by State" above, and by the percentage established in "NAICS Establishments by Size" above.

Changes in the Sample Design - When EJRC data become available, it will be possible to more efficiently

---

design the sample to meet targeted data needs. It is anticipated that changes in industry scope may be made.

EJR estimation requires a multi- step process. First, a rate is calculated for qualifying firms based on their response relative to conducting qualifying energy activities. This is referred to as the incidence rate and is used to identify the total number of QCEW establishments, by NAICS and State, that conduct energy activities. The second step of the process is to determine the average share of employment in qualifying establishments that are focused on energy-related portions of the business. This is called the energy share rate.

The incidence rate is then multiplied by the Known Universe establishment total, after reduction for churn. The total employment of these qualifying firms is then removed from the QCEW total by corresponding NAICS code (employments and establishments). The total employment of these qualifying firms is also multiplied by the energy share rate to determine energy employment in the Known Universe.

Private and Government (Known excluded) employment is calculated by multiplying the sum of QCEW employment by NAICS code, less Known Universe Employment by NAICS code, times incidence rate by NAICS code, times energy share by ANAICS code.

Total energy employment is calculated by adding Private, Government, and Known energy employment.

Value chain employment is allocated by NAICS industry. Technology employment is allocated by developing a “technology share” percentage in the instrument, by ANAICS.

### **B.3. Maximizing Response Rates**

**Describe methods to maximize response rates and to deal with issues of non-response.**

Employers will be mailed an advanced postcard or letter, followed by a letter with instructions that include the option to call into a call center, visit a web address (each requiring a unique establishment identifier) or wait for a phone call or email from the Contractor. The cover letter pledges confidentiality and explains the importance of the survey and the need for voluntary cooperation. After the mailings are complete, Contractor will send similar correspondence to Known Universe establishment contacts with email addresses, and then interviewers will begin calling establishments who have not responded and attempt to enroll them into the survey. Non-respondents and establishments that are reluctant to participate are re-contacted by an interviewer especially trained in refusal aversion and conversion.

To mitigate possible bias arising from nonresponse, weighting class adjustments to the weights will be made. The process for determining nonresponse weight factors, will differ, depending on whether we are adjusting the known or unknown sample.

#### **Weighting for nonresponse in the known sample**

##### **Step 1: Quantifying the Known Universe of Energy Firms**

The known universe of energy firms will be reverse matched, based on firm name and available contact information (phone number and address) against DATAAXLEUSA’s national database of businesses. This reverse match will provide a consistent designation by traditional industry code (4 to 6 digit NAICS code) and an initial estimate of employment size by firm location. In our experience, the reverse match will

---

identify 65 to 80 percent of known energy firms. The remaining 20 to 35 percent of known firms that are not matched with DATAAXLEUSA's national database will be evaluated individually and using publicly available data, and we will estimate NAICS industry for each firm and an estimate of employment size by location. The portion of known firms that cannot be matched to DATAAXLEUSA's national database or an estimate of industry classification, location information and when available employment by location<sup>1</sup> made will be discarded from the known universe. This analysis will provide a universe of known energy firms that are categorized by industry classification (NAICS code), location<sup>2</sup> (State and zip code), and employment by location<sup>3</sup> size.

### **Step 2: Stratifying the known sample by industry (NAICS), Geography, and Employment size by location**

The sampling plan for the known universe will be a census approach, with the objective of maximizing participation among all potential participants. Completed surveys will be categorized by industry, location, and employment size as is consistent with the known universe. The known universe will also be updated to reflect businesses that have gone out of business or are no longer involved in energy related work or whose original assumptions about industry, location and employment size identified in step 1 have been revised. This second step will allow the known sample to be compared against the known universe among energy firms.

### **Step 3: Comparing the known universe to the known sample and producing adjusted weights**

The adjusted weights will be based on the relationship between all energy employment in the known universe and total employment in the known universe for all energy firms in each stratum (i.e. each stratum will be based on Industry (I), Location (L) and Size of firm (S)) and all energy employment in the known sample and total employment in the known sample for each stratum.

$$\text{Known Universe Energy Employment (ALL FIRMS)} = \sum \text{KEN AF}$$

$$\text{Known Sample Employment (ALL FIRMS)} = \sum \text{Ken AF}$$

$$\text{Each stratum's total known universe employment} = \sum \text{KEN ILS}$$

$$\text{Each stratum's total known sample employment} = \sum \text{Ken ILS}$$

$$\text{Weight for each stratum within the known sample} = \left( \frac{\sum \text{KEN ILS}}{\sum \text{KEN AF}} \right) / \left( \frac{\sum \text{Ken ILS}}{\sum \text{Ken AF}} \right)$$

The weight for each stratum is then applied to each completed survey that falls into that stratum definition.

---

<sup>1</sup> To remain in the known universe file, a firm needs to be reverse matched to DataAxleUSA's national database or have some source of industry description. Employment by location is valuable but is not required to remain in the known universe.

<sup>2</sup> Location is determined to be the entire state, for smaller states, or regions within a state defined by zip code for larger states.

<sup>3</sup> In some cases employment by firm, among multiple locations is the only information that is available and in those cases employment is split equally among the locations it represents to provide an estimate of employment at each location.

---

## **Weighting for nonresponse in the unknown sample**

The unknown sampling plan is typically informed by what was learned in quantifying the known universe of energy firms. It provides some insight into what industries, areas and sizes of businesses energy businesses are more likely to be found. The analysis of the known universe along with a literature review of new and emerging energy industries, provides the basis for the traditional NAICS codes that are to be examined and sampled within the unknown universe.

### **Step 1: Removing known energy businesses from the unknown sample**

The known universe of energy firms will be removed from the relevant NAICS codes that are being sampled in the unknown sample. This will create a revised unknown universe (Total NAICS code universe minus known Energy Firm NAICS universe) in each of the NAICS codes on interest to be examined in the unknown sample, and ensure that employment is not double counted between the known and unknown universes. Estimate for unknown employment and establishments by Industry, Location and Size will be based on QCEW, CENSTATS and a proportional subtraction based on revised unknown employment and establishments.

$$\text{Known} = \sum \text{employment/establishments within a given NAICS code based on Known Universe Energy employment}$$
$$\text{Total} = \sum \text{employment/establishments within a given NAICS code based on CENSTAT's County Business Patterns estimate}$$
$$\text{Revised Unknown Universe Employment by Industry of Interest} = \text{Industry Employment Estimate}^4 * (1 - (\text{Known}/\text{Total}))$$
$$\text{Revised Unknown Universe Establishment count by Industry of Interest} = \text{Industry Establishment Estimate}^5 * (1 - (\text{Known}/\text{Total}))$$

### **Step 2: Develop a sampling plan based on the Industry (NAICS) categories of interest and develop stratum of sampling by Industry, Location and Size of Firm (ILS) based on Revised Unknown Universe of Interest**

The revised unknown universe for each industry of interest will serve as the basis for the sampling plan and the number of establishments and total employment within each stratum. Stratum are again based on Industry, Location and Size of the firm. The sampling plan will determine the size of the sample that is needed for each stratum of the revised unknown universe. By implementing the sampling plan, we will be able to determine the incidence of establishments and their corresponding employment of energy firms within each stratum of the unknown universe. Proper implementation of the sampling plan for the unknown universe will limit the need for adjusting weights for respondents to those stratum where firms identify they employ energy workers but are unwilling to complete the survey or do not finish the survey and their responses are not able to be used.

---

<sup>4</sup> Industry Employment Estimate is derived from QCEW

<sup>5</sup> Industry Establishment Estimate is derived from CENSTATS



---

### Step 3: Determining incidence and employment within each stratum and produce adjusted weights

In each industry stratum, data from the survey dispositions will provide the percentage of establishments and employment within each strata for energy workers and establishments from those that were sampled. For each stratum that is sampled in the unknown universe, the ratio between total employment for the sample and the revised unknown universe should be determined. The weighting adjustment only occurs within each stratum if the sample that is used for determining the disposition of energy establishments and employments differs from the completed surveys

Each stratum's total revised unknown universe employment =  $\sum UKEN_{ILS}$

Each stratum's total unknown sample employment<sup>6</sup> =  $\sum UKen_{ILS}$

Each stratum's total unknown employment from completed surveys =  $\sum UKens_{ILS}$

Weight for each stratum within the unknown sample only occurs when;

$(\sum UKen_{ILS}) / (\sum UKEN_{ILS}) \neq (\sum UKens_{ILS}) / (\sum UKEN_{ILS})$

Weight for each stratum within the unknown sample =  $(\sum UKen_{ILS}) / (\sum UKens_{ILS})$

## B.4. Test Procedures and Form Consultations

**Describe any tests of procedures or methods to be undertaken.**

The data collection forms being submitted for approval have been field tested and received the highest rating of satisfaction of the forms being tested. Cognitive assessments will be conducted by EIA for refinement to the survey. Similar instruments have been used with success (with and without federal support) at least 30 times over the past 10 years

## B.5. Statistical Consultations

**Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s) or other person(s) who will actually collect and/or analyze the information for the agency.**

Josh Williams, President, BW Research Partnership, 760-712-6592

Philip Jordan, Vice President, BW Research Partnership, 857-919-0774

Ryan Young, Research Manager, BW Research Partnership, 831-601-5639

Mitchell Schirch, Senior Research Analyst, BW Research Partnership, 603-204-4331

David Hiles (consulted previously), Supervisory Economist, Bureau of Labor Statistics, 202-691-6567

---

<sup>6</sup> Total Unknown Sample employment is determined from survey dispositions not a complete survey