Survey Respondents' Perceptions on Data Confidentiality and Data Sharing

By
Jacob Bournazian, U.S. Energy Information Administration

Abstract. The U.S. Energy Information Administration **(**EIA) is a federal statistical agency that collects energy information exclusively for statistical use. Protecting the confidentiality of energy data that EIA collects is vital for EIA to establish and maintain its trust relationship with survey respondents. Private companies that provide information to the government have a direct interest in protecting their investments and resources that generate revenue. For each survey, EIA seeks to balance its statutory obligations to gather and share energy information with other government agencies and Congress while maintaining a trust relationship with its survey respondents. This paper presents a quantitative assessment on the perceptions of survey respondents regarding a federal statistical agency's ability to protect the confidentiality of information it collects. This study was conducted during 2015 after the largest federal data breach in U.S. history occurred at the Office of Personnel Management.

Keywords: Data confidentiality, data security, data sharing.

Throughout the past two decades, survey respondents continue to express a growing concern about the privacy protections that apply to their data and the ability of the data collector to preserve the confidentiality of their responses.[1, 2, 3]

The issue of the public's trust in the government's ability to protect the confidentiality of their information plays an important role in preserving the government's ability to collect accurate information.[4] Recent studies have shown that trust in government has declined in the past 20 years.[5,6] One study showed that only 24% of the participants "trust the federal government in Washington to do what is right just about always or most of the time," while 75% responded that they trust the federal government only "some of the time" or "never."[7] A study in 2014 showed that the majority of adults feel that their privacy is being undermined in core areas such as data security and preserving confidentiality.[8] There have been previous studies that focused on business perceptions of the government's ability to protect information. One study showed that 55% of survey participants were either not too confident or not confident at all that their records would remain safe and private.[9] During the past twenty years, evidence shows that the public's trust in the federal government is declining.[10] When asked "How much do you trust the government in Washington to do what is right?" the percentage who responded "Almost Never" in 1995 was 19%.[11] That percentage rose to 21% in 2000 and increased to 34% in 2010.[12]

The study presented in this paper was conducted during June-October, 2015 after a major data breach at the Office of Personnel and Management (OPM) occurred. In June 2015, OPM discovered that the background investigation records of current, former, and prospective federal employees and contractors had been stolen**.**[13] OPM and the interagency incident response team concluded that sensitive information, including the Social Security Numbers of 21.5 million individuals, was stolen from the background investigation databases. This included 19.7 million individuals that applied for a background investigation, and 1.8 million non-applicants, primarily spouses or co-habitants of applicants.[14]

Prior to the 2015 data breach at OPM, 33% of the public already thought that the federal government was not competent, not honest, and would not keep their information

confidential.[15] The decline in public trust may impact the work of federal statistical agencies, especially if an agency does not have mandatory data collection authority.[16] A potential outcome is that if a respondent doubts that the government will keep their data confidential, they may report less accurate data or not report at all. [17,18] Prior research shows that the key factor affecting the respondent's level of trust is the confidence that the respondent has in the integrity of the data-collection agency rather than the substance of the words used in the pledge provided to survey respondents.[19]

1. Background

EIA collects energy information exclusively for statistical use. Protecting the confidentiality of energy data reported from private companies is vital for EIA to establish and maintain its trust relationship with survey respondents. Private companies that provide information to the federal government have a direct interest in protecting their investments and resources that generate revenue.

EIA is a unique principal federal statistical agency within the federal statistical community for two reasons. First, it has mandatory data collection authority to collect any information from any person engaged in any phase of energy supply or major energy consumption in order to assess energy supply conditions in the United States.[20] Second, EIA operates under statutory authority that obligates the agency to share energy data with other federal agencies and other offices within the Department of Energy.[21, 22] The Energy Independence and Security Act extended the obligation to share information with State governments by requiring EIA to share company level data with State governments under reasonable terms and conditions.[23] For each survey, EIA seeks to balance the competing objectives to provide accurate and independent objective energy data to the public, other government agencies, and Congress while maintaining a trust relationship with its survey respondents.

2. Methodology

Researchers conducted structured interviews with 57 participants from five different energy industry groups: Coal (10), Electricity (10), Natural Gas (12), Solar (6), and Petroleum (19). Within these groups, the Petroleum group consisted of four subgroups: Pipelines (4), Refiners (4), Terminals (5), and Wholesalers (6). Potential participants were selected to participate from lists of survey respondents in the frame files for the 11 surveys shown below.

1. Coal: Form EIA-7A, "Annual Survey of Coal Production and Preparation."
2. Electric Power: Forms EIA-860, "Annual Electric Generator Report" and EIA-923, "Power Plant Operations Report."
3. Natural Gas: Forms EIA-857, "Monthly Report of Natural Gas Purchases and Deliveries to Consumers" and EIA-176 "Annual Report of Natural and Supplemental Gas Supply and Disposition."
4. Petroleum: Forms EIA-810 "Monthly Refinery Report." EIA-812 "Monthly Product Pipeline Report," EIA-813 "Monthly Crude Oil Report," EIA-815 "Monthly Bulk Terminal and Blender Report," and EIA-863 "Petroleum Product Sales Identification Survey."
5. Renewable Fuels: Form EIA-63B, "Annual Photovoltaic Cell/Module Shipments Report."

These surveys were selected based on the common factors that they pledge to protect the confidentiality of the survey information and apply data protection methods to statistical

aggregate data that EIA releases to the public. One survey[24] also protects some of the information collected under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA).[25]

For each EIA respondent company, the person responsible for completing the relevant survey form was identified as a potential study participant. The potential participants were sent an introductory email explaining the purpose of the study. Participation in this study was voluntary. A total of 243 companies were contacted, resulting in 57 interviews, or a "willingness to participate" rate of 23.4%.

3. Procedure

Participants completed a structured telephone interview which averaged under 20 minutes. The interviewer explained to participants that their participation was voluntary, and that data would be kept anonymous and confidential. The interview consisted of 15 main open-ended questions with 18 follow-up questions that the interviewer asked when relevant. Out of 33 total questions, this paper highlights findings from 19 of the 33 questions which had the highest level of responses.

Two separate raters read each response and applied a coding scheme to capture the open-ended data categorically. For example, responses about a participant's trust in the government indicated either "Trust," "Distrust," "No Opinion," or "Declining Trust." Each rater independently coded participants' answers and then resolved any differences in coding. The frequency of each response was then calculated across all 57 participants as well as across all energy industry groups.

4. Results

Participants described confidentiality in many ways, with two main themes emerging from their definitions. Specifically, 53% of the participants interpreted the word "confidential" as meaning that the government will not release the information to the public or share it with their competitors, while 37% had a more conservative interpretation and felt that confidentiality means keeping the data only between the respondent and EIA. The two most common responses show that participants commonly think of confidentiality in two main ways:

1) Some participants endorse a strict definition of confidentiality—information is shared only between two parties for EIA's use, with no qualifications; and
2) Others see confidentiality as limiting who EIA can share the information with—data sharing is limited, and information may be further protected by safeguards or principles of trust and privacy.

When asked why they wanted to keep their survey information confidential, 77% of the participants cited a need to remain competitive as the most common reason they wanted EIA to protect their data. Among the other reasons that participants provided, 12% wanted to protect the privacy of their employees or customers; 5% felt that keeping information confidential is a basic principle of a trust relationship with EIA; and 7% felt that it was a legal obligation that was binding on the companies, such as obligations in written contracts not to disclose information to third parties.

Figure 1 below shows the participants responses when asked "In general, how much do you trust the federal government when it comes to the privacy and confidentiality of your information?"

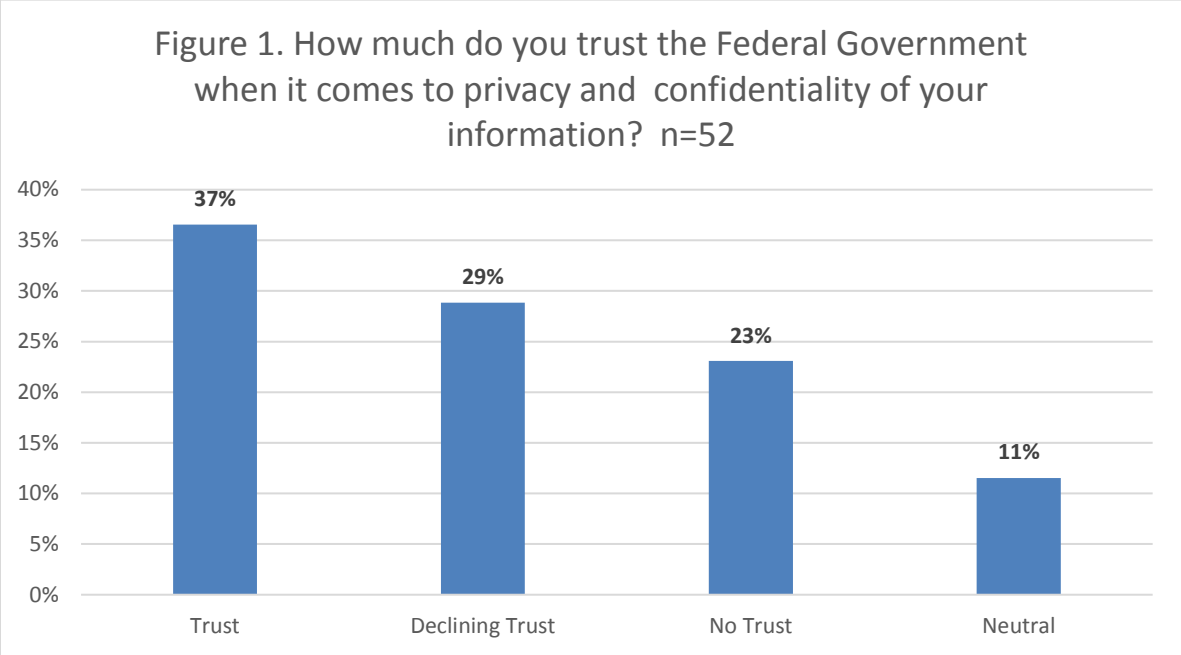

Figure 1 shows that only 37% of the participants stated that they trusted the government to protect their information. 29% of the participants stated that their trust was declining or that they were unsure of trusting the government. Another 23% stated that they had little or no trust in the federal government's ability to protect information with the remaining 11% stating they were neutral or had no opinion. Differences in perception existed across energy industry groups. Most notably, a majority (75% of the participants) in the Natural Gas group said that they trusted the federal government as a whole, while participants in other groups reported mixed levels of trust.

The combined 52% (29% that have declining trust and 23% that have no trust) shows the low business perceptions of trust in the federal government to protect the confidentiality of their information. The 52% that had declining trust or no trust may also show the possible impact of the breach at OPM on the energy industry's trust in government.

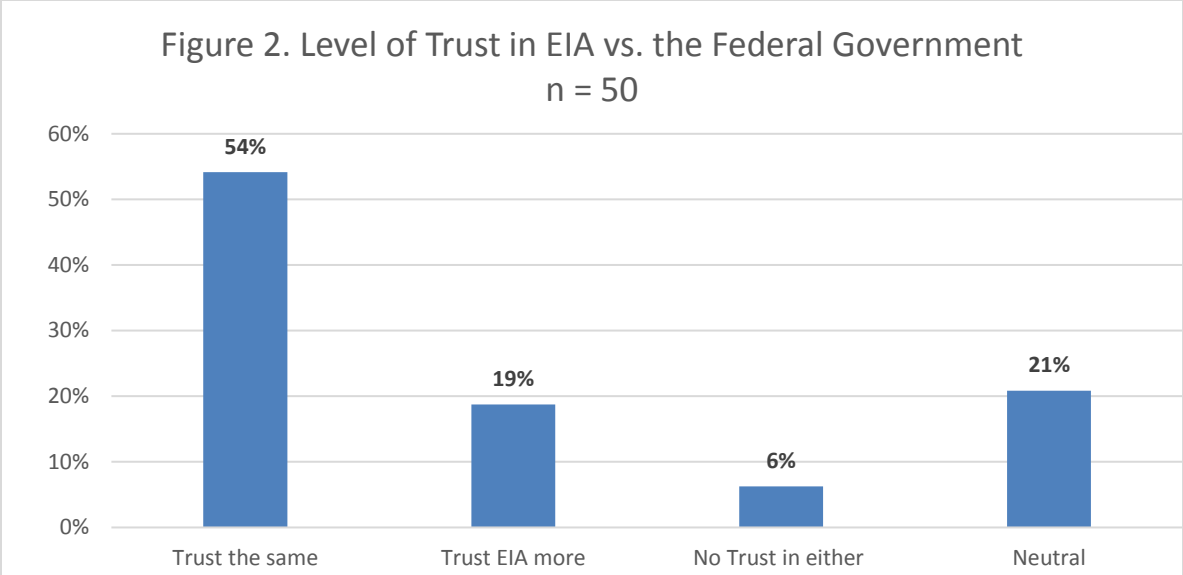

Figure 2 above shows the participants' responses when asked "In comparison to the federal government, how much do you trust EIA when it comes to protecting your companies' survey responses?" Approximately half the participants (54%) stated that they trusted EIA as much as they trusted the federal government. 19% trusted EIA more than the federal government. No participants reported that their trust in EIA was declining or that their trust in both EIA and the federal government was declining. No participant reported that they had less trust in EIA than they had in the federal government.

For the 19% that trusted EIA more than the federal government, several participants mentioned that EIA applied data security procedures and noted that EIA had no record of computer system breaches by data intruders. These results indicate the benefit from a statistical agency's efforts to build and maintain a trust relationship with respondents by applying strong data security protections and limiting its use of survey data to statistical purposes. No participant mentioned data sharing as a criterion that affects their perception of confidentiality. The 21% that had no opinion or were neutral shows that there is a significant percentage of survey respondents that may form a positive opinion if they knew more about the data safeguards that EIA applies to protect their information. EIA may be able to strengthen its trust relationship with this category of participants by communicating more frequently and providing more information on its data security procedures.

When asked "Are you satisfied with the data security that EIA applies to protect your data from unauthorized access by data intruders and hackers?" 59% of the participants stated that they were satisfied with the steps that EIA takes to protect data from hackers. 33% were not satisfied or did not know, and 8% said their confidence was declining. The fact that some participants had declining trust when discussing data security measures suggests that applying appropriate data safeguards is a component of the trust relationship between survey respondents and EIA that the previous questions did not completely measure. The fact that 33% of the participants were not satisfied or did not know shows that EIA could strengthen its trust relationship by increasing communication on this topic with survey respondents.

Approximately 80% of the survey information that EIA collects is protected to the extent that it satisfies the criteria for an exemption under the Freedom of Information Act

(FOIA).[26] 72% of the participants stated they were familiar with FOIA and the remaining 28% of the participants responded that they were not familiar with FOIA or did not know what it meant. Contrastingly, the majority of the participants had not heard of CIPSEA, although EIA uses CIPSEA to protect approximately 20% of its survey data. Only one survey, partially protected under CIPSEA was included as part of this study. Of the 72% that were familiar with FOIA, each participant was asked whether they were confident that their data would be protected by applying the exemptions under FOIA. Of those familiar with FOIA, 61% indicated that they had confidence that the FOIA exemptions could protect their information while 39% indicated that they did not have confidence that their data would be protected by applying FOIA exemptions.

When asked, "Are you familiar with the term "disclosure limitation methodologies?" 79% of participants stated that they were not familiar with this term. Furthermore, of the 21% that stated that they were familiar with this term, only 5% correctly understood that the published data are checked to make sure that the statistics do not identify specific companies. This shows that participants do not understand the technical jargon used by statisticians in the federal statistical community. The use of plain English to explain data safeguards may improve the comprehension of the subject on the risks of re-identification by the general public. The public will not pay attention to statements made by an agency if they do not understand the words that are used in the communications.

When participants were asked if they were aware that EIA has a legal obligation to share energy data with other departments in the federal government?" 33% of the participants stated that they were aware and 67% of the participants stated that they did not know that EIA shared data with other departments in the federal government. EIA dedicates half of the text on its data confidentiality provisions in its survey instructions to explain the obligation to share company level data with other federal agencies for any non-statistical purposes such as administrative, regulatory, law enforcement, or adjudicatory purposes. The fact that the majority of the participants were not aware of EIA's obligation to share energy data with other federal agencies indicates that placing that type of information in the survey instructions is not an effective way of communicating that activity to survey respondents. This outcome was consistent across energy industry groups, except for the Natural Gas group where 42% of the participants stated they were aware of this data sharing activity.

The participants were probed on whether they have concerns that EIA shares company level data with other departments in the federal government. 74% of the participants stated that they did not have any concerns with EIA sharing company level data with other federal agencies while the remaining 26% of the participants stated that they had concerns with this data sharing activity. The most common concerns were that the information reported to EIA is for statistical purposes and other departments would use it for various non-statistical purposes. Participants also had less trust that the non-statistical use of their data by other agencies was for a proper purpose. Of the same 26% that had concerns, some participants also felt that they should be notified when EIA provides data to other departments. Other participants felt that EIA should only provide aggregate data to other federal departments and not provide identifiable company-level data. One reason for their concern was the perception that EIA does a better job of protecting their data than other agencies. In fact, only about half of the participants (56%) believed that other federal agencies have the same ability as EIA to protect their survey data.

When asked "Do you have any concerns with EIA sharing your data with researchers at universities?" 39% of the participants expressed concerns with this type of activity. The most common concern was that researchers could misuse the data to support a private agenda that may be adverse to the industry. Other concerns related to the terms and conditions of the oversight and whether the researchers were subject to penalties. The remaining 61% of the participants did not have concerns about this data sharing activity. One-third of that 61% conditioned their response on the data being de-identified and appropriate data safeguards being in place to prevent misuse.

5. Conclusion

The objective of this study was to assess the perceptions of survey respondents regarding EIA's ability to protect the confidentiality of the information that they report to EIA. Information was gathered from a representative group of survey participants across all energy industry groups. Results from this study may assist a principal federal statistical agency to identify practices that strengthen their trust relationship with survey respondents relating to data confidentiality protection policies and data sharing activities.

The findings from this study show that the participants have more trust in EIA than other departments in the federal government to protect the confidentiality of their information. The data breach that occurred at the Office of Personnel Management during 2015 negatively affected the public's trust in the federal government's ability to protect their information, but it did not affect EIA survey respondent's perception of EIA's ability to protect their information.

Responses indicated that EIA could strengthen its trust relationship with survey respondents if it communicated and informed respondents about the data security procedures it applies to protect survey data. Respondents do not understand technical statistical terminology. EIA, as well as the rest of the federal statistical community, should explore alternative wording using plain English to describe data protection methodologies so that the concepts are understandable to non-statisticians and the general public. One recommendation is to replace the term "statistical disclosure limitation methodologies" with the term "data protection methods."

EIA applies the exemptions under the Freedom of Information Act (FOIA) and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) to protect the survey information it collects. Although a majority of the participants had heard of FOIA, either through EIA survey forms or other sources, the responses indicate that EIA should communicate more with survey respondents and explain the laws it uses and data safeguards it applies to protect survey information. Providing this type of information annually or semi-annually would strengthen EIA's trust relationship with survey respondents.

The majority of the participants were not aware about EIA's data sharing activities. Currently, EIA shares company level data with other federal agencies and colleges and universities. EIA grants restricted access to researchers only for statistical purposes and carefully reviews the resulting papers for potential statistical disclosures prior to any public release. However, study participants had more concerns with EIA sharing company level data with academic researchers than they did with EIA sharing data with other federal agencies.

An important finding from this study is that EIA needs to communicate more information and more frequently with survey respondents regarding the data security procedures that it applies to safeguard survey information. A strong trust relationship with respondents provides a protective shield against negative perceptions whenever a breach may occur in another part of the federal government. By increasing the respondents' understanding of the agency's data security procedures, it can build and maintain a strong productive trust relationship with its survey respondents.

---

[1] Westin, A., Social and Political Dimensions of Privacy. Journal of Social Issues, Vol. 59(2), p. 431-453, (2003).
[2] Madden, M., Public Perceptions of Privacy and Security in the Post-Snowden Era. Pew Research Center, Washington, DC. Available: http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/ (November 12, 2014).
[3] Singer E., Bates, N., Van Hoewyk, J., Concerns about Privacy, Trust in Government, and Willingness to Use Administrative Records to Improve the Decennial Census. AAPOR (2011).
[4] Directive No. 1, Office of Management and Budget, Vol 79 FRN p. 71610, December 2, 2014.
[5] House, C., More Transparency? It's Not Clear! Presentation at the Federal Committee on Statistical Methodology's December Policy Conference. Washington, DC. (2014).
[6] Singer E., Bates, N., Van Hoewyk, J. (2011). Concerns about Privacy, Trust in Government, and Willingness to Use Administrative Records to Improve the Decennial Census. AAPOR 2011.
[7] Pew Research Center. (2014). Public Trust in Government: 1958-2014. Section 2 Views of the Nation, the Constitution and Government, Views of the Government. Washington, DC. Available: http://www.people-press.org/2014/11/13/public-trust-in-government (June 26, 2014).
[8] Madden, M., Public Perceptions of Privacy and Security in the Post-Snowden Era. Pew Research Center, Washington, DC. (2014). Available: http://www.pewinternet.org/2014/11/12/public-privacy-perceptions
[9] Rainie, L. (2015). Networked Privacy in the Age of Surveillance, Sousveillance, Coveillance. Harvard University - "Privacy in a Networked World" Workshop. Cambridge, MA.
[10] Singer E., Bates, N., Van Hoewyk, J., Concerns about Privacy, Trust in Government, and Willingness to Use Administrative Records to Improve the Decennial Census. AAPOR 2011, p. 5681 (2011).
[11] .Id.
[12] Id.
[13] Cybersecurity Incident Reports, Cybersecurity Resource Center, Office of Personnel Management. https://www.opm.gov/cybersecurity/cybersecurity-incidents/ Accessed August 15, 2016.
[14] Id.
[15] Nardone, T, COPAFS Remarks. Council of Professional Associations on Federal Statistics Quarterly meeting, Washington, DC, December 2013.
[16] Prewitt, K., Why It Matters to Distinguish Between Privacy & Confidentiality. Journal of Privacy and Confidentiality Vol. 3, No. 2 (2011) p. 41 at p. 45.
[17] Childs, J.H. and Smirnova, M., Using Cognitive Interviewing to Detect Privacy and Confidentiality Concerns. JSM 2012 Proceedings of the Survey Research Methods Section, ASA, p. 3599 at p. 3604. (2012).
[18] Prewitt, K, Why It Matters to Distinguish Between Privacy & Confidentiality. Journal of Privacy and Confidentiality p. 46.
[19] Singer E., Mathiowetz N., Couper M., The Impact of Privacy and Confidentiality Concerns on Survey Participation: The Case of the 1990 U.S. Census. The Public Opinion Quarterly, Vol. 57. No. 4 p. 479. (Winter, 1993).
[20] 15 U.S.C. § 772(b) and 15 U.S.C. § 779a.
[21] 15 U.S.C. § 771(f).
[22] 42 U.S.C. § 7135(f).
[23] 42 U.S.C. § 17284 (c) (1) (B).

[24] Form EIA-863, the name and address information reported in Part 1 item 9-18, Parts 2, Preparer Information, and Part 3, Total Sales Volumes by State are protected under CIPSEA. This survey is currently suspended.
[25] Enacted in December, 2002 as Title V of the E-Government Act of 2002 (Pub. L. 107–347, 116 Stat. 2899, 44 U.S.C. § 101).
[26] 5 U.S.C. § 552.