Clean Data Means Better Decisions

Webinar

February 26, 2025

NATIONAL INSTITUTE ON
Retirement Security
Reliable Research. Sensible Solutions.

Agenda

- Logistics and Introductions
- Presentation and Polls
- Q&A

Speakers

Laurie Mitchell

Senior Business Consultant

Tegrit

Jonathan O’Reilly

Director of Operations Audit & Business Intelligence

Teacher Retirement System of Texas

Margaret Rogers

Director of Communications and Member Relations

National Institute on Retirement Security

Getting Clean Data

Laurie Mitchell, Senior Business Consultant

TEGRIT

Why is Clean Data Important?

- Today, pension systems are finding that their IT admin solutions either turn over or need major upgrades about every 20-25 years.
  - Due to dramatic advances in technology security and systems management.
- Your IT systems will need to adapt often and expertly to keep data secure and operations running smoothly.

Why is Clean Data Important?

- This is especially true for data.
- Rapid changes continue to occur in the volume of data that can be stored, storage methods and security solutions.
- Clean, well-managed data is paramount for smoothing out the multiple upgrades and transitions your system will experience in its lifetime.

POLLING QUESTION #1

Does what we just described resonate with your experience?

Which of these applies best to your pension agency?

- We have upgraded/replaced our PAS within the last 20 years.
- We plan to upgrade/replace in the next 20 years.
- Both
- Neither

Types of Data to Find and Clean

- Incomplete
- Inaccurate
  - This includes faked and unnecessary data
- Missing
  - This includes rogue or fugitive data

Incomplete Data

The most common type is created when agencies choose to store summary information (like totals) instead of detailed information.

This is especially true for wage and contribution information.

Incomplete Data - *Resolution*

1. When you face a transition, store your data at the most granular level that is available.
2. Rely on an experienced data architect to map your data into a logical, accessible solution.

Inaccurate Data

The most common type is produced by deferred and inactive members who do not keep their contact information current.

Inaccurate Data - *Resolutions*

1. A robust contact campaign – especially for members where a benefit or a refund will eventually be due.
2. Include these members in your annual member statement process.
3. Reach out regularly by email to keep both the email and physical addresses current.

Inaccurate - Faked Data

- Occurs when needed data isn’t immediately available and the admin system allows for blank or free-form entry.
- Almost never complies with business rules.
- Interferes with logical processing.

Faked Data - Examples

- Fake SSNs, like 111-11-1111
- Member and spouse with the same SSN
- Forced birthdates, like 01/01/2000
- ZIP codes as 00000
- Name placeholders, like John Doe

Faked Data - *Resolutions*

1. Write/run queries to identify fake data.
2. Review and prioritize reports.
3. Assign resources to correct the data. This could involve reaching out to the employers or the members to get correct information.

Inaccurate - Unnecessary Data

- Status fields
  - e.g., refunded, vested, returned to work
- Total fields
  - e.g., service totals, contribution totals
- Maintenance can be problematic.

Unnecessary Data- *Resolution*

1. Letting go can be hard – especially those status fields!
2. Ask your data architect to show how you can get the information you need. It's important that you are confident that you can trigger the workflows and run the reports you need.

Missing Data

- This simply doesn’t exist. You never collected it and never stored it. Here are some examples.
  - You need to contact a beneficiary, but you only have name and relationship.
  - You want active members to know about a retirement seminar at their school district, but you don’t have district.
  - You want to connect retirees to county services but didn’t collect their county of residence.

POLLING QUESTION #2

Who has experienced missing data?

Do you have an example?

Submit using the Q&A button.

Missing Data- *Resolution*

1. Expand your dataset to include the information you wish you had. Be inclusive with your internal team to capture data needs.
2. Ask your data architect to ensure that your database is flexible enough to add fields as your business matures and new ideas evolve.

Fugitive Data

- Data that is hard to find or access.
  - Spreadsheets
  - Old Access databases
  - Microfilm/Microfiche
  - Stored off-site?
- Outdated formats – or formats that are just different.

Fugitive Data – *Resolution*

1. Digitized data
    a) Reformat, clean-up and move into the central PAS database.
2. Non-digitized data
    a) Assess the quality and value of the data against the effort to convert.

Impact of Low-Quality Data

- Operational issues
  - Incorrect service credit totals affect eligibility.
  - Address inaccuracies can impact payment delivery.
- Reporting issues
  - All operational reports rely on good data.
  - Bad data makes for un-informed decisions.

Making Great Decisions

Jonathan O'Reilly, Director of Operations Audit and Business Intelligence
Teacher Retirement System of Texas

TRS
TEACHER RETIREMENT SYSTEM OF TEXAS

POLLING QUESTION #3

Do you have a data analytics team?

Select the answer that best reflects your organization's analytics maturity level.

- Proficient - We use operational dashboards and analytical scripts/models as an integral part of our operations.
- Getting Started - We've started a few analytics projects, but it is not on a regular basis or highly complex.
- Plan to start - We're looking to break ground on a data analytics initiative.
- Data analytics? Why would we ever do that?

POLLING QUESTION #4

What percent of data science projects get completed?

- Over 75%
- 50-75%
- 25-50%
- Less than 25%

Only 15-20% are successful!

- Source: Gartner. Of those, only 8% generate value according to CEOs.
- Reasons
  - Unclear objectives or lack of business engagement
  - Data projects conceived by data teams and without buy-in or guidance from a business sponsor
  - The business opportunity was fleeting and the data product not timely.
  - Wrong metric selection to answer the business question
  - Over-investment to create a perfect model when a 90% model would unlock incredible business value in shorter time

Building Your Data Analytics Program

Beginnings

- Know how data analytics will support your business objectives
  - Depending on where your org is at, this can be at an enterprise, departmental, or project level.
- Build off existing or accessible resources.
- Don’t expect to prove your ROI all at once.

Building Your Data Analytics Program

Role of Leadership

- Ask questions that data can answer
- Give voice to data-driven decisions
- Celebrate – and promote – successes
- Create a culture of data literacy by encouraging and sponsoring data analytics projects.

Building Your Data Analytics Program

Proof of Concept Projects

- Understand the business question/use case, and necessary program rules.
- What data is needed to address the question.
- What does the output need to look like?
- Stay connected to the business owner – meet regularly and share early solutions (e.g., fail fast).

Building Your Data Analytics Program

Tools and Techniques

- "Many roads lead to Rome"
  - Data reporting - SQL
  - Data integration – SAS, Altair, Alteryx, Python
  - Data visualization - Power BI, Tableau

Building Your Data Analytics Program

Measure Successes

- Use data to prove that data analytics initiatives work.
- Establish Key Performance Indicators (KPIs).
- Measure project success against those.
- Examples:
  - Average time to process.
  - % of cases processed outside of SLA.

Applications for Data Analytics

Continuously ensuring quality

- Service-level agreement (SLA) reports, KPI measurement
  - Quarterly and Monthly are common
  - Weekly/Daily less common
    - Identify problem areas sooner and address it before it hits your quarterly report!

Applications for Data Analytics

Continuously ensuring quality

- Trend and Risk measurement
  - Changes in mix of paper vs. electronic applications
  - Changes in mix of call center issues over time
  - Monitoring your risk tolerance

Applications for Data Analytics

Continuously ensuring quality

- Exception/QA reports
  - Sanity checks for assurance
    - Recalculate annuities, COLAs, valid SSNs, deceased members
  - Vendor compliance, regulatory compliance
  - Cross-system checks
    - Data migration from legacy to new system
    - Pre-implementation accuracies of new vendor

Application for Data Analytics

Case Example: Targeted Training Analysis

- Analyzed employer's late submissions and errors with uploading payroll reports.
- Developed key metrics that indicate ongoing, problematic behavior with certain aspect of the report submission process.
  - Extent of days late
  - Affected members on the late reports
  - % of reports with late errors

Application for Data Analytics

Case Example: Dental/Vision Launch

- Independent testing of accuracy of new enrollments and premiums.
- Simulated enrollments in IT test environments, verified system premium amounts against expected premium amounts.
- Helped provide assurance to a process that took in 100k+ enrollments.

Application for Data Analytics

Understanding Customer Behavior

- Funnel analysis
  - 9 step process: How many start at step 1 and drop off thru to step 9?
- Slicing the data to understand a trend
  - Increase in X – what's driving that?
    - Check the mix of likely indicators: sex, years of service, account status.
    - Look for over/under-representation.
    - What's normal vs. not normal.

Free Resources

- For leaders
  - SAS Data Literacy Essentials
  - INFORMS Building Successful Analytics Teams
  - IIA Data-enabled Internal Auditing
  - IIA Increase your Data Analytics ROI
  - IIA The New Basics of Data Analytics

- For practitioners
  - INFORMS Job Task Analysis
  - Alex the Analyst SQL Tutorial YouTube Playlist
  - Coursera 2025 Top 7 Data Analytics Certifications

Questions

What are some of your data successes?
What are your data challenges?
What are your data goals?

Laurie Mitchell
Laurie.Mitchell@tegrit.com

Jonathan O'Reilly
Jonathan.oreilly@trs.Texas.gov