

We found a data quality issue. Now what?

Presenter: Kate Mullins

Co-Author: Hailey DuBreuil



MEET OUR DATA QUALITY TEAM



Kate Mullins

kmullins@hsri.org

Project Manager



Hailey DuBreuil

hdubreuil@hsri.org

Project Coordinator



Katie Howard

khoward@hsri.org

Data Scientist



Yuan Zhang

yzhang@hsri.org

Research Analyst

AGENDA

01

HSRI and Our
Approach

02

Presentation
Objectives

03

What are data
quality issues?

Issue
Resolution
Framework

04

Future
Directions

05

Questions

06

HSRI AND OUR APPROACH



Human Services Research Institute

We are a nonprofit, mission-driven organization. We use our data expertise—developed over 40+ years—and our understanding of the complete health and human services landscape to help agencies and communities improve the health, well-being, and economic stability of the populations they serve.

Housing & Homelessness | **Population Health** | Aging & Disabilities
Child, Youth & Family | Behavioral Health | Intellectual & Developmental Disabilities

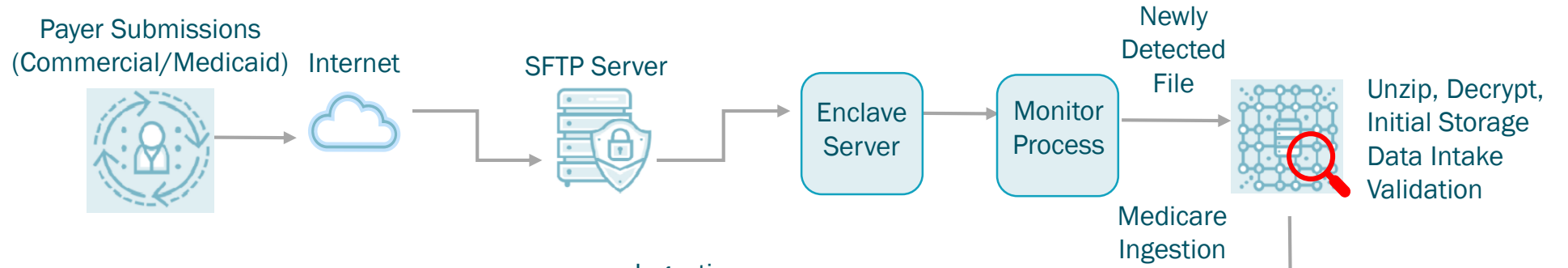


Population Health Team: What We Do

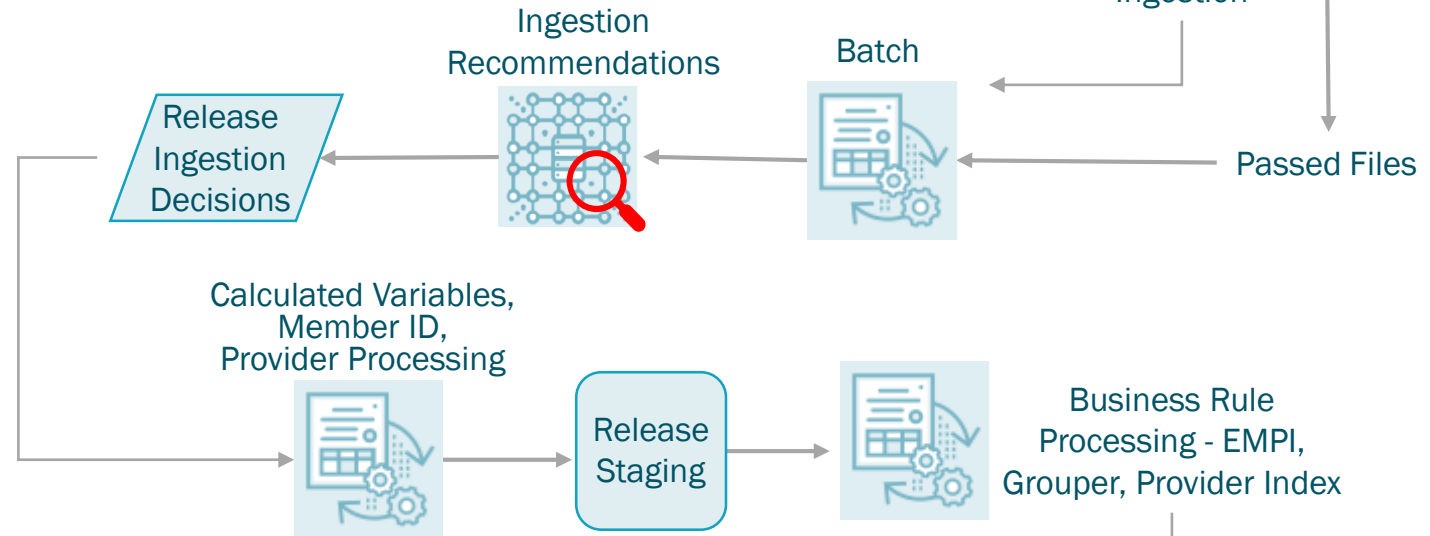
We develop and maintain nonproprietary data collection and reporting systems, custom analytics, state-level health data warehouses, data quality improvement procedures, and healthcare transparency websites.



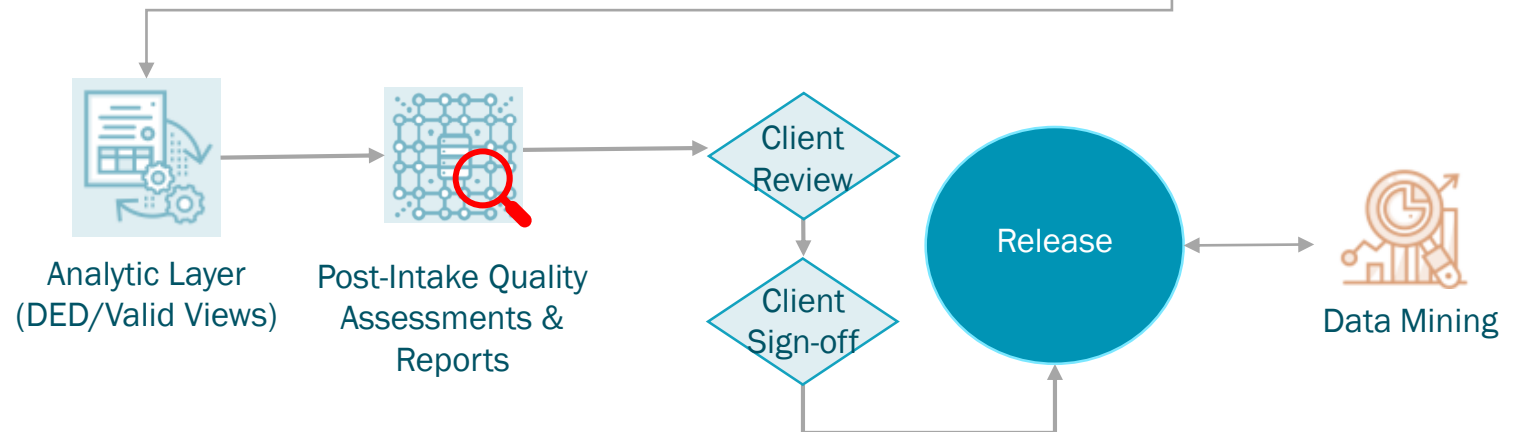
1. Data Submission & Staging



2. Data Warehouse Processing & Enhancement



3. Extracts, Analysis-Ready Datasets, and Reporting



Data Quality Approach

Continual Improvement and Flexibility



- Feedback loop with internal and external stakeholders
- Regular process improvement procedures with flexibility to address more immediate issues

HSRI Help Desk Provides Data Intake Support

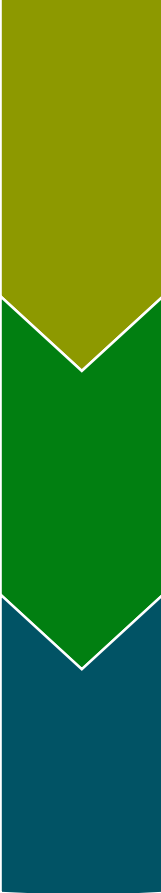
- Detailed technical support to resolve validation issues and ensure data are submitted in a timely manner
- All support requests received during business hours (Mon.–Fri.) are responded to in 2 hours or less
- Accessible via:
 - Toll-free phone number
 - Email
 - Web contact form in the HSRI Data Submission and Quality Portal



PRESENTATION OBJECTIVES



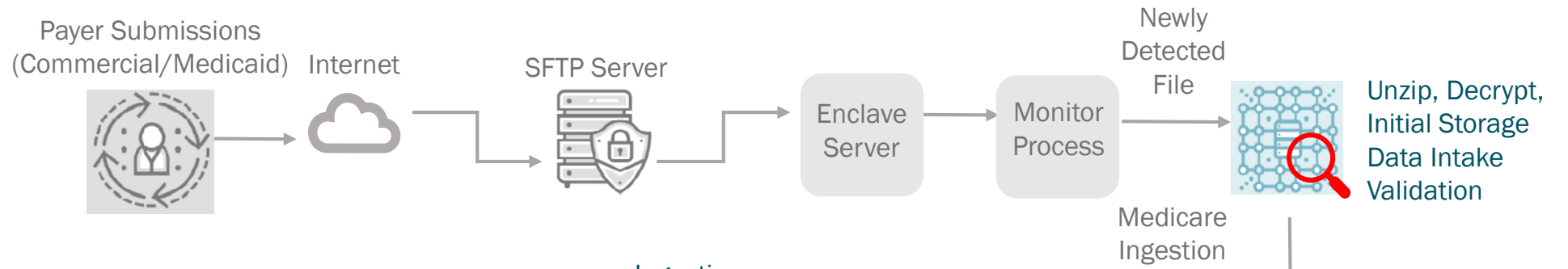
Today's Objectives

- 
1. Inform data users on the complexity and challenges of resolving APCD data quality issues
 2. Provide a framework for states navigating the process:
 - a) Where to focus limited resources
 - b) How to approach decision making and resolve issues
 3. Make recommendations for future directions

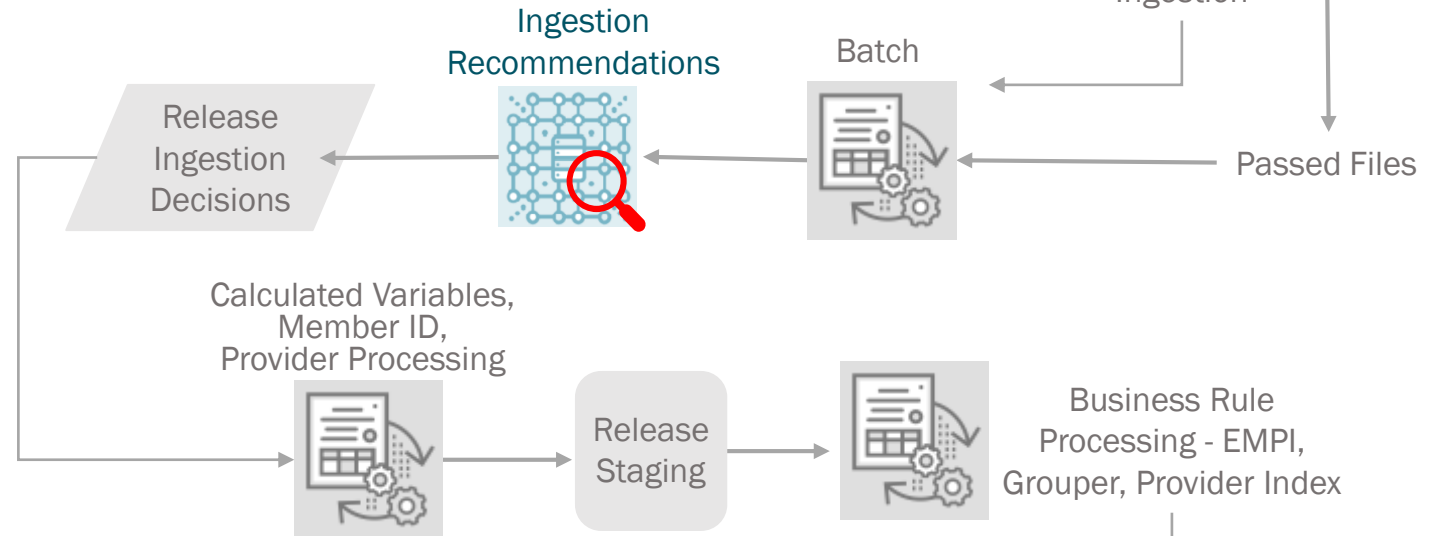
WHAT ARE DATA QUALITY ISSUES?



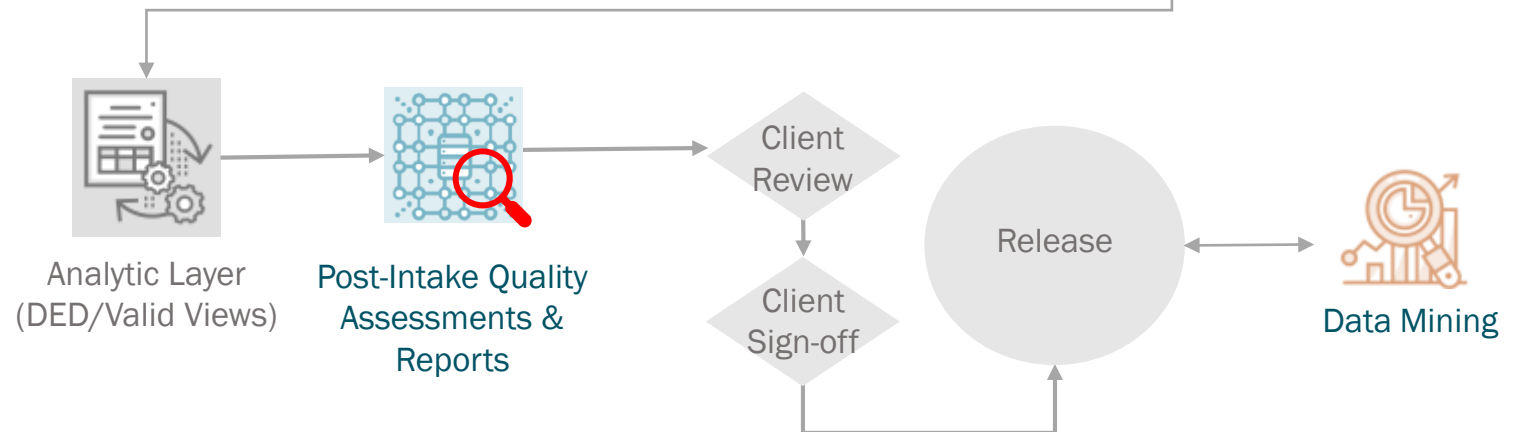
1. Data Submission & Staging



2. Data Warehouse Processing & Enhancement



3. Extracts, Analysis-Ready Datasets, and Reporting



What are data quality issues?

- Inconsistent claim and/or encounter volume over time
- Inconsistent PMPM over time
- Low match rates for Patient/Provider/Encounter identifiers
- Inconsistent population of fields over time
- Mismatch of results when compared with external sources



How do we find out about issues?

- Data Submitter Self-Report
- Post-Intake Quality Assessments & Reporting
- Data Mining
- Data Users (internal and external)
- Implementation of Third-Party Tools



Why are the issues difficult to address?

	APCD Administrator	APCD Data Vendor	APCD Data Submitter
Limited Resources	●	●	●
Competing Priorities	●	●	●
Revolving Door of Data Quality Issues	●	●	●
Staff Turnover and Training	●	●	●
Varying Requirements Across States <ul style="list-style-type: none">• Tolerance for Issues• Policies for Resubmission• Data Submission Platforms			●

ISSUE RESOLUTION FRAMEWORK



Framework for Resolution

Prioritize
Issues

Is this issue
worth pursuing?
How urgent is it?

Identify
Resolution
Options

What are options for
resolving the issue?

Make
Decisions

What questions
should be asked to
choose the best
resolution option?

Implement
Resolution

How can we prevent
a similar issue in the
future?

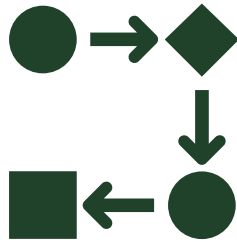
How can we best
communicate the
resolution?

Considerations for Prioritizing Issues



Priority Payers

Issue impacts a high number of covered lives or high percentage of APCD, payer types, etc.



Status of Data

Data impacted are in use



Key Fields







Used for Member Identification, Claim Versioning, etc.



Time Periods

Issue impacts multiple months or years

Options for Resolving Issues

	PROS	CONS
 Update Documentation (e.g.: submission guide or rule)	Future standardization across payers	No immediate impact Lengthy approval process
 Modify Data Quality Identification Processes	Resolution in future submissions Future standardization across payers	No resolution in data
 Educate Submitters	Resolution in future submissions	Historical issues remain
 Request Resubmission from Submitter	Historical issues resolved	Potentially time- and resource-intensive
 Remediate Data by Administrator/ Vendor	Historical issues resolved	Patchwork code
 Modify Data User Documentation	Users can work around issue based on use case	No resolution in data

Decision-Making Questions

- Is this a significant issue that makes the data unusable?
- Can data users code around the issue easily?
- Is there time sensitivity to resolving the issue?
- How many payers does the issue affect?
- Does the issue occur in recent data (last 5 years)?
- Does the issue span more than a short amount of time (e.g.: 1 month)?



Decision-Making Questions

- Can the submitter fix the issue and resubmit?
- Will resubmission cause unintended consequences/other quality issues?
- Is there other specialty code that needs to be considered when processing?
- Will remediation cause unintended consequences/other quality issues?
- Does the issue warrant the resources necessary for resolution?



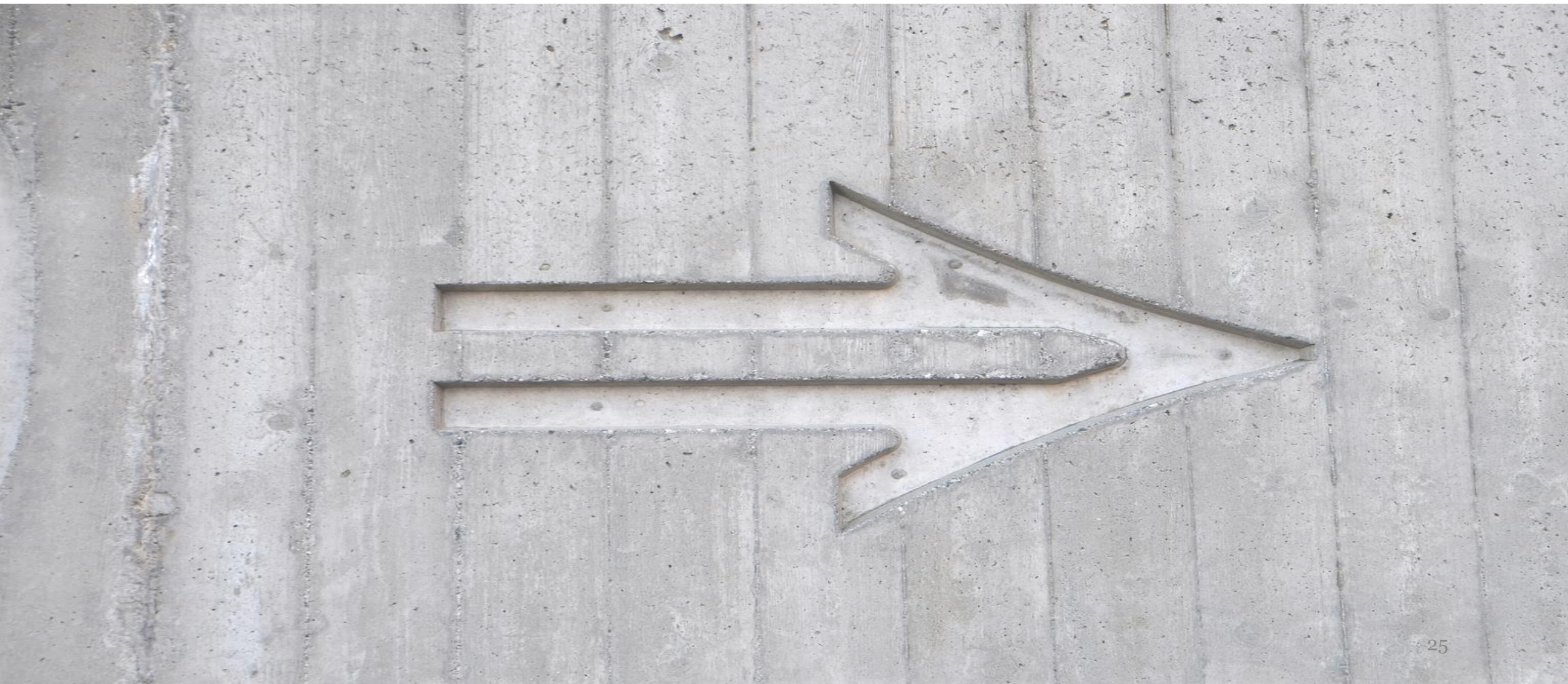
Example 1: Insurance Product Type

Issue:	All claims are submitted with the same Insurance Product Type (IPT) code while the eligibility IPT has variation
Identification Method	Data analysis and reporting
Decision-Making Considerations	<ul style="list-style-type: none">• Data released to users• Payer has a large number of covered lives and makes up large proportion of APCD• Issue occurs in recent years (past three years)• Time constraints to fix the issue for data users• Data resubmission may cause shifts/unintended consequences
Resolution	<ul style="list-style-type: none">• Administrator/vendor data remediation in historical submissions• Submitter education & resolution in future submissions• Exploring modification of data quality identification processes• Timeline from issue identification to resolution implementation: 2 months

Example 2: Provider Network Unknown

Issue:	Many payers submitted a high volume of claims with “Unknown” for Provider Network Indicator, which indicates if the servicing provider is participating in vs. out of network
Identification Method	Data Mining
Decision-Making Considerations	<ul style="list-style-type: none">• Data released to users• Targeted payers with the largest impact• Issue occurs in recent years of data• Relatively easy issue to fix if the payer has the data and time
Resolution	<ul style="list-style-type: none">• Requested resubmission of historical data from high-impact payers• Payers without capacity to resubmit files corrected the issue in future submissions• Modified data intake validation processes for earlier issue detection in the future• Timeline from issue identification to resolution implementation: 8 months

FUTURE DIRECTIONS





Future Directions for the APCD Community

- Continue efforts toward data submission and intake validation standardization (APCD-CDL™)
- Establish venue for:
 - States to share data quality policies
 - States to discuss common data quality issues
 - An APCD data user learning community

Thank You.

