Balancing Statistical Principles—the U.S. System

For

Energy Data in Canada Workshop

HEC Montreal

September 28, 2021 | Montreal, Quebec (virtual)

By

Tom Leckey, Assistant Administrator for Energy Statistics

Overview

- Structure of the U.S. Statistical System
- Energy Information Administration, Statutory Authority
- Regulation through the Office of Management and Budget
- Data Suppression Rules and Examples
- Data Access and Users

The U.S. Statistical System is operationally decentralized

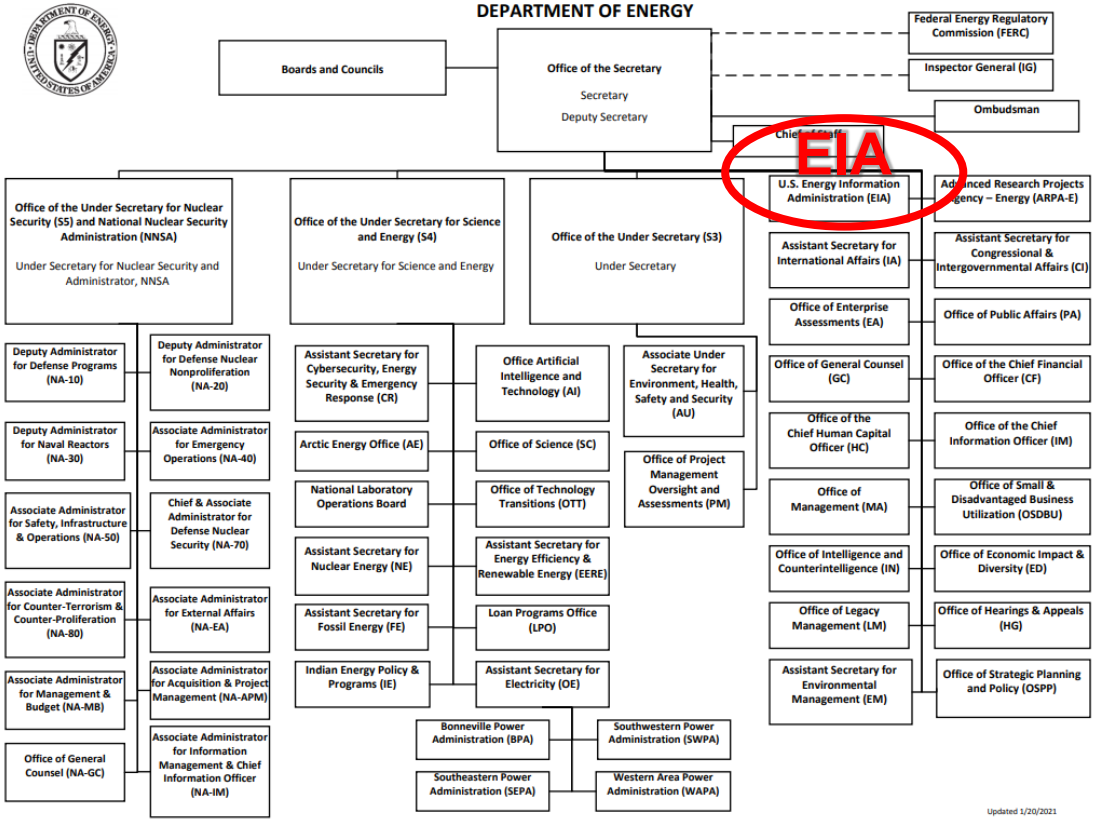
- Three Branches of Government: Executive, Legislative, and Judiciary
- Executive Branch – 15 Departments
- 190 “statistical agencies” within 15 Departments
- But only 90 of these “agencies” conduct statistical collections
- And only 14 of those 90 are *Principal Federal Statistical Agencies*
- EIA is among the 14 principal agencies

Common characteristics of Principal Federal Statistical Agencies:

- Produce objective data that are relevant to policy issues
- Achieve and maintain credibility among data users
- Achieve and maintain trust among data providers
- Achieve and maintain a strong position of independence from the appearance and reality of political influence and control
- Continual improvement and innovation

EIA is the statistical and analytical agency within the U.S. Department of Energy

- EIA's information is independent: the EIA Administrator is a direct report to the U.S. Secretary of Energy, but EIA's data and reports are released at the discretion of the Administrator.
- EIA's role is unique: by providing an unbiased view of energy markets, EIA increases transparency and promotes public understanding of important energy issues.



EIA's Statistics programs rely on mandatory, but negotiated, collection authority

Legal rights to collect

- Federal Energy Administration Act of 1974 (Public Law 93-275)
- Department of Energy (DOE) Organization Act of 1977 (Public Law 95-91)
- Other legal mandates

Legal obligations to protect

- Confidential Information Protection and Statistical Efficiency Act (CIPSEA), Title V of the E-Government Act of 2002 (Public Law 107-347)
- Freedom of Information Act, 5 USC. 552, exemptions 3, 4, and 6
- Paperwork Reduction Act, 44 U.S.C. 3501
- Information Quality Act, P.L. No. 106-554; H.R. 5658, Section 515(a)

Trust

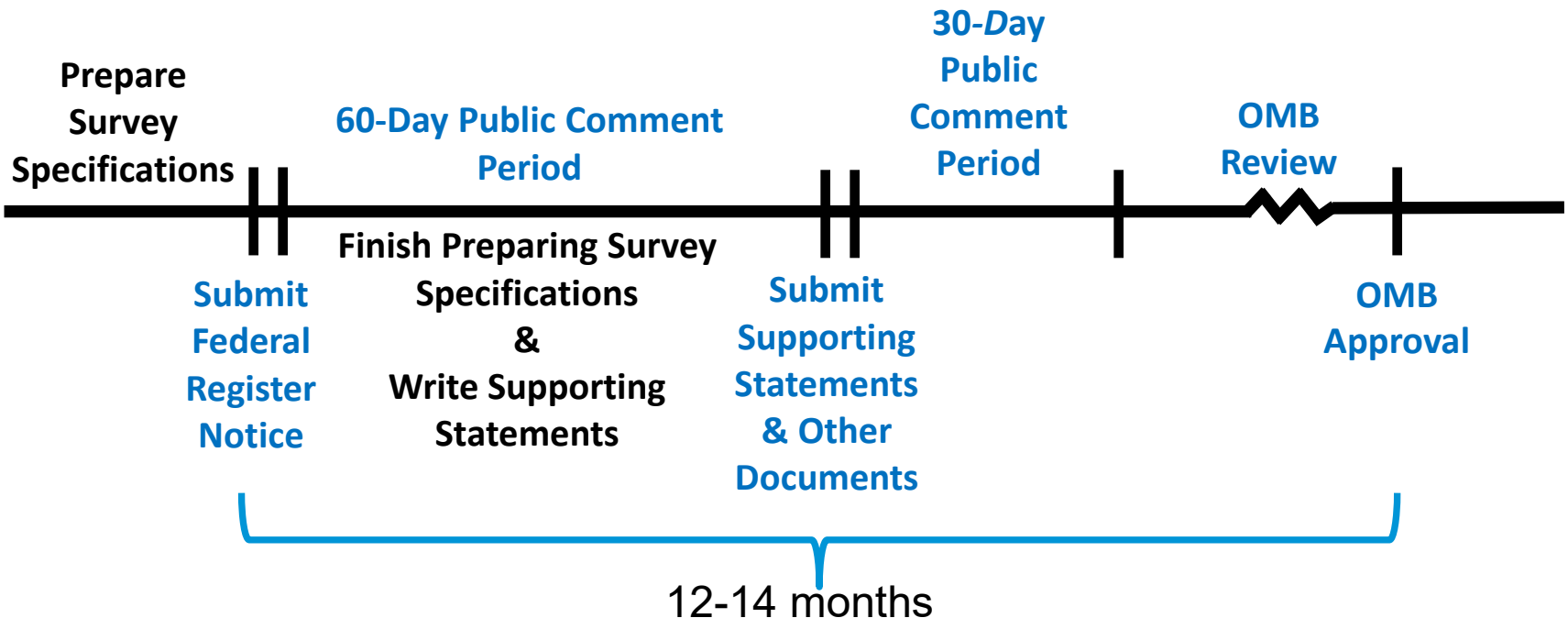
Amicable relationship with data suppliers

Maximizing public access to information while maintaining confidentiality

Integrity and transparency

- Transparency with data users about data sources and survey methods
- Transparency with data suppliers on use and purpose for collecting the information and how the data will be protected

The OMB clearance process – normal timeline



Need for Disclosure Protection Methodology at EIA

- Title 15 U.S.C. authorizes mandatory data collection by EIA
- EIA informs respondents to our surveys when disclosure limitation procedures (typically, cell suppression) are applied to statistical aggregates:
  - These promises are important to keep in order to maintain the trust of EIA's survey respondents and the high quality of our data products
  - All data items collected from 12 surveys protected under the Confidential Information Protection and Statistical Efficiency Act of 2018 (CIPSEA)
    - EIA, and outside data users, must use for statistical purposes
    - CIPSEA does not allow for time limits on disclosure protection
  - Select data items from almost 40 surveys protected under an exemption to the Freedom of Information Act (FOIA)—"Competitive harm"
    - Data may be used for nonstatistical purposes
    - FOIA exemption allows for expiration of time limits on disclosure protection

Primary and Complementary Cell Suppressions

- Prior to publishing data tables at EIA, we suppress cells of data in a given published table to prevent disclosures of identifiable and sensitive data

- Sensitive cells are typically identified using the P% Rule, where the value of P depends on the survey and is confidential, but other rules may be used instead of, or in combination with, the P% Rule
  - If we denote the total value for a given cell by T and the values for the top two contributors to the cell by C1 (largest) and C2 (second largest) then a cell is considered sensitive and is a *primary suppression* if the remainder, T – C1 – C2, is less than P% of C1
  - We do not want the second largest contributor to be able to derive the largest contributor's value within less than P% of C1, which is the amount of required protection

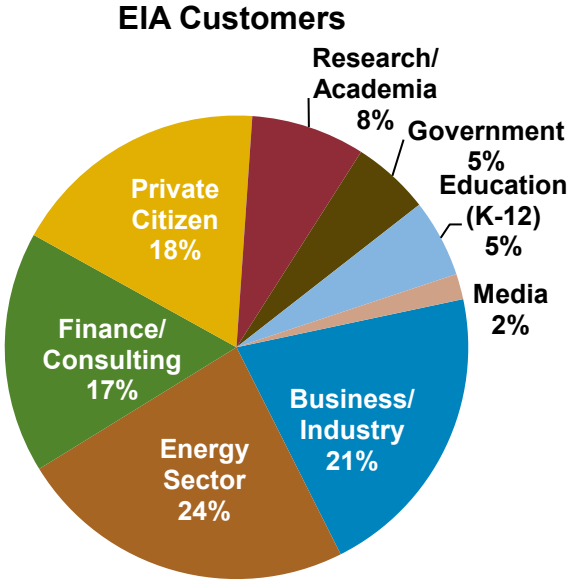
- Additional cells called *complementary suppressions* may need to be suppressed to protect the primary suppressions

Example of Primary Suppression

Monthly and weekly U.S. and regional stocks of total gasoline[1] by type of facility (thousand barrels)

East Coast (PADD 1)	2-Jul-21	9-Jul-21	16-Jul-21	23-Jul-21	30-Jul-21	6-Aug-21	13-Aug-21	20-Aug-21	27-Aug-21
Total	69,031	68,846	67,886	64,704	61,197	61,610	58,696	56,381	57,662
Bulk terminal[3]	49,488	50,088	49,834	46,983	44,389	44,756	42,466	40,804	40,732
Pipeline	16,402	15,827	15,231	15,106	14,275	14,277	13,782	12,842	14,340
Refinery	3,141	2,931	2,821	2,615	2,533	2,577	2,449	2,735	2,589
New England (PADD 1A)	**2-Jul-21**	**9-Jul-21**	**16-Jul-21**	**23-Jul-21**	**30-Jul-21**	**6-Aug-21**	**13-Aug-21**	**20-Aug-21**	**27-Aug-21**
Total	5,183	4,881	4,669	4,600	4,924	4,888	4,464	3,802	3,351
Bulk terminal	W	W	W	W	W	W	W	W	W
Pipeline	W	W	W	W	W	W	W	W	W
Refinery	-	-	-	-	-	-	-	-	-
Middle Atlantic (PADD 1B)	**2-Jul-21**	**9-Jul-21**	**16-Jul-21**	**23-Jul-21**	**30-Jul-21**	**6-Aug-21**	**13-Aug-21**	**20-Aug-21**	**27-Aug-21**
Total	35,828	36,400	35,944	32,442	29,997	28,897	27,607	26,942	26,783
Bulk terminal[3]	27,187	27,676	27,639	25,135	22,445	21,986	20,829	20,069	19,346
Pipeline	W	W	W	W	W	W	W	W	W
Refinery	W	W	W	W	W	W	W	W	W
South Atlantic (PADD 1C)	**2-Jul-21**	**9-Jul-21**	**16-Jul-21**	**23-Jul-21**	**30-Jul-21**	**6-Aug-21**	**13-Aug-21**	**20-Aug-21**	**27-Aug-21**
Total	28,020	27,565	27,272	27,662	26,276	27,825	26,626	25,637	27,527
Bulk terminal	W	W	W	W	W	W	W	W	W
Pipeline	W	W	W	W	W	W	W	W	W
Refinery	W	W	W	W	W	W	W	W	W

EIA's information is used by a range of stakeholders



Source: 2020 Customer Satisfaction Survey

Customer-focused Performance Results

- **Quality**: 90% of customers are satisfied or very satisfied with the quality of EIA information
- **Timeliness**: 95% of select recurring products meet their release date target

Examples of Activities

Government
- Executive Agencies – WH, DOE, & EPA use EIA data to track energy markets and program performance and to analyze policy proposals
- Congress – policy development and agency funding
- State Governments – planning and program development

Energy Sector
- Consumers – monitor actual and projected prices
- Producers – track inventory statistics

Business/Industry
- Manufacturers – market research

Finance/Consulting
- Commodities Analysts – market response to supply data

Media/Research/Education
- Journalists – cite energy statistics
- Teachers – use *Energy Kids* materials
- Researchers – energy forecasting and modeling

Private Citizens
- Public – *Energy Explained* provides a primer on energy topics

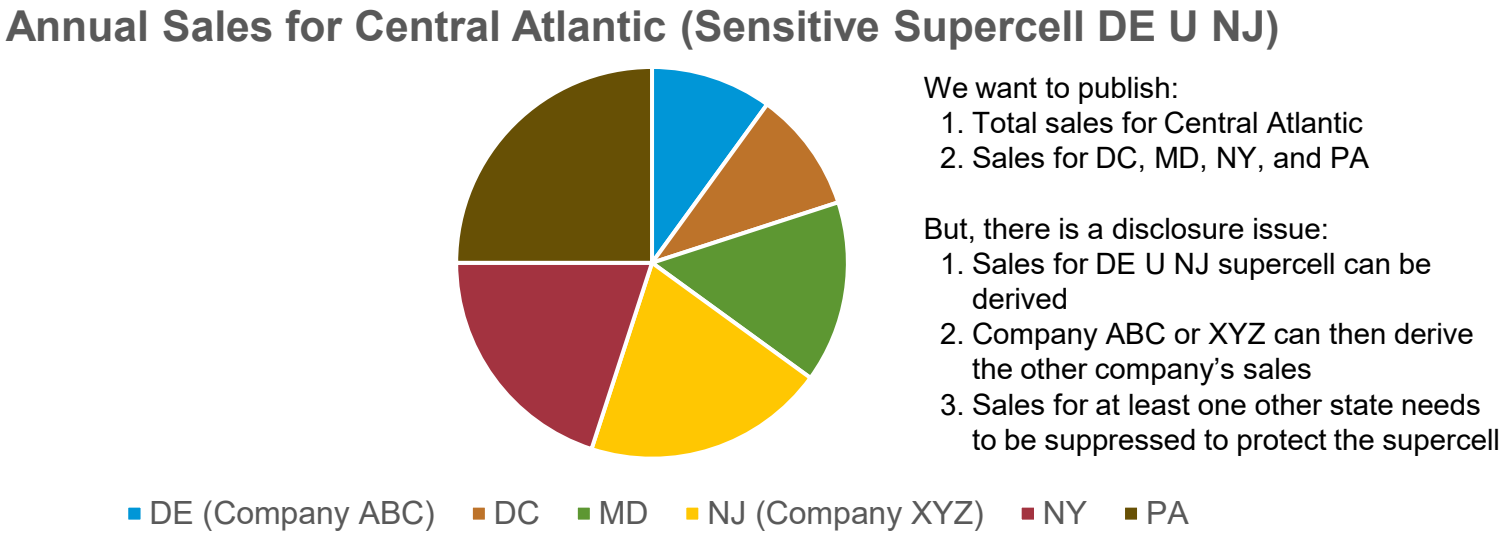
Suppression for Quality Reasons

Table B16. Employment size category, floorspace, 2012

	Total floorspace (million square feet)							
		Number of workers						
	All buildings	Fewer than 5 workers	5 to 9 workers	10 to 19 workers	20 to 49 workers	50 to 99 workers	100 to 249 workers	250 or more workers
All buildings	87,093	17,751	8,973	9,623	14,514	13,476	10,941	11,815
Building floorspace (square feet)								
1,001 to 5,000	8,041	5,362	1,768	794	117	N	N	N
5,001 to 10,000	8,900	3,843	2,389	1,793	869	Q	N	N
10,001 to 25,000	14,105	4,334	2,454	3,135	3,314	719	Q	Q
25,001 to 50,000	11,917	1,791	1,437	2,017	3,615	2,339	686	Q
50,001 to 100,000	13,918	1,332	447	1,051	4,036	4,388	2,328	337
100,001 to 200,000	12,415	677	Q	574	1,903	4,152	2,861	1,879
200,001 to 500,000	10,724	Q	Q	Q	595	1,544	4,078	3,842
Over 500,000	7,074	Q	Q	Q	Q	Q	845	5,718
	RSEs for total floorspace							
	All buildings	Fewer than 5 workers	5 to 9 workers	10 to 19 workers	20 to 49 workers	50 to 99 workers	100 to 249 workers	250 or more workers
All buildings	4.7	6.6	6.9	8.7	6.2	6.5	8.5	6.8
Building floorspace (square feet)								
1,001 to 5,000	5.2	6.1	7.3	9.1	24.5	0.0	0.0	0.0
5,001 to 10,000	8.2	8.1	11.8	13.5	13.8	70.2	0.0	0.0
10,001 to 25,000	6.9	9.9	9.9	11.3	10.8	18.1	44.4	100.0
25,001 to 50,000	8.0	22.2	13.2	14.2	10.1	13.9	18.7	96.5
50,001 to 100,000	7.7	22.4	25.1	18.6	10.8	10.1	14.5	18.6
100,001 to 200,000	7.2	28.6	26.6	22.6	12.6	11.5	15.1	18.2
200,001 to 500,000	6.5	33.6	43.1	32.0	23.7	14.8	11.7	7.9
Over 500,000	9.6	45.5	111.9	67.8	56.1	37.8	21.0	10.6

Hypothetical Example of a Sensitive Supercell

- The simplest example of a sensitive supercell is when each individual component cell has only one contributor, which is often referred to as *company protection*



We want to publish:

1. Total sales for Central Atlantic
2. Sales for DC, MD, NY, and PA

But, there is a disclosure issue:

1. Sales for DE U NJ supercell can be derived
2. Company ABC or XYZ can then derive the other company's sales
3. Sales for at least one other state needs to be suppressed to protect the supercell

Resources—U.S. Statistical Policy

- One page primer on Data Protection and Accessibility
- One page summary of Principles and Practices
- Managing Information as a Strategic Resource, OMB Circular A-130
- Confidential Information Protection and Statistical Efficiency Act of 2002
- Evidence Based Policymaking Act of 2018 – requires federal agencies to apply statistical principles to program information; increases access to data.