



# AUDITING DATA QUALITY & DATA INTEGRITY

---

Bill Hallinan and Miki Cestnik

# ABOUT US

## **Bill Hallinan, MS Software Engineering, IS Auditor**

- Worked with data quality and integrity as software developer and manager
- Celebrating his 2-year anniversary as an auditor and professional skeptic in the service of the public good
- Outdoor enthusiast – hike, bike, ski, sail. Trail name: Monk.
- Office nick-name: Boogle (rhymes with Google).

## **Miki Cestnik, CISA, IS Audit Manager**

- Bill's Supervisor (comes with extra duty pay)
- Newly appointed elementary teacher, bus driver, cafeteria lady, school counselor, school nurse, and PE teacher.

# ABOUT MONTANA . . .

The Constitution of the State of Montana,  
Article V, Part V, Section 10 Organization and procedure

The legislature **shall establish a legislative post-audit committee** which shall supervise post-auditing duties provided by law.



# WHY?

- Last year, NSAA conference attendees suggested the topic of **auditing data reliability and integrity**.
- Also, we talked about the **need for data reliability**, so now we can talk about **how** you do that.

## What we hope you bring home:

- The **when, where, what, and how** of auditing data quality and integrity.
- Confidence to ask for and to assess the quality and integrity of data used in your audits.
- Time saving tips and tricks, e.g. staying on track vs off track.

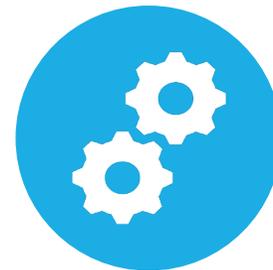
# DISCUSSION OVERVIEW: DATA QUALITY & DATA INTEGRITY\*



**WHEN SHOULD  
WE LOOK AT  
THEM?**



**WHERE TO  
SEARCH  
FOR THEM?**



**WHAT ARE  
THEY?**



**HOW DO WE  
MEASURE  
THEM?**

\* For purposes of this discussion “data” means any files or database records stored electronically

# WHEN?

The line between reliability of data for other analysis and the integrity/quality of data is based on **risk**.

- Is there a risk that the agency is using “bad” data?
  - **If this is an audit, it’s done during fieldwork.**
- If the risk that the agency is doing something “bad”, can the data prove it?
  - Need to make sure, for standards, that the data we use is reliable.
  - **Needs to be done during planning the audit so methodologies of analysis take in to account findings and need for sufficiency.**

Our discussion going forward is going to focus on the risk of the agency using bad data and fieldwork addressing data quality/reliability.

# ACTIVITY



Write down **three instance** of when your audit work was delayed or possible findings diminished because of data quality or integrity issues. Could any of these **have been avoided** by asking about or addressing data quality sooner?

## For example:

- Not able to do your analysis because data was missing or improperly stored.
- Sorting good data from bad made any analysis too costly or impossible.
- Spent too much time looking for data that should have been easy to find.
- Your audit interviews showed nobody had confidence in what was collected.

# WHERE?

“We know it when we see it, **but we cannot see it when we need it.**”

Have a conversation about data. It's an opportunity to educate your clients. Audits work better when clients are good data steward and proud of it. Please show me where you keep <blank>. They are important to the audit and business because <blank>.

For example, where are the:

- Data collection plans
- Data definitions and dictionaries
- User access records and error logs
- Support tickets for bug, data errors, and release notes
- The staff that work closely with, and **care most about**, the data being right

# ACTIVITY



Write down an instance of where you **unexpectedly** found good data (a treasure trove), and an instance where you were **sorely disappointed** and found embarrassingly awful data.

## For example:

- **Good:** double blind, independently lab-certified, statistically assurance, including data collection plan, soil data for every lead smelter in the Western US.
- **Bad:** the **only copy** of the publicly-traded company's prior years accounting and human resource records on a backup tape that was "mostly" readable.

# ATTENDANCE CHECK



We interrupt our remote presentation for a number  
from our sponsor 😊!

# WHAT IS DATA QUALITY?

Do you need to know if this is “good” data or “bad” data, then ask, “Can I use it **right now** to make an important business decision?”

Long cycles (a year to collect and clean data) vs short cycles (good to go as we receive it)

Things to think about (see graphic to right)

Examples:

African HIV data collection program

Engineering data used for bridge construction

Property data used for taxation purposes



Completeness

Consistency

Conformity

Accuracy

Integrity

Timeliness

Completeness

# WHAT IS DATA INTEGRITY?

Assuming all data was quality data once, is it still good enough today to make important decision?

For example, “For my audit I need today’s data and the data from five year ago for the audit objective.”

Things to think about (see graphic to right)

Overtime, many things change: people, rules, data systems, knowledge, definitions.

- Those numbers differ because two years ago we started recharacterizing x as y.

Good examples of stable data: whether, finances, census



Legible  
Original  
Enduring  
Accurate  
Available  
Complete  
Consistent  
Attributable  
Contemporaneous

# ACTIVITY



From your audit experience, write down one example of **bad-to-begin-with data** masquerading as good data, e.g. Cinderella-Sister-Data. (Quality).

Write down an example of previously **good-data-gone-bad**, e.g. Walter-White data. (Integrity).

# HOW?

**Standards already prescribe** the elements for and levels of quality and integrity: ISO, EPA, FDA, GSA, DHS, DOD, OMB, IAAO, Census, NIST, etc. Beyond that:

**Ask** the auditee, “What **data quality standards** do they use formally or informally use?”

**Ask**, “What was the **effort and cost** to collect your data?” and “Would there be any business impacts if you lost or corrupted your prior years’ data?”

**Ask**, “What **measure** they used to determine their data is good enough?”

**How** can you tell if you are on track with data quality? **The good data and its relationships are easy to find, understand, and yield consistent, impactful results.**

**Great resource:** [Methodologies for Data Quality Assessment and Improvement](https://www.researchgate.net) at <https://www.researchgate.net>

# EXAMPLES



- **Data analysis: gain full access** to examine data schemas. Perform interviews to reach a complete understanding of data and related architectural and management rules. Use **agency/auditee tools** to explore. **PeopleSoft, Retirement, Property Tax, Environment.**
- **Data quality requirements:** What is sufficient quality? Who leads? Who has the last say? Examples: layered management. **Property Tax, FC Auditor requirements**
- **Critical areas:** Is the data used in critical areas? e.g. payments, safety, privacy, health, allocation formulas? **Environment, Education, Taxation, Health and Human Services**
- **Process modeling:** How is the data received, manipulated, stored, retrieved? **Property Tax. Environment.**
- **Measurement of quality (objective or subjective):** Who can see it, change it, updated it? What do the users say about it? Error Logs? **Retirement, Property Tax, People Soft**

# DISCUSSION



**Thank you** for your attention to our presentation!

Here's the point in our program where **we stop**, and **you start asking us questions and/or share your best (or most outlandish) question response.**

**You are up!** Let the Q&A begin! Just write your question/response to the moderator in the chat box. She will read them for us.

# BRING IT HOME

**Thank you again for your time and attention!**

**Auditing data quality and integrity isn't easy because it's a moving target and every audit and auditee have different thresholds, but we can't shy away from it.**

It takes a lot of work with everyone to agree on something, or at least discuss it. The effort has benefits for both communication and impacts.

Together we've shared the when, where, what, and how. We encourage you to bring it on home! (Well you might already be at home, but you know what we mean 😊!)

**Good evening, everyone!**