Working Document of the NPC Study
Dynamic Delivery –
America's Evolving Oil and Natural Gas Transportation Infrastructure
Made Available December 12, 2019

Topic Paper #4-7

USE OF DATA INTEGRATION TO SUPPORT INTEGRITY ASSESSMENT

Prepared for the
Technology Advancement and Deployment Task Group

On December 12, 2019 the National Petroleum Council (NPC) in approving its report, *Dynamic Delivery – America's Evolving Oil and Natural Gas Transportation Infrastructure*, also approved the making available of certain materials used in the study process, including detailed, specific subject matter papers prepared or used by the study's Permitting, Siting, and Community Engagement for Infrastructure Development Task Group. These Topic Papers were working documents that were part of the analyses that led to development of the summary results presented in the report's Executive Summary and Chapters.

These Topic Papers represent the views and conclusions of the authors. The National Petroleum Council has not endorsed or approved the statements and conclusions contained in these documents, but approved the publication of these materials as part of the study process.

The NPC believes that these papers will be of interest to the readers of the report and will help them better understand the results. These materials are being made available in the interest of transparency.

The attached paper is one of 26 such working documents used in the study analyses. Appendix C of the final NPC report provides a complete list of the 26 Topic Papers. The full papers can be viewed and downloaded from the report section of the NPC website (www.npc.org).

This page is intentionally left blank.

Topic Paper

(Prepared for the National Petroleum Council Study on Oil and Natural Gas Transportation Infrastructure)

4-7	**Use of Data Integration to Support Integrity Assessment**
Author(s)	**David Heaney (Phillips 66 Company)** **Bryan MacKenzie (Enbridge Energy Partners)** **Thomas Bubenik (DNV GL USA)**
Reviewers	
Date: June 11, 2019	**Revision:** Final

SUMMARY

The alignment and integration of in-line and field inspection data is a critical part of determining the location and severity of potential threats. Operators use a wide range of data to provide a combined understanding of the conditions affecting a given inspection result. Performing assessments using integrated data has challenges related to the data sources, data alignment, spatial and measurement accuracy, variability in input parameters, and analysis methods. These challenges can lead to overly conservative results and unnecessary mitigations. This topic paper addresses industry advancements—including technology advancement, research, and industry best practices—in integrating data to better support asset integrity programs.

I. INTRODUCTION

The effectiveness of an integrity program relies not only on having complete, valid data sets and a sound understanding of threat mechanisms, but also effective data integration to support decision making related to integrity assessments. Accurately integrated data sets, which are normalized and aligned, can be leveraged through advanced analytics to predict threats not otherwise easily identified. This paper describes the challenges to data integration, new technologies to support data alignment, data integration opportunities, and operators' recommendations regarding the inclusion of data integration into industry practices.

II. CHALLENGES TO DATA INTEGRATION

Asset integrity engineers evaluate large data sets to find critical threats on assets such as pipelines, facility piping, and tanks. On its own, a data point identified by an in-line or field inspection is not necessarily a threat or benign to an asset. For example, an in-line inspection anomaly may be identified based on a change in gauss readings measured by an inspection tool; if there is a known magnet on the pipeline, then the risk may be minimal. However, if there is known coating disbondment issues in this area, the anomaly may indicate severe corrosion. Integrating multiple data sets to identify the actual threats to the asset can be a complex endeavor and there are many challenges that should be considered to ensure accuracy in pipeline assessments. The following factors should be considered when aligning and integrating data:

a. Disparate data sources

Many data sources exist that may provide information that can highlight a specific threat on a pipeline. Some of the sources are internal to a company while others come from external or publicly available sources. Sources of internal data can include in-line and field inspection data; as-built drawings; pipe material properties; maximum operating pressure profiles; operator-identified high-consequence areas; historical operating pressures; flange ratings of piping systems; cathodic protection readings; and close interval studies. External data sources can be more challenging to integrate into in-house data sets as operators must understand the data set in terms of how often it is updated, the general accuracy, and usable formats. Sources of external data can include the National Pipeline Mapping System[1], the National Hydrography Dataset[2], census data[3], satellite imaging, topography maps, and LiDAR surveys. Combined, these multiple internal and external data sets can be used to provide a complete understanding of the pipeline's design, location, structural condition, and operating conditions, which can be used to support holistic integrity assessment. However, these data sources may have different levels of accuracy, precision, and availability and there may be challenges with identifying and integrating all of the required detailed information for a given assessment.

b. Data Alignment

One challenge associated with data integration is data alignment, which refers to matching up different information to the same point on the pipeline. While alignment and integration of data can appear synonymous, data integration refers to the more complex development of an understanding that cannot be otherwise developed based on individual data sets. When aligning data, pipeline-specific data, such as pipe grade and wall thickness, are typically aligned to pipe centerline coordinates. Other data, such as population statistics, crossing locations, and cathodic protection readings are typically aligned to above-ground references, such as Geographic Information System coordinates. Still other data, such as operating pressures and coating condition reports, are available only at discrete sites and must be interpolated or extrapolated to estimate values away from the measurement sites.

Aligning and interpreting discrete data, above-ground data, and centerline coordinates can be challenging, but mapping technologies are improving. The need for small locational tolerances depends on the intended use of the data. For example, aligning integrity assessment data, such as the results from multiple in-line inspections, is a challenge. Here, the required uncertain-

---

[1] PHMSA, National Pipeline Mapping System, https://www.npms.phmsa.dot.gov, accessed May 27, 2019.

[2] USGS, National Hydrography Dataset, https://www.usgs.gov/core-science-systems/ngp/national-hydrography, accessed May 27, 2019.

[3] United States Census Bureau, https://www.census.gov, accessed May 27, 2019.

ty can be less than the locational tolerance of the in-line inspection tool itself, making it difficult to determine where and when changes to the pipe are occurring. Mapping technology improvements can reduce the locational uncertainties to sub-meter and sometimes sub-centimeter accuracies depending on the characteristic being investigated.

c. Spatial and Measurement Errors

Some operators axially align data based on a common coordinate system, such as the centerline coordinates of the pipeline. Errors in spatial coordinates, typically caused by odometer slippage and incorrect circumferential positioning, can be significant challenges for effective data integration. Where necessary, individual data streams are "rubber banded" along the centerline to ensure point-to-point comparisons are correct. Data can also be aligned circumferentially or errors in this positioning can be accounted for when comparing results from two or more in-line inspections.

In-line inspection data are inherently uncertain, which means the reported dimensions (e.g., depth, length, and width) of an anomaly may be higher or lower than the actual values. Operators take the potential for bias and data uncertainty into account when comparing results from multiple inspections when performing these assessments.

d. Variability in input parameters

Some of the input parameters used in assessing the severity of anomalies present along a pipeline naturally vary with time. For example, variability in operating pressure affects not only the maximum discrete point pressure along a pipeline but also the cyclic loading severity. When assessing time varying parameters, operators strive to account for maximum discrete points by making conservative (worst-case) assumptions regarding the time-varying profiles. In some cases, techniques such as rain flow cycle counting can be used to characterize the cyclic loading severity when assessing the potential for fatigue and/or environmentally assisted cracking.

Material properties also vary from pipe joint to pipe joint, even when the grade and wall thickness are the same. Operators make conservative estimates of material properties based on an understanding of the likely distribution of parameters such as yield strength and Charpy V-notch impact energy. When material properties are not known, operators are sometimes forced to make very conservative assumptions when calculating critical flaw sizes and expected remaining lives of existing anomalies.

e. Analysis Methods

Pipeline operators face uncertainties when choosing analysis methods for assessing features such as environmentally assisted cracking or mechanical damage. There is a need for a clear consensus on which analysis methods to use when dealing with, for example, low toughness welds. For most cases, such as for metal loss in pipe with reasonable toughness values, there is consensus-based guidance[4] for anomaly assessment. However, for cases when these techniques are not well defined or clearly recommended by industry best practices, operators will use conservative assumptions to ensure safety, potentially leading to overly conservative results.

f. Conservatism of Inspection, Validation, and Repair Decisions

Many of the fitness-for-purpose models or failure pressure calculations regularly employed by operators incorporate conservatism into the calculations. When employing those calculations, a safety factor will be added depending on the application. For example, a factor of 1.39 is applied to burst pressure calculations to determine the safe operating pressure or mitigation requirements and most companies use a safety factor of two when determining reinspection intervals based on anomaly time-to-failure calculations. As described in the prior challenges, operators may include additional levels of conservatism to account for uncertainty in the data sources, data alignment, spatial and measurement accuracy, input parameters, and analysis methods. The combination of all of these assumptions can lead to "stacking" of uncertainties which provide a level of conservatism beyond the original intent of the pre-defined safety factor. High levels of conservatism can result in an excessive number of features that meet actionable criteria causing a potential misallocation of resources towards mitigation in unnecessary areas. Improved accuracy of data and analysis, accurate quantification of known uncertainties, and use of calibrated safety factors or probabilistic/risk-based analysis can help to enhance the accuracy of assessments and minimize overly conservative inspection, validation, and repair decisions.

III. DATA ALIGNMENT TECHNOLOGY

Data alignment is an essential part of data integration, as accurately identifying and locating the key features of interest is core to an effective integrity program. Traditionally, data alignment has been managed through a manual and time-consuming process of analyzing in-line inspections to identify matching features across multiple inspections that were performed over years of pipeline operation. The manual approach can be prone to errors and may leave large quantities of data unprocessed, so vendors now leverage a combination of automation (using a variety of algorithms) and manual verification processes. An automated approach to data processing and feature alignment is a possible means to support improvements to pipeline integrity management and efforts have been made by in-line inspection vendors towards supporting this goal. This push towards automation is enhanced by recent technological advancements in the fields of cloud computing, data science, and machine learning. Through leveraging these advancements, hundreds of thousands of data points can be aligned and analyzed in a matter of hours.

As inspection data may be provided by multiple sources, it may come in different formats (i.e. naming conventions, units, etc.) and may require significant manual intervention to convert the data into a common environment before the feature alignment process can begin. This usually relies on classifying data fields (e.g. anomaly type, orientation, depth) based on standard input

---

[4] For example, American Society of Mechanical Engineers B31G: Manual for Determining the Remaining Strength of Corroded Pipelines.

rules to ensure the data is vendor agnostic. Machine learning classification algorithms have been developed that can be used to automatically interpret in-line inspection data, identify errors or ambiguities, and process data into a consistent and useable format, minimizing required human intervention. A further benefit of machine learning algorithms is that they can be trained based on operator specific requirements, potentially minimizing the need for complete standardization between vendors or operators.

Alignment of defects requires multiple known data points to ensure relative spatial accuracy. When automating this alignment of two (or more) inspections, algorithms often first align the girth welds then match features within the aligned joints or pipe lengths. Girth welds are often used with alignment algorithms because they provide a local common reference, with enough data frequency to allow the algorithm to account for odometer calibration discrepancies, pipe changes (repairs and re-routes), erroneous sensor values, or missing data. Once a common reference is established based on girth weld alignment, reported features such as corrosion spots, deformations, and cracks will often overlap and can be matched within each joint. The extent to which the anomalies overlap not only depends on the quality of the match but also on the characteristics of the anomaly or group of anomalies. A one-to-one, many-to-one, or one-to-many relationship may exist and can be resolved through predetermined rules. Uncertainties such as measurement or tool position errors can be accounted for to help ensure that the most probable matches are reported and matches can be reviewed and adjusted based on tolerance parameters. This type of analysis allows for matching to be performed across any number of assessments as long as they represent the same physical pipe in the ground.

With scalable, automated, and accurate alignment, it is possible to perform in-depth analysis over the entire historical set of inspections for a pipeline. Software that has access to the data from multiple pipeline operators and in-line inspection vendors can leverage machine learning to effectively identify threats and recognize patterns that could help to enhance overall integrity management.

IV. DATA INTEGRATION OPPORTUNITIES

a. Metal Loss Growth Analysis

Pipeline operators face numerous uncertainties when estimating metal loss degradation rates for remaining life calculations, which are essential for determining reassessment intervals (i.e., the time required between successive in-line inspection runs). It is generally accepted that it is difficult to match and compare two or more in-line inspections, and that uncertainties or inaccuracies in reported anomaly characteristics can make direct comparisons subject to large errors. An example of overlapping features identified by a series of six inspections over almost fifteen years are shown in Figure 1. For these cases, operators can use statistically valid methods for determining when, where, and how much growth has occurred between multiple in-line inspections. This need will become more critical as the pipeline infrastructure ages and inspection data continues to accumulate.

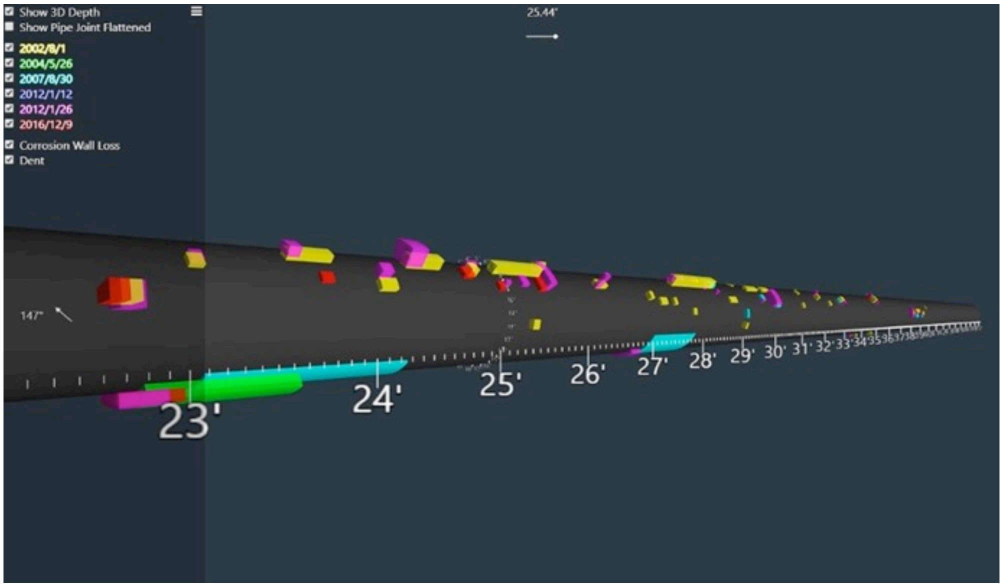

Figure 1: Three-Dimensional View of Metal Loss Comparison of In-Line Inspection Data

Outlier investigation is important as corrosion growth is typically managed well by pipeline operators, but historical failures have often been found to result from errors in in-line inspection results and/or unusually high degradation rates. Operators need to understand the conditions under which in-line inspection errors or extreme degradation rates can occur to develop suitable remediation and mitigation strategies. A better understanding of extreme values in natural corrosion and defect degradation rates is required, especially when dealing with these outliers. In addition, the pipeline industry needs to remain vigilant for new examples of in-line inspection characterization errors and/or extreme degradation rates, so they can be properly evaluated and mitigated.

b. Field Data Collection

Maintenance activities on pipelines involve documenting the work that took place, as findings from these activities can be very important in integrated data analysis. These findings provide validation of the inspection tool performance, help to identify outliers, and can provide specific condition data not necessarily available from in-line inspections alone. Collecting field information is typically performed by the project manager or the non-destructive evaluation technicians during or after the field work. The data from these field visits that is relevant to an integrity program may include defect classification, defect dimensions (such as depth, length, width, and orientation), failure pressure calculations, coating conditions, pipe-to-soil voltage readings (assessing cathodic protection effectiveness), and the start and end locations of recoating and/or repairs. Collecting data on a timely and consistent basis has its challenges as work may take place in remote areas where external communication is non-existent. Once a defect has been repaired and the dig has been back-filled, any as-found data that was not appropriately documented is no longer available.

Consistent data collection processes and procedures enable an operator to have access to relevant field information to help make sound decisions on the results. Untrustworthy data is a cause for concern in many datasets and undermines the operator's ability to validate their in-line inspection results, impairing the process[5] of providing dig feedback to in-line inspection vendors for improvement. Additionally, results that are not incorporated in a timely manner may be a lost opportunity to identify and address other time-dependent threats on a pipeline. A reliable system and/or platform to gather and store their field data and understand its limitations is important for managing data collection processes.

c. Crack Threat Data Management

Managing crack threats requires integrating multiple data sets and sources to make informed decisions on reassessment cycles, technology selections, performance assurance. The industry has incorporated projects from the American Petroleum Institute (API) and Pipeline Research Council International (PRCI) to guide decision-making in crack management. Figure 2 shows how industry published documents can be leveraged and merged with in-line inspection vendor and operator-controlled data.

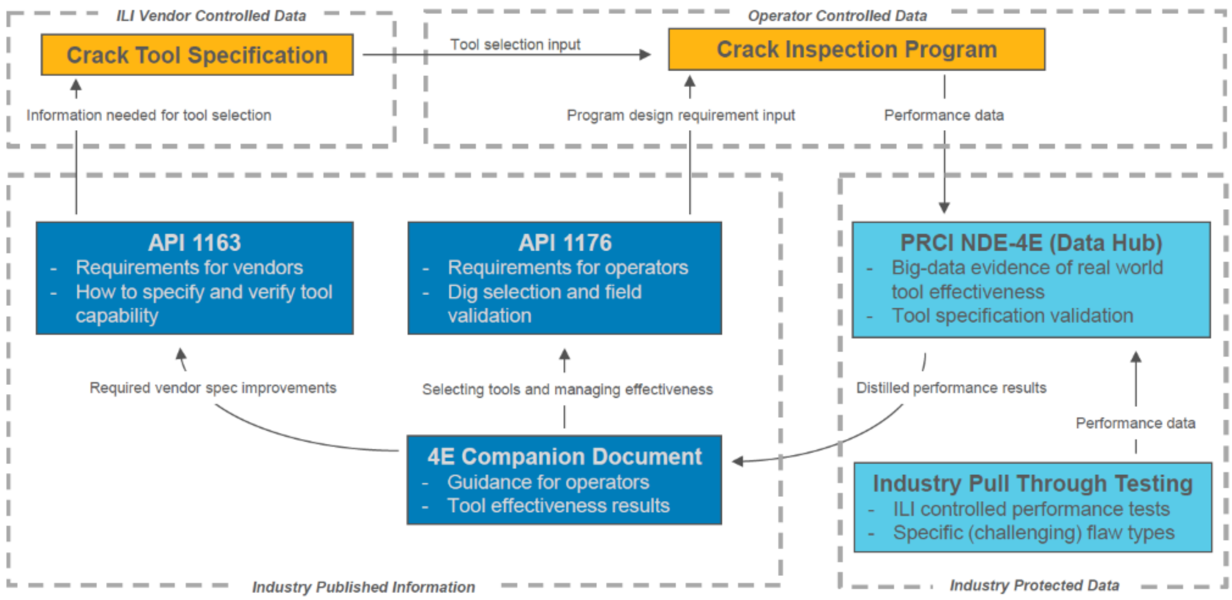

Figure 2: Crack Data Management Integration[6]

The PRCI NDE-4E[7] tests validated tool performance using thousands of in-the-ditch data points and these results are housed in a "data hub" available to PRCI members to consult. Indus-

[5] American Petroleum Institute Standard 1163: In-line Inspection Systems Qualification.

[6] Figure provided by Enbridge Pipelines.

[7] Pipeline Research Council International Project NDE-4E: ILI Crack Tool Reliability and Performance Evaluation.

try published documents like the PRCI NDE-4E companion document, API 1176[8], and API 1163[9] outline guidelines and requirements for operators and vendors alike regarding tool technology effectiveness, selection, and validation standards. Each in-line inspection technology will have specific identification, detection, and sizing specifications developed by the vendors to help operators select the appropriate tools to assess a pipeline. The specifications are developed through a series of pull tests and updated through real-world data results from operators. API 1163 outlines the requirements for in-line inspection vendors and operators to continuously provide feedback and validate performance through data sharing.

Based on the guidance outlined in the recommended practices, operators are able to integrate the data sets generated from a crack tool into their integrity program. Decisions to investigate specific threats or anomalies are made according to risk-based thresholds defined in the program. API 1176 and the NDE-4E document may provide guidance for these processes and decisions.

V. DISCUSSION AND RECOMMENDATIONS

There exist limitations to the current hazardous liquid regulations around anomaly identification and repair. Current regulations[10] do not explicitly allow advanced analytics to determine if anomalies are a threat. The industry has developed recommended practices and guidelines to help incorporate data integration into operators' processes and procedures, with the goal of improving overall integrity management.

Regulators could help to address current limitations to the adoption of advanced analytics by providing clear consensus guidance on how to compare and use data from multiple in-line inspections to estimate degradation rates and thereby establish a rational basis for reinspection intervals. For example, allowing analysis methods for anomalies in low-toughness pipe and welds that help to remove the stacking of conservatisms will enhance an operator's ability to prioritize potential defects for remediation in a practical manner which will lead to more effective and efficient integrity programs.

---

[8] American Petroleum Institute Recommended Practice 1176: Assessment and Management of Cracking in Pipelines.

[9] American Petroleum Institute Standard 1163: In-line Inspection Systems Qualification.

[10] American Petroleum Institute Technical Report 1178 – Data Management and Integration Guideline.