

# **Big Data or Not: Determining the Use of Big Data**

ASPE Generic Information Collection Request  
OMB No. 0990-0421

## **Supporting Statement – Section B**

**Submitted:** March 21, 2018

Contracting Officer Representative

James Sorace MD MS

Medical Officer

U.S. Department of Health and Human Services

Office of the Assistant Secretary for Planning and Evaluation

200 Independence Avenue SW, Washington DC 20201

202-205-8678

James.sorace@hhs.gov

## **Section B – Data Collection Procedures**

### **1. Respondent Universe and Sampling Methods**

Critical to the success of this project will be the identification of a diverse pool of big data subject matter experts (SMEs) and their successful recruitment to participate in the study. SMEs will be drawn from industry, academia, and the government. They may include people with substantive experience analyzing big data, methodological experts, and those who have produced big data. As described in this section the contractor (NORC) will work closely with ASPE to identify and assess a broad pool of potential SMEs, taking a systematic approach to develop the pool of experts. The contractor has worked with their internal SMEs at NORC to identify a preliminary list. Additional experts have been identified through a limited literature review conducted by NORC. This literature review examined the benefits and challenges of using big data (BD) for evidence-building and reviewed different uses of BD, using search terms related to health policy topics (“population health”, “health planning”, etc.), BD and data science (“data quality”, “data science”, etc.), and data types (“health administrative data”, “Medicare enrollment”, etc.). NORC librarians and University of Chicago’s library database were used to collect BD articles, within the last two to three years, using our set of search terms. We anticipate conducting up to 25 interviews with SMEs from the industry, academia, and government. Each SME on the list of interviewees has been researched to ensure we are interviewing individuals with a variety of experience in BD.

### **2. Procedures for the Collection of Information**

With guidance from ASPE, NORC has developed an interview protocol, which focuses on the following topics regarding big data: data types, data quality, data types and algorithms, experiences with big data and decision-making, and training their organization uses for the analysis of various data types.

### **3. Methods to Maximize Response Rates Deal with Nonresponse**

We will recruit participants via email and will send follow up emails if there is no response. We do not anticipate problems with response rate or nonresponse given that we have developed a list of SMEs that is sufficient to accommodate any reasonable amount of attrition. In the event that a subject matter expert does not respond or chooses not to participate, we will seek out another individual that is representative of a similar BD background. The SME list consists of individuals that work with BD in academia, industry, and the government. For example, if an expert in BD analytics is not available NORC would find another individual with a similar background. Similarly, if an expert in the application of BD to environmental sensing becomes unavailable NORC would approach an expert with a similar background.

### **4. Test of Procedures or Methods to be Undertaken**

The subject matter expert interview protocol was developed by the study project team (Drs. James Sorace, Joshua Williams, Felicia LeClere, and Zachary Seeskin). NORC then used internal SMEs to build initial summary questions and confirms that 60 minutes will be adequate to discuss their opinions.

### **5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data**

The study project team (Drs. James Sorace, Joshua Williams, Felicia LeClere, and Zachary Seeskin) is comprised of doctoral-level researchers trained in qualitative data analyses. The study team has extensive knowledge of the subject matter. This is an information gathering exercise among experts, which will provide important insight into their view of data quality and BD. It is

not designed to be representative of national opinion nor of the full potential range of expertise as this area is rapidly evolving.