

**Report of the Big Data Working Group  
to the  
Census Scientific Advisory Committee  
March 31, 2017**

**Formation and purpose**

The Big Data Working Group (“WG”) was created on April 7, 2015. The purpose of the group was set forth in the *CSAC Working Group Guidelines*: The group will address Census policies, research and methodology, test, operations, communications/messaging, and other activities to ascertain needs and best practices to improve censuses, surveys, operations and programs. The group will provide scientific and technical expertise from the following disciplines: business and data analytics, operations research, statistics, information technology, marketing and other fields of expertise, as appropriate, to address Census Bureau program needs and objectives.

**Membership and meeting history**

The Working Group’s membership includes Ken Simonson, Chair; Noel Cressie, Jack Dangermond, Peter Glynn, Willie Jasso and Roberto Rigobon. Members met by teleconference with Census staff on June 1, July 9, September 2, and November 9, 2015; January 11, March 14, May 16, and July 26, 2016; and March 9, 2017.

**Topics covered**

Over the past two years, staff reported on several projects to investigate the feasibility of using retail transactions data from private data processing and analytics firms to augment Census retail sales data with greater timeliness and detail as to type of product, vendor or location. Ultimately, these sources may enable Census to produce data with less respondent burden, less imputation and less cost to the Bureau. However, a variety of confidentiality and methodological hurdles remain. WG members provided suggestions and cautions about spatio-temporal, classification and correlation issues.

In addition, Census staff provided updates on their participation in research meetings, conferences and exploratory talks with other government agencies, academic groups and data processing and analytics firms; and on organization, staffing and budget expectations for a Big Data Research Center within Census. WG members suggested other conferences, researchers and papers that could be useful to Census.

**Recommendations**

Now that some of the data tests mentioned above have been concluded and the research center has been established, the nature of Census staff’s use of a Big Data WG is changing. Nevertheless, it is the consensus of WG members that there is value in maintaining a consultative body such as the WG.

It is essential that Census continue to pursue ways of harnessing “big data” and to develop the in-house expertise to be a useful resource to other parts of the agency and, ideally, to other government agencies, academics and private sector users of big data. Census and other statistical offices will undoubtedly continue to use various combinations of surveys and big data to produce the next generation of statistics. Aspects such as privacy and frameworks to evaluate data quality

need to be addressed moving forward. It is conceivable that surveys (small data) will be used to correct the likely biases that arise from a big-data source.

A WG or other consultative body holding regular teleconferences or in-person meetings is useful as a way to encourage Census staff to provide a well-structured overview or update of internal and external initiatives relating to big data. Scheduling one or two calls between CSAC meetings, and holding them with a limited number of CSAC members, is an efficient way of enabling interested members to stay current and provide input regarding methodology, related research, other conferences and individuals or organizations for Census to engage with.

Although the terms of most of the current WG will expire shortly, it is worthwhile to maintain some form of big-data oversight group. We strongly encourage the Center's staff to consult with such a group, and with other individuals who have relevant expertise. A WG could help the Center and its director set some direction and priorities, and be a sounding board for the projects the director has planned. The Center and other staff should use a WG to get feedback on the use of specific data sets, methodological issues and ways of making the Center a valuable resource within Census, across government statistical agencies, and for outside researchers. Such contacts should not be limited to formal meetings of the full WG or limited to its members.