Department of Energy
Washington, DC 20585

January 7, 2016

Chad Whiteman
Department of Energy Desk Officer
Office of Information and Regulatory Affairs
Office of Management and Budget
Washington, DC 20503

SUBJECT: National Energy Related Jobs Survey

I. Purpose of this collection:

A major objective of the Secretary's Energy Jobs Strategy Council is to identify and quantify the effect that the rapidly changing nature of energy production, distribution, and consumption throughout the U.S. economy is having on job creation and economic competitiveness. Too often, these impacts are inadequately understood and, in some sectors, incompletely measured.

The proposed data collection will explore employment related to energy, including: electric power generation; electric power transmission, distribution, and storage; energy efficiency, including heating, cooling and building envelope; fuels; and motor vehicles production. The purpose is to quantify that employment, and its distribution among different sources of energy and the technologies employed. This effort will collect data necessary to extrapolate employment by primary value chain activity, including: research and development; manufacturing; sales and distribution; installation, repair and maintenance; and professional services. It will also collect data on workforce demographics and the establishment's perception on the difficulty of recruiting qualified workers.

The Bureau of Labor Statistics (BLS) collects data on establishments and jobs. Establishments are classified according to type of economic activity, or process of production, by the North American Industry Classification System (NAICS), and jobs are classified into occupations categories using the Standard Occupational Classification (SOC) system. BLS data in many cases adequately attributes employment in more traditional sectors of the energy sector, such as fossil fuel exploration, extraction, storage and transportation. However, the distribution of employment among different technologies within these sectors, such as nontraditional extraction of oil and gas, electric generation by specific fuel type (i.e., coal, gas, wind, etc.), and research and development of energy related technologies, cannot be determined.

Moreover, newer energy sources, such as solar, wind, and hydropower, and the technologies they employ, are largely included in NAICS and SOC codes for other industries and cannot be separated into their respective production processes or occupations. In 2010, the BLS was tasked by the Congress with reporting and tracking

"Green Jobs". In order to accomplish this task, BLS initiated two new data collections – the Green Goods and Services and the Green Technologies and Practices data collections. However, those collections were discontinued due to sequestration.

It is important to note that this data collection is additive, rather than duplicative, of current BLS data collections. Rather than creating new, exclusive categories of employment, the proposed data collection would ascertain specific activities and technologies within the BLS structure, so allocation of employment by activity and technology uses existing BLS data as a baseline. In this way, the proposed data collection reduces the burden on businesses because it relies on existing baseline employment totals by NAICS code that are already collected and released by BLS, therefore allowing statistical estimation procedures to be used (requiring fewer total responses) rather than replicating a census of businesses.

RESPONDENT UNIVERSE

II. Description of the reporting frame (include Eligibility or threshold levels for determining in scope respondents) What the sources for building your frame files:

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 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).

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, InfoUSA, 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.

About 10,000 in-scope "Known Universe" establishments were pre-identified as having "energy" activity. A database of likely energy-related establishments will be developed 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/InfoUSA dataset and a "known" indicator will be used to assist in oversampling "known" establishments. The Known Universe sampling is a census with up to eight (8) contact attempts by phone, mail, and email as available.

SAMPLING METHODOLOGY

III. Briefly describe the sample sizes, sampling methodology (cutoff, probability, etc). Describe the sampling stratum or classification to be applied to implement the sample design

Contractor will contact about 30,000 establishments per year. The total survey completion targets will be based on a sample selected using the QCEW/InfoUSA 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.

EJR panels will have a probability-based sample aimed at satisfying data needs at both the State x 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 24,900; complex stratification using state, industry

Known Sampling – All establishments in the Known Universe will be contacted up to eight times. The responses will be treated separately, and the overall employment from

the Known Universe sample will be unduplicated 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.
- Remove duplicate 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.

ESTIMATION PROCEDURES

IV. Describe any imputation you may do for nonresponse or if you will be applying any special estimation methods

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

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

Weighting Adjustment for Nonresponse in Annual Survey for Energy Related Jobs Report

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-related firms will be reversed matched, based on firm name and available contact information (phone number and address) against INFOUSA'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 35 to 20 percent of known firms that are not matched with INFOUSA's national database will be evaluated individually and using publically available data, 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 INFOUSA's national database or an estimate of industry classification, location information and when available employment by location1 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), location2 (State and zip code), and employment by location3 size.

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

---

[1] To remain in the known universe file, a firm needs to be reverse matched to InfoUSA'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.

[2] Location is determined to be the entire state, for smaller states, or regions within a state defined by zip code for larger states.

[3] 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.

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-related 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-related 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-related employment in the known sample and total employment in the known sample for each stratum.

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

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

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

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

Weight for each stratum within the known sample = ($\sum$ KEN ILS / $\sum$ KEN AF) / ($\sum$ Ken ILS / $\sum$ Ken AF)

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

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-related 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-related 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-related 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 – known Energy-Related 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.

Known = $\sum$ employment/establishments within a given NAICS code based on Known Universe Energy employment

Total = $\sum$ employment/establishments within a given NAICS code based on CENSTAT's County Business Patterns estimate

Revised Unknown Universe Employment by Industry of Interest = Industry Employment Estimate4 * (1 – (Known/Total))

Revised Unknown Universe Establishment count by Industry of Interest = Industry Establishment Estimate5 * (1 – (Known/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-related 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-related workers but are unwilling to complete the survey or do not finish the survey and their responses are not able to be used.

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-related 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 = $\sum$ $UKEN_{ILS}$

[4] Industry Employment Estimate is derived from QCEW
[5] Industry Establishment Estimate is derived from CENSTATS

Each stratum's total unknown sample employment[6] $= \sum$ 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 $_{ILS})$

APPENDIX

V. Copy of survey instrument as a supplemental document.

VI. Summary results at the national level from the previous collection as an example and attach that as another supplement document.

Ruth Samardick is the point of contact for questions and may be reached at 202-586-5544 or by email: ruth.samardick@hq.doe.gov.

Sincerely yours,

Natasha Campbell for David Foster

David A. Foster
Senior Advisor
Office of the Secretary

[6] Total Unknown Sample employment is determined from survey dispositions not a complete survey