

Dr. Strangedata, or: How I Learned to Stop Worrying and Love O&S Data Validation

Presenter: [Patrick Shelton](#)

Co-Authors: [Daniel Puentes](#), [Alexis Lewandowski](#)

May 13, 2025

The Big Question

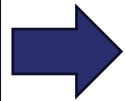
Presented at the ICEAA 2025 Professional Development & Training Workshop - www.iceaaonline.com/atl2025

How much do you trust your data?

Four Simple Axioms

Presented at the ICEAA 2025 Professional Development & Training Workshop - www.iceaaonline.com/atl2025

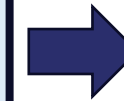
Slow Down!



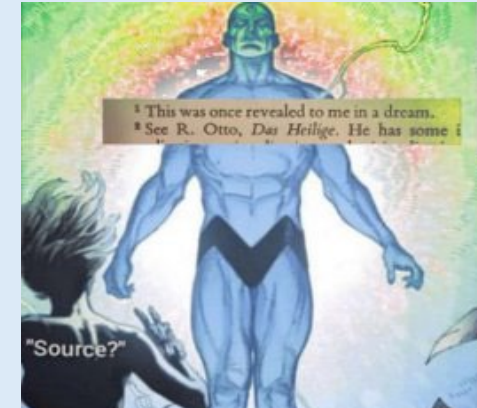
Trust Nothing



Comparison
is King 👑



Get it Straight
from the Source



Bottom Line Up Front

By conducting ***simple, common sense data validation*** best practices,
we uncovered:

**The Department of Energy's real property database of record
contains inconsistencies and errors which could limit its utility**

and

**The scale of the problem has not been widely understood
until now**

- **The Federal Property Management Reform Act of 2016**
 - All federal agencies must **annually collect and report** data on all of their **real property** assets, including:
 - *Actual annual operating cost*
- **Federal Information Management System (FIMS)**
 - The Department of Energy's database used to aggregate real property data
 - *Tracks annual operating cost*
 - *~170 other data fields*
 - Considered an example for other agencies to follow

Public Law 114–318
114th Congress

An Act

Dec. 16, 2016
[H.R. 6451]

Federal Property
Management
Reform Act of
2016.
40 USC 101 note.

40 USC 621 note.

To improve the Government-wide management of Federal property.


Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the “Federal Property Management Reform Act of 2016”.

SEC. 2. PURPOSE.

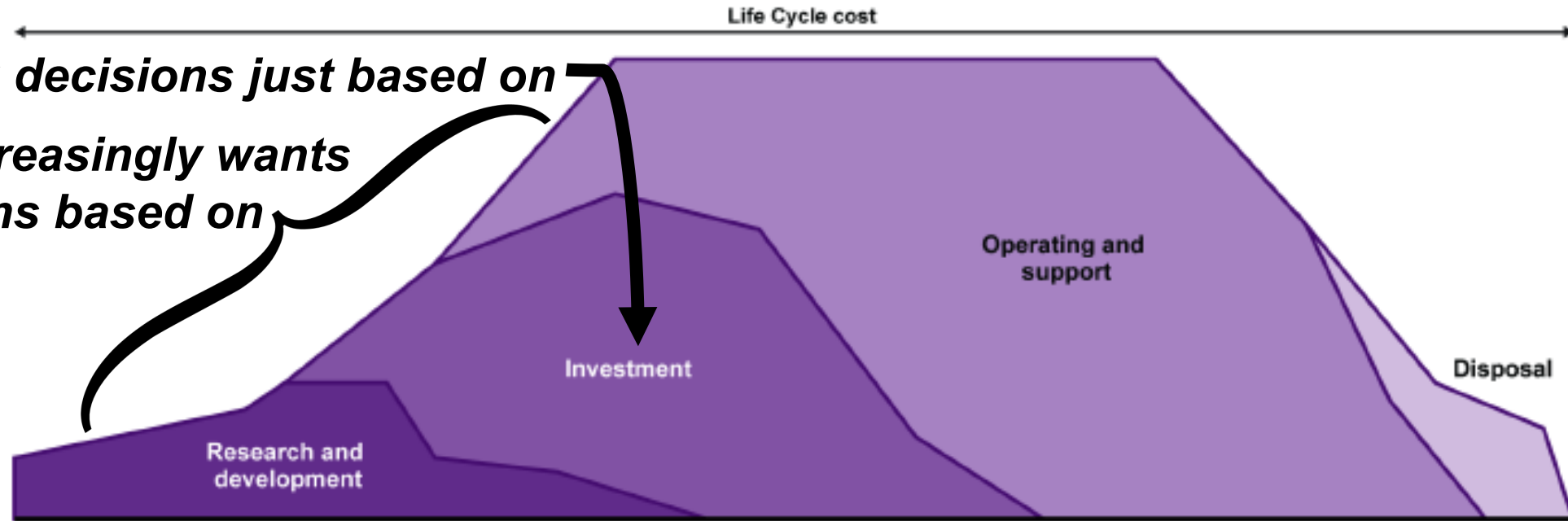
The purpose of this Act is to increase the efficiency and effectiveness of the Federal Government in managing property of the Federal Government by—

U.S. Department of Energy 

Facilities Information Management System

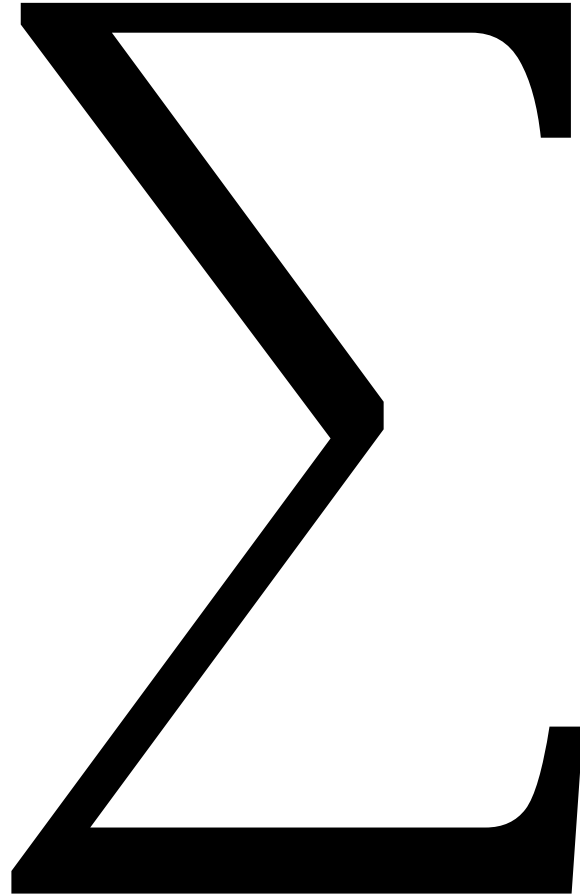
Why Are We Looking at This?

Rather than making decisions just based on DOE leadership increasingly wants investment decisions based on



PA&E (us) are the NNSA's programmatic cost estimating office:

- **Problem:** Lack of methods to defensibly estimate the Operations & Support portion of Total Lifecycle Cost
 - **Step 1:** Assess the O&S data that the DOE already collects



FIMS total O&S cost

=

Annual Actual Maintenance

Electricity

Water/Sewer

Gas

Heating

Cooling

Pest Control

Refuse/Recycling

Snow Removal

Grounds

Janitorial

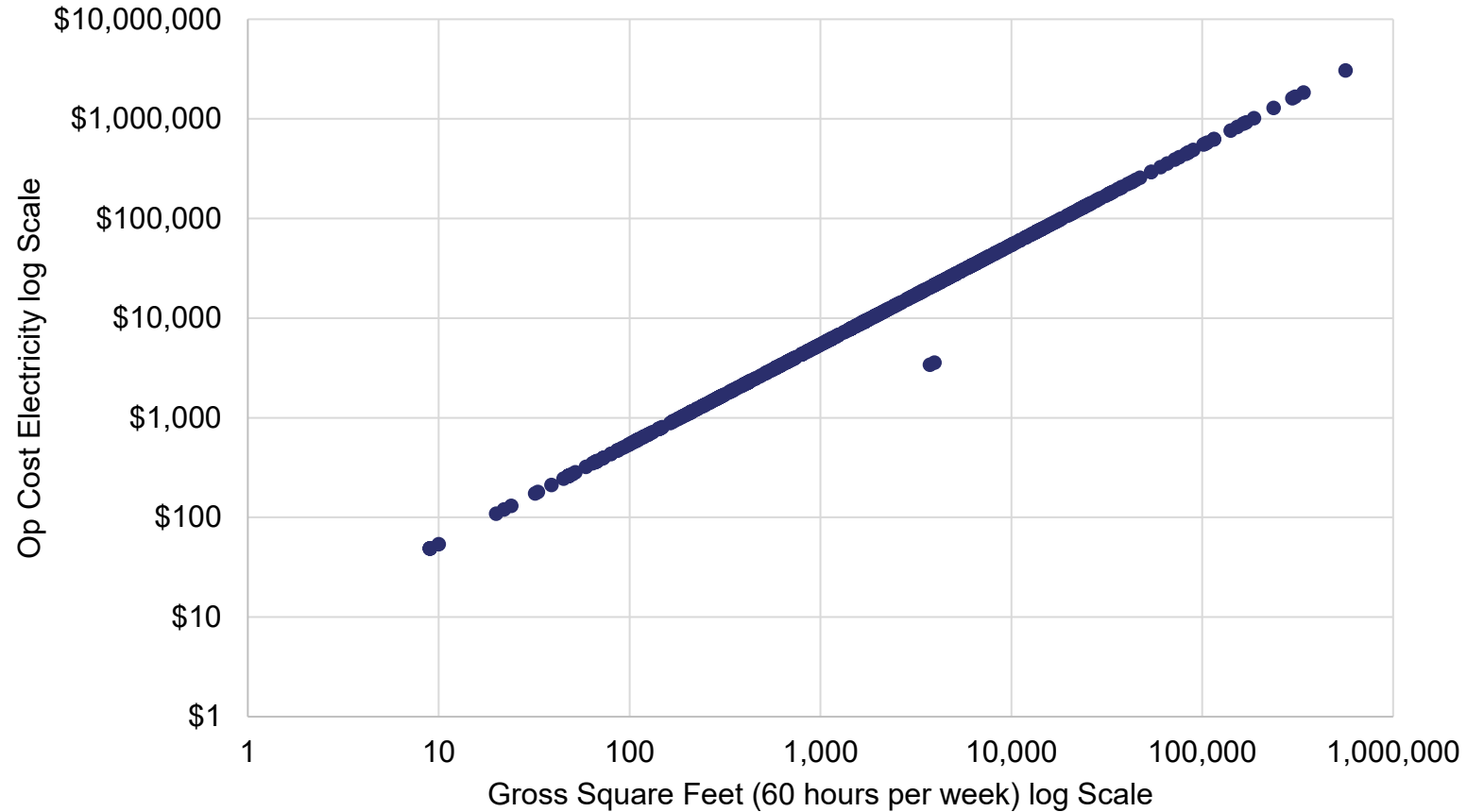
FIMS Dummy Data Example

Presented at the ICEAA 2025 Professional Development & Training Workshop - www.iceaaonline.com/atl2025

Asset Name	Size	Hazard Category	Core Capability	Electricity Cost
Pollos Hermanos	4,000 sq ft	06 Nanoparticle Facility	Product Distribution	\$12,000
The Laundry	10,000 sq ft	05 Chemical Hazard Facility	Cooking	\$30,000
A1A Car Wash	12,500 sq ft	00 Eyebrow Facility	Financial Services	\$37,500

FIMS has a big list of buildings, technical parameters, and an electricity cost assigned to each one – this data is perfect!

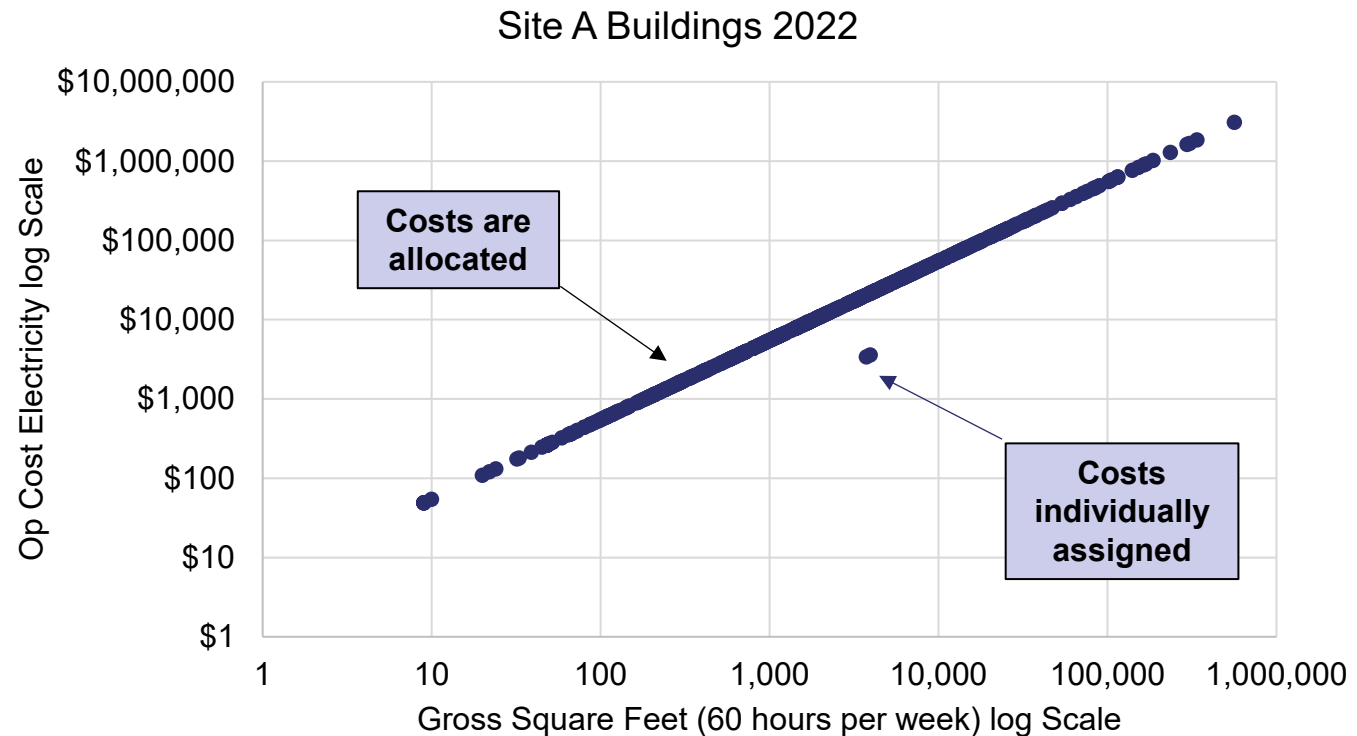
Site A Buildings 2022



Lesson 1: Slow Down!

Slow Down! Allocations, Not Actuals

- **FIMS** allocates all O&S costs (except maintenance, maybe) from the site total down to each building based on gross square footage (GSF) and operating hours per week.
- **Take the time to:**
 - *Read the data definitions*
 - *Understand how data is collected*
 - *Explore!*



**You cannot make good decisions from data that you do not first intimately understand!
Invest the time upfront to prevent spending even more time fixing it later**

Lesson 2: Trust Nothing

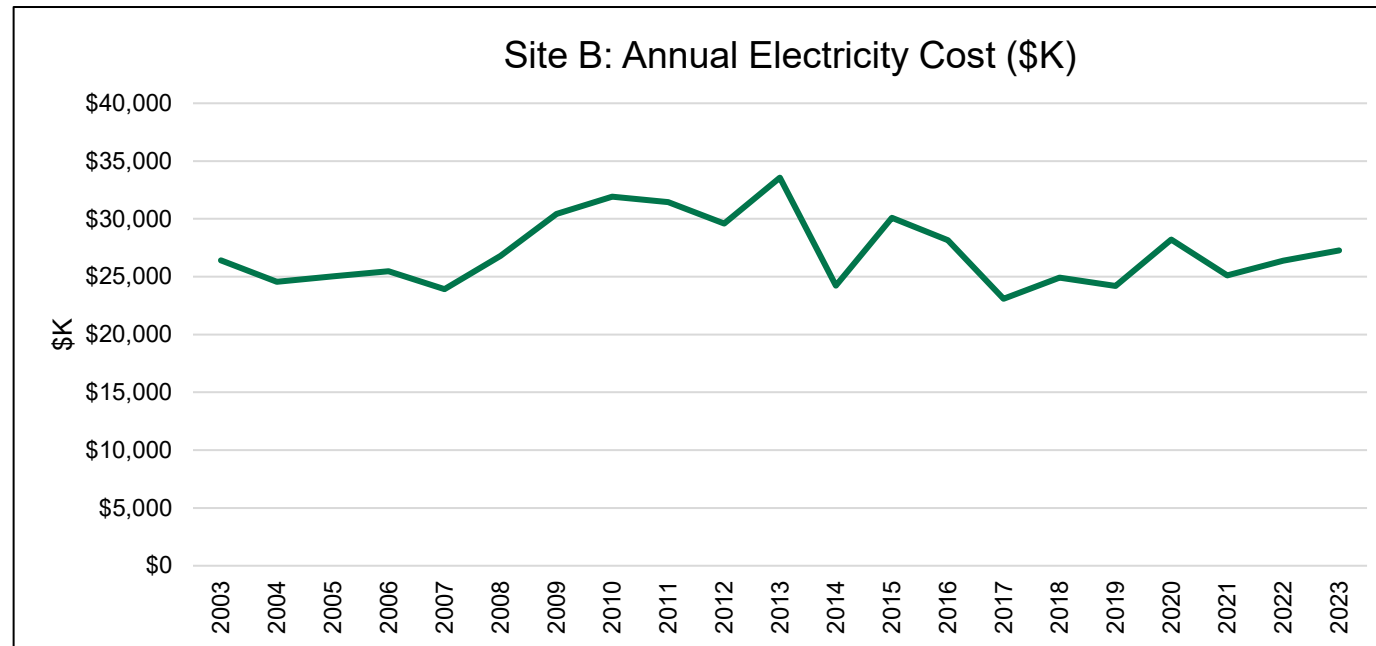
- **Now we know:** the electricity cost for the vast majority of the buildings in **FIMS** is an estimate based on the site total
- **As long as the site total is accurate, we can:**
 - Just use the **FIMS** method of estimating electricity cost via GSF and hours of operation
 - Create a more robust method that incorporates more parameters
- **Key phrase:**
 - *“As long as the site total is accurate”*



Having trust issues in a relationship is bad. Having trust issues in a relationship with data is a virtue. Any data that you have not personally collected should be met with a healthy skepticism. Do not let someone else's mistakes and assumptions become your mistakes and assumptions.

Introducing - DOE Sustainability Dashboard

- Collects various parameters related to sustainability goals
- One of those parameters just happens to be total electricity cost



Two independent sources collecting the same parameter means we can...

Lesson 3: Comparison is King

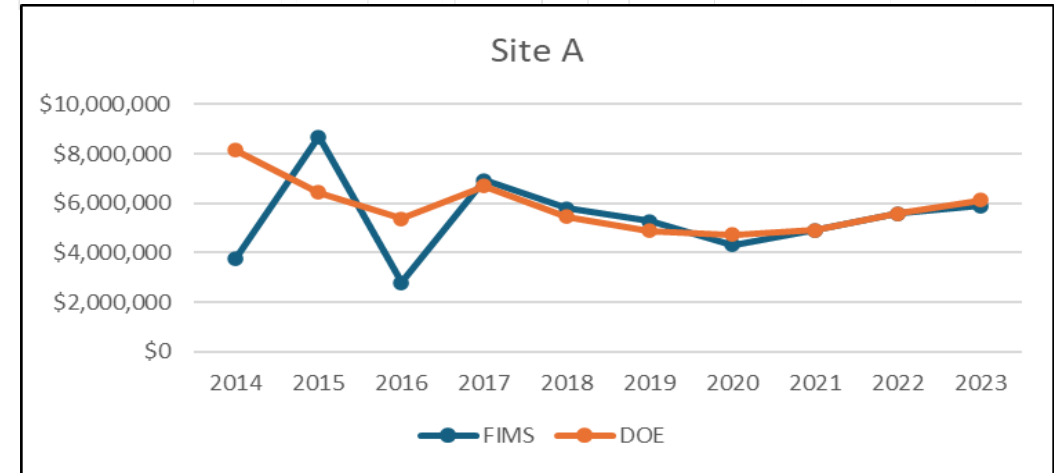
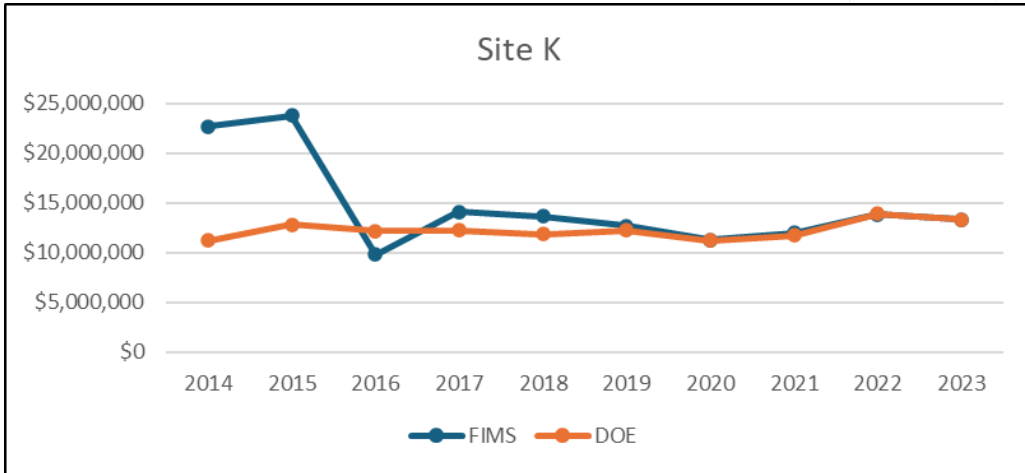


Such a simple question:

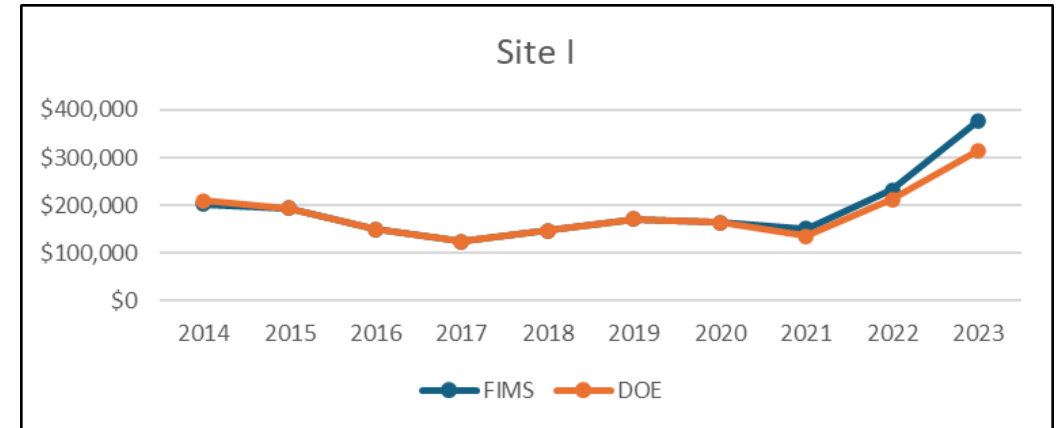
*Do the site total electricity costs in **FIMS** match the site total electricity costs in the **DOE Sustainability Dashboard**?*

It's not great for your mental health to sit on social media comparing your life with others. It's also not great for your mental health to compare data sources, but it works wonders for the health of your estimating methods.

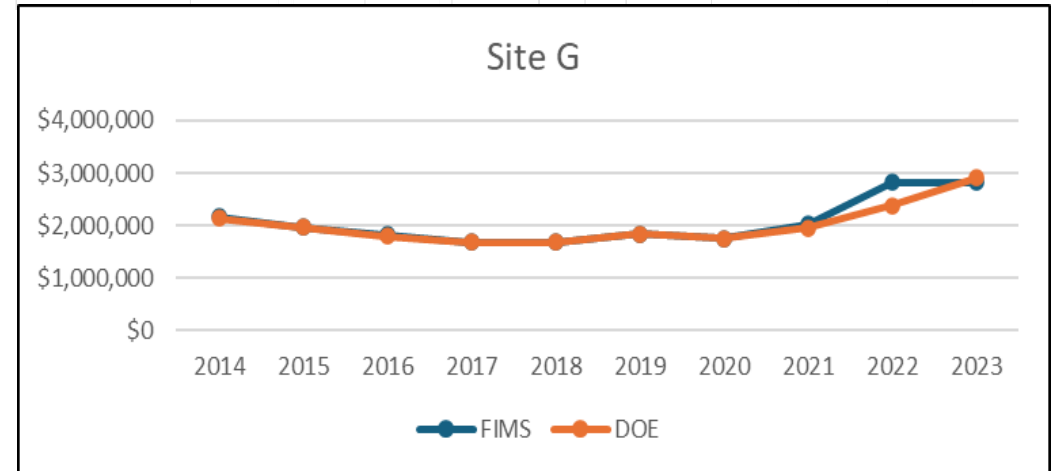
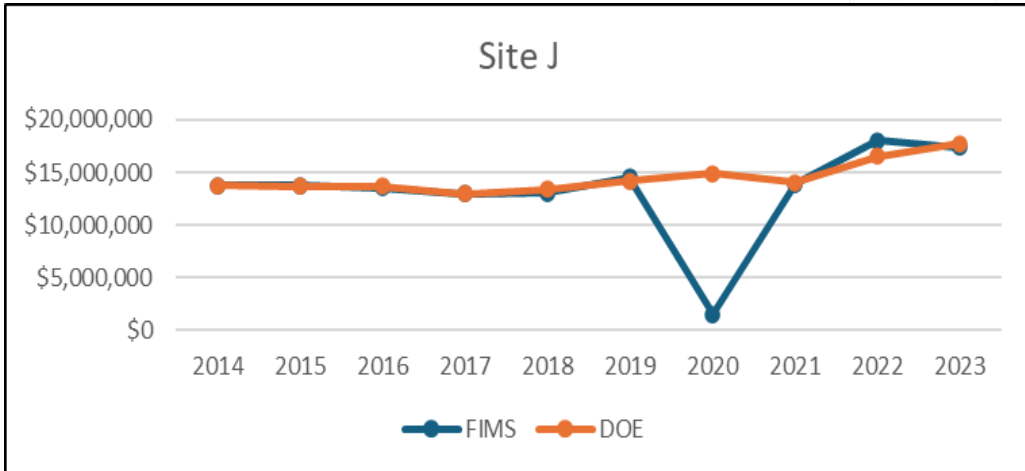
Discrepancy 1



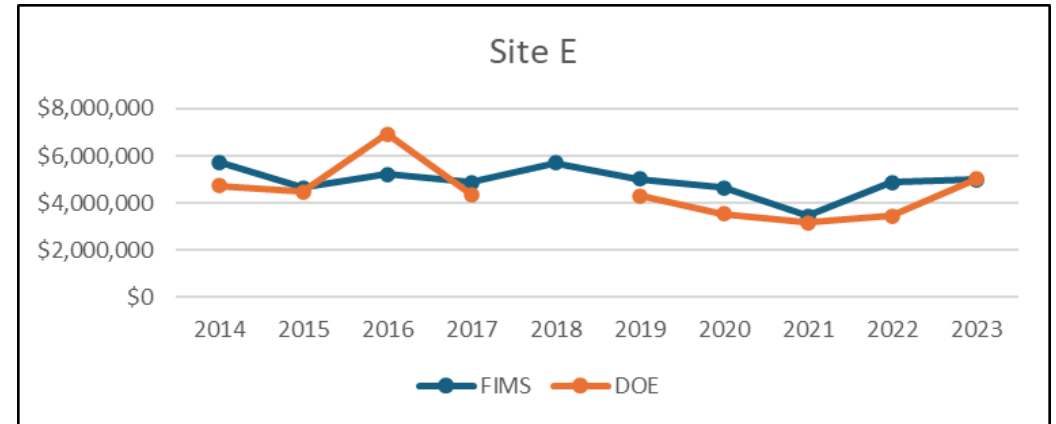
- **Site K and Site A:** previous years not aligned, but data (mostly) matches for the last several years
- **Site I:** previous years matched, but data has since fallen out of alignment



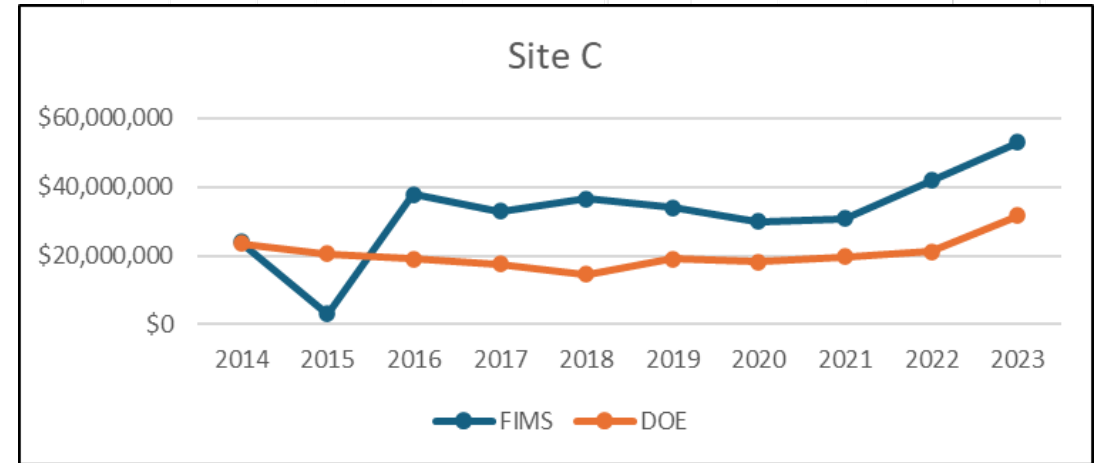
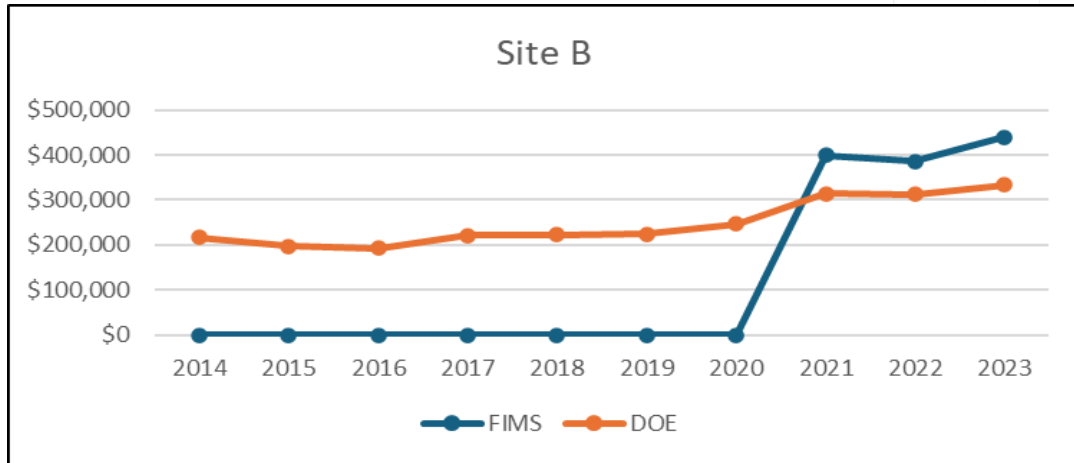
Discrepancy 2



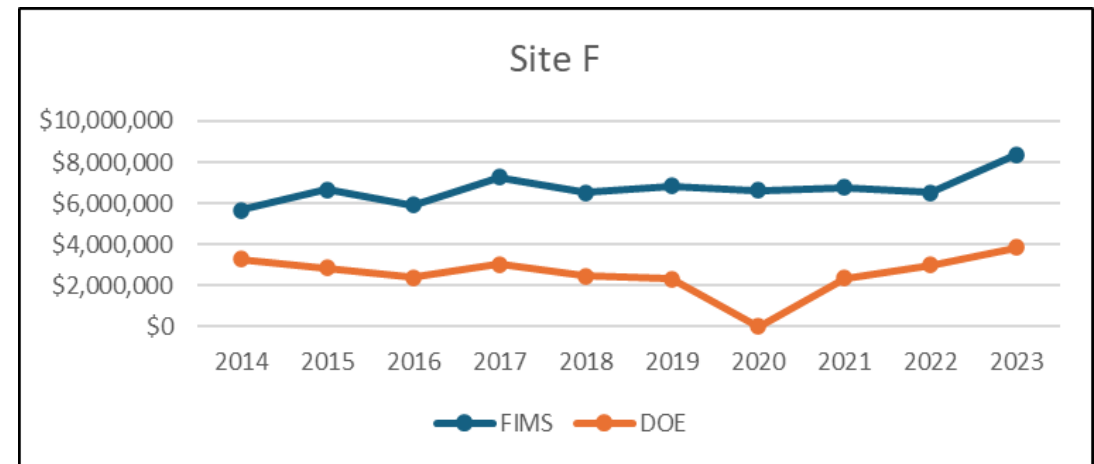
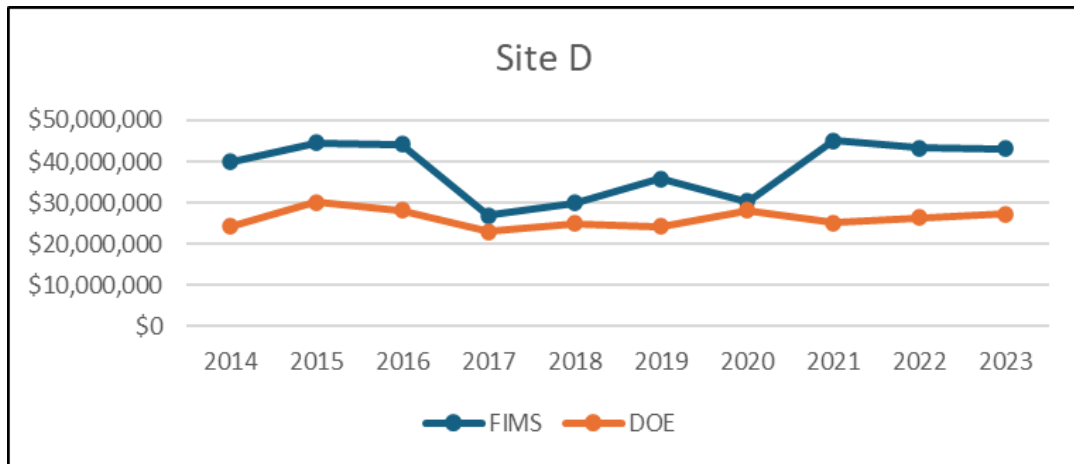
- **Site J & Site G:** data aligned except for outlier years, possibly input error or single instance of miscommunication between **FIMS** and **DOE Sustainability**
- **Site E:** data missing for **DOE Sustainability** and consistently different from **FIMS**



Discrepancy 3



FIMS cost is consistently higher than the DOE Sustainability cost



Lesson 4: Get it Straight From the Source

- Ask the **FIMS** and **DOE Sustainability** data owners –
 - ***What is going on here??***
 - Could you walk us through how your site calculates electricity cost?
 - Could we take a look at your electricity bill?
 - Could you help us understand why there are differences between the data sources?

Primary Sources >>> Secondary Sources

What's the Deal?

- **Discrepancy 1: change in reporting practices**
 - *“Turnover among the reporting staff has caused inconsistencies and changes to methodology in the past, since sites can largely determine what and how they report. There is little documentation on what changed or why.”*
- **Discrepancy 2: data errors**
 - *“Yeah, it looks like we missed a zero that year. Once we submit the data, we don't look back at it or correct it.”*
- **Discrepancy 3: multipliers?!**
 - This requires its own slides

Multipliers?! Multipliers.

DOE Sustainability: raw cost of electricity

FIMS: electricity cost * multiplier

Multiplier justification:

1. *FIMS* does not have a discrete data field for every site operating cost
2. Sites want the *FIMS* reported total operating cost = financial total operating cost

Solution:

Add a multiplier to the operating costs that are in FIMS

Result:

All of the individual operating costs are higher than the true cost of that element, but the total site operating cost in FIMS (hypothetically) matches the total site operating cost in the financial system

Multipliers?! Multipliers.

DOE Sustainability: raw cost of electricity

FIMS: electricity cost * multiplier

One site's senior management dictates the multiplier every year, so:

FIMS / DOE Sustainability should = multiplier

FIMS data for this site is inconsistent with the site's own stated data definitions and methodology

Year	FIMS / DOE Sus.	Site Reported Multiplier
2014	1.02	"internal reporting error"
2015	0.15	"internal reporting error"
2016	1.99	1.82
2017	1.87	1.76
2018	2.50	1.75
2019	1.79	1.67
2020	1.65	1.66
2021	1.57	1.64
2022	1.96	1.67
2023	1.68	1.67

Electricity Conclusion

- Based on our current understanding what data is reported to **FIMS** vs. **DOE Sustainability** and the quality of that data, the **DOE Sustainability Dashboard** is the more reliable database of utility costs for use in cost estimating
 - Raw costs with no multipliers across all sites
 - Consistency over time with minimal obvious data errors
 - Clear reporting practices and understanding of what data is being reported

How Did this Happen?

- **FIMS** claims pride in its data validation and internal auditing process
- How did we manage to break it with a couple line charts and phone calls?



Disclaimer

Presented at the ICEAA 2025 Professional Development & Training Workshop - www.iceaaonline.com/atl2025

Everything we are about to present is based on our current understanding of **FIMS' policies and procedures, as derived from official written guidance, training materials, and many detailed conversations with subject matter experts and FIMS administrators at all levels within the DOE.**

All of the following may not represent or be applicable to all parties in all years.

Straight to the Source

The Good:

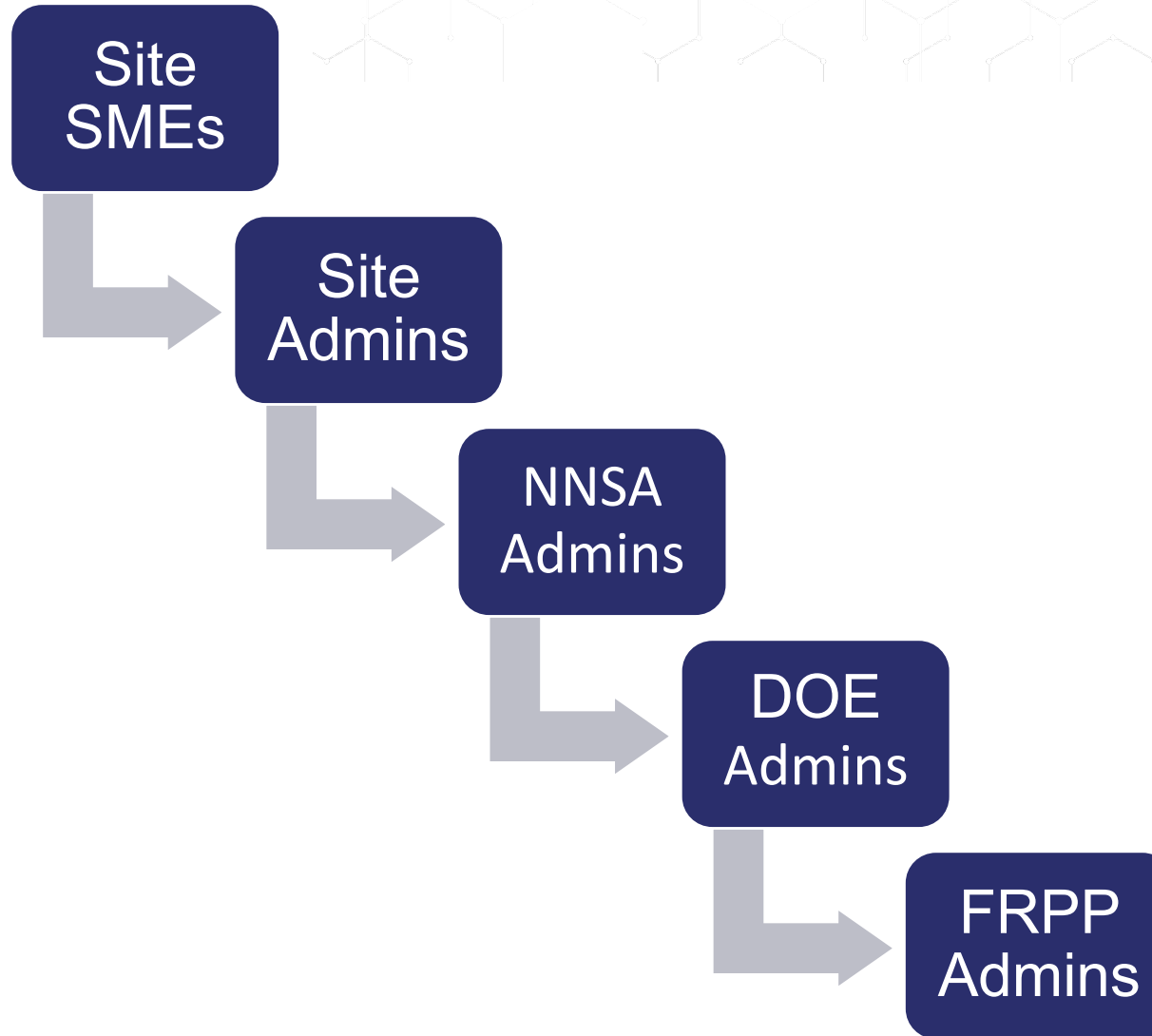
- **All data needs to be backed up with a primary source document**
 - Number in **FIMS** needs to match the number in the primary source
 - *What is a primary source?*
- **If any calculations are made:**
 - Methodology needs to be documented
 - Replicable by an outside auditor

BUT:

- **Every piece of data could be a single point failure**
 - Produced, controlled, and only understood by a single SME at each site

FIMS Reporting Chain

Presented at the ICEAA 2025 Professional Development & Training Workshop - www.iceaaonline.com/atl2025



Trust Me Bro

Case study:

Annual Actual Maintenance

One of the most scrutinized FIMS columns

NNSA admins:

Could not identify a single example of a primary source document for maintenance

Site admins:

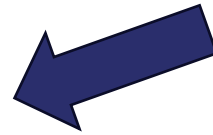
“I just enter the numbers that the SMEs give me. I don’t know how they come up with the numbers.”

Existing validation practices are more data entry checks than actually auditing the accuracy of the numbers

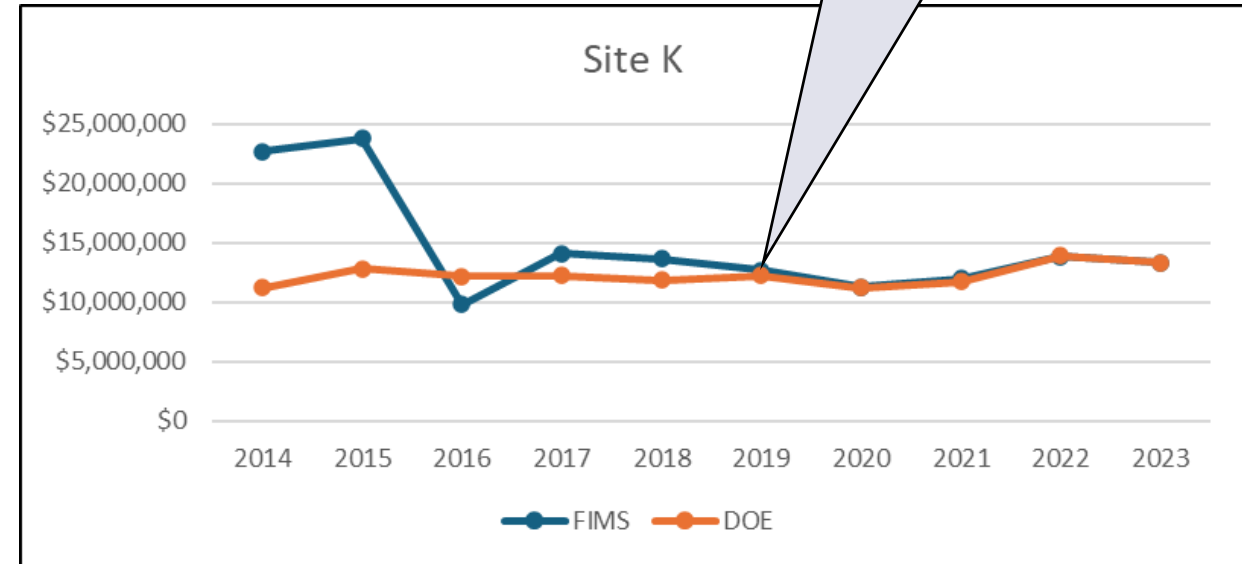


Knowledge Transfer

- **Data quality at any site:** function of the quality of the SMEs
- Admins are SME and data entry jockeys more than analytical, QA staff
- **Wide open to knowledge transfer issues when there is turnover**
 - If a new SME doesn't know how previous SME calculated electricity costs → assumptions change → numbers change
 - If a new administrator doesn't know all of the SMEs and doesn't know all of the data definitions → SME directs admin rather than admin directing SME → single point of failure created



Site K Admin:
"Hey, that's when I started!"



- **Communication with other data owners is limited**
 - **Example:** **FIMS** admins don't talk to **DOE Sustainability Dashboard** admins
 - *Different assumptions*
 - *Different data definitions*
 - *Different numbers*

Sites are rarely, if ever, talking to each other about standardizing best practices, and HQ should be facilitating these discussions or sharing information



Definition Disagreement

- **FIMS**' data dictionary does a very good job characterizing each data element, however:
 - **Some sites may choose to ignore the written definition when it doesn't neatly fit their data**
 - Total operating cost:
 - **Multiplier sites:** *"what if we disagree with **FIMS**' definition of total operating cost?"*
 - **Definitions can leave things open to interpretation**
 - *What is maintenance anyway?*



Wrong Forever



Incorrect data are *NEVER* corrected



because

Roughly half of all data audits result in at least one corrective action plan for incorrect data, but **by *FIMS* policy, that data *cannot be updated or fixed after the audit***

meaning

All corrective action plans only impact data for the following year

and there is

No accessible way of knowing what historic data had a corrective action plan issued for it

Very difficult to trust any of the historic data!

Audit Timeline

Presented at the ICEAA 2025 Professional Development & Training Workshop - www.iceaaonline.com/atl2025



HQ shows up to audit each site's **FIMS** data:

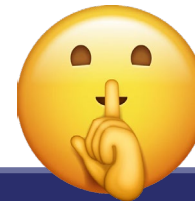
ONCE EVERY FIVE YEARS



Sites “**self-audit**” every other year, and they:

*Often give themselves **higher data quality scores** and
fewer corrective action plans*

Suggests *self-audits* are treated as a **check-box activity**



**80% of FIMS data could be unreliable, or at least more
unreliable than the other 20%**

Audit Accuracy

Presented at the ICEAA 2025 Professional Development & Training Workshop - www.iceaaonline.com/atl2025

HQ audit stated goal:

90% accuracy

But CAPs only issued when data falls below

80% accuracy

Meaning

All of **FIMS** data, even in the HQ audit years, could be **20% off** from the real number **with no documented issue**

For the audit to uncover a problem, they need access to the primary source documents and the SMEs to prove a data discrepancy/error

- Documented pattern of **SMEs “not available”** when HQ shows up
- Site admins usually unable to speak to data specifics or complex primary source documents
- HQ auditors **may skip over a data field if a SME or document is not available** rather than hold up the audit process or issue a CAP just because a SME is not available

FIMS Validation and Audit Takeaways

Good data validation policies for primary sources are in place –
just ensure they are followed!

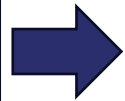
Limit single points of failure and maximize information sharing –
increase communication and collaboration across the board.
Intra-site, inter-site, between sites and HQ, between data owners

This is a full-time job for us and for the admins, so we should have the time and resources to...

Four Simple Axioms

Presented at the ICEAA 2025 Professional Development & Training Workshop - www.iceaonline.com/atl2025

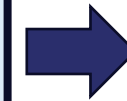
Slow Down!



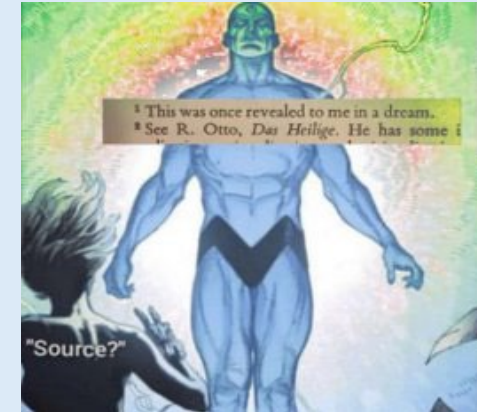
Trust Nothing



Comparison
is King 



Get it Straight
from the Source



What's the Point?

Presented at the ICEAA 2025 Professional Development & Training Workshop - www.iceaonline.com/atl2025

***It doesn't take a SME to do good
data validation.***

What's the Point?

So, if a couple of junior analysts can unravel one of the DOE's largest, most scrutinized databases with some simple visuals and questions...

The Big Question

Presented at the ICEAA 2025 Professional Development & Training Workshop - www.iceaaonline.com/atl2025

How much do you trust your data?