Independent Statistics & Analysis
U.S. Energy Information Administration

2018 CBECS Data Center Pilot Results

May 2021

Independent Statistics & Analysis
www.eia.gov

U.S. Department of Energy
Washington, DC 20585

This report was prepared by the U.S. Energy Information Administration (EIA), the statistical and analytical agency within the U.S. Department of Energy. By law, EIA's data, analyses, and forecasts are independent of approval by any other officer or employee of the U.S. Government. The views in this report should not be construed as representing those of the U.S. Department of Energy or other federal agencies.

2018 CBECS Data Center Pilot Results

Executive summary

- Data centers are currently not published as a separate building type for *Commercial Buildings Energy Consumption Survey* (CBECS); they are included in the *Other* building type category.
- The U.S. Energy Information Administration (EIA) cannot publish data centers as a separate building type for a few reasons, including lack of frame, small sample size with current sampling, and low cooperation rates.
- EIA conducted a pilot study of 50 data centers as part of the main 2018 CBECS data collection to assess the feasibility of publishing data center estimates in future CBECS.
- CoStar was used as the pilot frame.
- The pilot response rate was 26%, less than half that of the main sample.
- The pilot cases required two times more contacts per completed case compared with the main sample.
- Item nonresponse was very high for items that are important to data centers.
- Interviewers noted that security, finding a respondent, and the voluntary nature of CBECS were obstacles to receiving responses.
- Lack of quality frame source continues to be an obstacle to producing data center estimates.
- Data center estimates are likely not feasible with current methods. EIA would need a quality frame and cooperation from the industry in order to consider data centers as a separate building type.

Background and motivation

Data centers are one of the most energy-intensive building types, using 10–40 times more energy per square foot than the typical office building and accounting for almost 2% of the electricity use in the United States (Office of Energy Efficiency and Renewable Energy, n.d.). Data center computing capacity has rapidly grown in the past decade, but consumption has grown more slowly because of efficiency measures (Shehabi, et al., 2016). Therefore, data center buildings are of great interest to the building energy data community.

The *Commercial Buildings Energy Consumption Survey* (CBECS), the only national-level source of energy consumption data and energy-related building characteristics data for commercial buildings, includes data centers in the survey. However, because of the small number of data center buildings in the respondent dataset,[1] EIA does not publish data centers as a separate building type.[2] Instead they are grouped into the *Other* building type. This *Other* category includes buildings that don't fit into the list of CBECS principal building activity categories, such as airplane hangars and public restrooms.

[1] The 2012, CBECS had six responding data center buildings in the final respondent dataset.

[2] EIA collects data on data center space that is part of a building type such as an office or education building in CBECS and publishes this data because this space is much more common than separate data center buildings.

CBECS data users inquire frequently about data center estimates. In response to the data user community, EIA conducted a Data Center Pilot along with the 2018 CBECS buildings survey data collection. EIA aimed to gather data that would assess the feasibility of collecting data and publishing estimates for data center buildings in future CBECS. The pilot was a way to test the waters on a small number of buildings for a limited cost, instead of investing in a large data collection operation that potentially would not give useful results.

EIA considered many known or suspected obstacles to getting enough data center completed cases to publish estimates, including not having a frame of data centers; not being able to identify data centers on the current frames to sample enough of them; and high unit nonresponse rate, likely because of company privacy concerns. Although item nonresponse is corrected with imputation and does not directly prevent publishing estimates for a building type, a very high item nonresponse rate for important questionnaire items is a data quality concern. The main goals of the pilot were to test a potential frame, measure unit nonresponse, and measure item nonresponse for important questionnaire items.

Methods

Frame

To select a sample of data centers for the pilot, EIA needed a frame (a list of data centers with information needed for sampling and contacting the buildings). Currently, no known comprehensive list of data centers is accessible to EIA.[3] EIA used CoStar, a commercial real estate database to which EIA purchased a one-year license for experimental purposes, as the frame because of the database's easy file creation process and the completeness of its square footage and address information. CoStar's coverage of data centers is unknown but sufficient for a pilot study in which estimates are not needed.

To create the frame, all buildings larger than 50,000 square feet[4] identified as data centers were downloaded from CoStar. EIA and Westat, EIA's data collection contractor for the 2018 CBECS, identified which of those buildings were in the 2018 CBECS primary sampling units (PSUs) and did further file cleaning to prepare for sampling. The final frame used for the sample consisted of 115 buildings from CoStar that were identified as data centers.

As part of the preparations for sampling, EIA categorized all 115 data centers on the frame as either a colocation data center or a private data center. A colocation data center is a data center that hosts and operates multiple organizations' servers. A private data center is a data center building used by just one organization, for example, Google or Facebook. The data center type was used as a sampling stratification variable because EIA suspected that type would have an effect on both the response rate and energy use patterns.

[3] The CBECS area frame does cover data centers; however, because there are so few of them in the population and because they are difficult to identify in the area frame listing process and are more geographically dispersed than other buidings, the area frame does not provide enough data centers to sample to provide separate estimates.
[4] EIA chose a minimum size of 50,000 square feet because smaller buildings are less likely to be whole building data centers.

Sample

EIA chose a sample size of 50 data centers in considering costs and desired accuracy (20 private data centers and 30 colocation data centers). Essentially, EIA selected all private data centers from the 115 on the frame because out of all 115 data centers, 22 were private and 3 were on the same campus, so one of the buildings on that campus was selected, along with the remaining 19 private data centers. EIA grouped the colocation data centers into two building size categories, small (50,000 square feet–200,000 square feet) and large (more than 200,000 square feet). We selected the colocation data centers with proportional allocation in size and and census region strata to ensure geographic and size distribution.

Data collection

The data collection protocol for the pilot cases was the same as the main CBECS sample. The interviewers and respondents were unaware that the buildings were part of a pilot.

Areas of assessment

In evaluating the success of the pilot, we selected five criteria for assessment, calculating metrics for the first four and performing a qualitative review of the fifth:

- Unit nonresponse, compared with the main CBECS sample
- Misclassifications from frame
- Measure of effort for completed cases (number of interviewer contacts per completed case)
- Item nonresponse rates for important questionnaire items
- Qualitative assessment of interviewer notes from contact history file, looking for patterns in contacts and attempted contacts

Results

Unit nonresponse

Of the 50 sampled pilot buildings, 12 completed the survey, 3 were out of scope, and 35 did not complete the survey. These results indicate a response rate[5] of

$$\frac{Respondents}{Respondents + Nonrespondents} = \frac{12}{12 + 35} = 25.5\%$$

In comparison, the main CBECS sample response rate was 54.7%.

Only one data center identified on the frame as private responded out of the 20 private data centers sampled.

Frame misclassification

[5] The actual response rate calculation of the main CBECS sample is more complicated, and it is documented in the 2018 CBECS Final Report. For this report, EIA used a simplistic calculation.

EIA discovered that 6 out of the 50 sampled data centers were not true data centers, or were not eligible for CBECS. Two of the respondent buildings were offices, and one was an education building. Three of the sampled buildings were not eligible for CBECS: one building was under construction, one was demolished, and one was not found by the interviewer.

Interviewer effort for completed cases

The average number of interviewer contacts per completed case from the pilot was 12; in comparison, the average number of contacts per completed case from the main sample was 6.

Item nonresponse

EIA calculates item nonresponse rates[6] for select questionnaire items by using the 22 responding pilot and main sample data centers (9 confirmed as data centers from the pilot[7] and 13 from the main sample). These rates were compared with office buildings in the main sample.

Square footage and number of workers, two key items for all building types in CBECS, were compared in terms of item nonresponse. In addition, item nonresponse for servers, a key item for data centers, was compared with the main CBECS sample. Servers can be reported as a number or a category; if the respondent does not give a number, it receives a response set with ranges of numbers and an opportunity to report the range in which the number of servers belongs. We calculated item nonresponse for power usage effectiveness (PUE),[8] a key item for data centers, but we did not compare it with offices because this question was only asked of data center buildings.

Table 1 shows the item nonresponse rates for data centers and offices for these five questionnaire items. Square footage and number of workers item nonresponse was 3% lower for data centers compared with offices. However, item nonresponse for the number of servers was much higher for data centers. Of the data center building respondents, 86% answered *don't know* or *refuse* or skipped both the numeric value and categorical server question; 90% answered *don't know* or *refuse* or skipped the numeric value. The server questions were some of the only questions in the entire questionnaire that allowed a *refuse* option. Because 21% of offices and 68% of data centers selected the *refuse* option, EIA infers that number of servers is a sensitive building characteristic that respondents do not want to share, even if they know the answer.

Item nonresponse was 80% for PUE. We did not ask this question for offices, so there is no basis for comparison.

[6] Item nonresponse rate is calculated as the number of responding buildings that did not answer the question, divided by the total number of responding buildings for which the question was asked.

[7] Though 12 sampled buildings from the pilot responded, three were not actually data centers, as described in the *Frame misclassification* section.

[8] Power usage effectiveness (PUE), an energy efficiency measure for data centers, is the ratio of the total energy use for the data center divided by the energy use for the IT equipment. The ideal ratio is 1.

Table 1. Item nonresponse rates for key items

Questionnaire item	Data centers	Offices
Square footage	18%	22%
Number of workers	23%	26%
Number of servers (numeric)	90%	35%
Number of servers (numeric or categorical)	86%	33%
Power usage effectiveness	80%	N/A

Qualitative assessment of interviewer notes

EIA reviewed the interviewer notes available on the contact history paradata file for the pilot cases. In reviewing the notes, EIA determined that finding a respondent was a challenge—a respondent was identified in only 18 of the 50 cases. Interviewers faced obstacles in either finding a respondent or getting a completed interview, including

- High security: interviewers were often denied access to the building.
- Appointment passive refusal: if interviewers gained access to the building, the front desk would often say a potential respondent is available by appointment only. This approach often led to a passive refusal.
- Voluntary nature of survey: large companies with many data centers saw that the survey was voluntary and refused.
- Lack of activity at the data center: several data centers in the frame appeared to have few or no employees present.

Strategies and interviewer qualities that seemed more effective in gaining cooperation included

- Talking to security: even if security denied access to the data center, they often could provide phone numbers, email addresses, or addresses of a corporate office to the interviewer.
- Talking to corporate office: a respondent with suitable knowledge to answer the survey was not always located at the data center. Interviewers were often referred to an employee at a corporate office.
- Corporate response: some companies had multiple data centers in the sample. One corporate respondent was able to provide 3 of the 12 completed cases.
- Persistence: after identifying a respondent, some interviewers were effective at ensuring the respondent completed the interview.

Conclusion and recommendations

The 2018 CBECS Data Center Pilot confirmed that the suspected obstacles to data center building estimates are, in fact, present:

- Lack of a quality, comprehensive frame
- Low cooperation rates for data center buildings sampled
- High item nonresponse rate for important questionnaire items

EIA used CoStar as the pilot frame; however, CoStar's coverage of the population of data center buildings is unknown. EIA could look into other potential frame sources, such as Data Center Map and Wired Real Estate Group, for potential use. However, without external verification of their completeness and potential bias, the quality of those resources is unknown. EIA would need to use a web scraping operation to pull a file of them for sampling, which could potentially present complications with website terms of use and conformity of available information into a standard file.

The cost and burden for completed data center building interviews using current methods is high. Pilot buildings responded at less than half the rate of buildings in the main CBECS sample, and they required two times as many interviewer contacts as the main sample. Difficulty finding respondents, high security, the voluntary nature of the survey, and the lack of employees at some buildings were common barriers to getting completed interviews.

For questionnaire items that are important to data centers, the rate of nonresponse was high. More than 80% of data center buildings did not answer the questions about the number of servers and the PUE, indicating a sensitivity to these questions.

If getting data center building estimates is of high importance to future CBECS, EIA would need more cooperation from industry to get access to a frame (if one exists) and to help encourage buildings to respond. A different contact strategy may be required for these buildings. Cognitive research and pre-tests could be helpful in identifying response barriers and ways to get around them.

Although separate data center building estimates will not be available from the 2018 CBECS, the main questionnaire included questions about small data center spaces within other buildings, such as offices, and these data will be published similar to how they were in the 2012 CBECS (Lewis, 2016). Small room data centers receive much less attention than large data centers, but they account for a significant fraction of the total number of servers in the United States and consume 13 billion kilowatthours of energy each year (Ganeshalingham, Shehabi, & Desroches, 2017).

References

Ganeshalingham, M., Shehabi, A., & Desroches, L.-B. (2017, June). *Shining a Light on Small Data Centers in the U.S.* Retrieved from Lawrence Berkeley National Laboratory: https://eta.lbl.gov/sites/default/files/publications/lbnl-2001025.pdf

Lewis, K. (2016). *Today in Energy*. Retrieved from eia.gov: https://www.eia.gov/todayinenergy/detail.php?id=28232

Office of Energy Efficiency and Renewable Energy. (n.d.). Retrieved from energy.gov: https://www.energy.gov/eere/buildings/data-centers-and-servers

Shehabi, A., Smith, S. J., Sartor, D. A., Brown, R. E., Herrlin, M., Koomey, J. G., . . . Lintner, W. (2016). *United States Data Center Energy Usage Report.* Lawrence Berkeley National Lab.