

## **Breakout: Data Discrepancies, Quality Control, Universally Recognized Definitions, Data Enhancement ("Other Category")**

*Views and information are ideas that were presented by those that attended the breakout session...we didn't always agree, but these were the issues that were raised.*

*There was an important point made at the beginning of our session...that we all have the ultimate goal of improving pipeline safety and as Cynthia pointed out in her opening remarks data will help us get there. We are all focused on how data can help.*

### **Quality Control**

- Liquids – “Total # of anomalies remediated in a year”
  - Some interpret this differently as it could be what met your excavation criteria as opposed to what you ended up repairing (could be an order of magnitude higher). There are many operators documenting their internal interpretations to ensure consistency and it was discussed that operators need to look internally first for consistency, but that all must understand the context and limitations of reporting data that people have interpreted differently.
- PPTS has common interpretations/definitions. They don't only ask for “cause” but instead ask a series of questions that may indicate a wrong selection
- On distribution leaks there is inconsistency in what is being reported, and AGA has created a document on definitions for their members
- For incident reporting and incident causes...incidents don't necessarily have one cause, there is a need for more data fields to describe other additional causes
- For insertions into cast iron – The question is not asked if it is plastic or cast iron?
- There are issues with nomenclature where individual operator experiences or incidents shape their understanding of what is being sought by data collection efforts
- Field/regional employees are filling out incident forms, therefore there needs to be adequate time to train employees on new forms when released

- All stakeholders (operator, regulator, and the public) need to understand the limitations of a data field.
- As there are more and more requests for data...resources get spread thin...and sometimes this translates to more errors (transcription or otherwise)
- Regional PHMSA interpretations vs. DC interpretations

There was a common message that there is an outstanding need to understand under what context this data is being collected and what it will be used for. This will help shape the definition and further refine what it is that is collected. Much of the misunderstanding stems from a lack of direct sight to what the data will be used for.

There is a significant need for a list of common definitions/interpretations. A lot of the presentations yesterday and today looked at performance measures, but in order for trends to be meaningful we must first agree on definitions.

There is a potential to bring together a group of industry, public and regulatory representatives to discuss definitions and ensure common understanding. They would need to look through listing of already collected data and understand where existing discrepancies are. Then as new data elements are created define them as well.

Need to focus on what does the data tell us? All data, after all, should tell the same story, yet when different people analyze the data we get different results. This discrepancy when different people analyze the data boils down to differences in QA/QC and context.

### **Data Enhancements**

- One comment was made about API's PPTS and that it guides you in the right selection of cause...doesn't only ask one question.
- Pre-populating annual forms with relevant information. Use logic from last year to check
- Is \$ appropriate to determine significance?

### **Use of "other"...other causes**

- Maybe intentionally uninvestigated as you can't send every distribution leak to forensics
- Cast iron leaking and you replace it or you do an insertion you may not look for the leak source