

Integrity Data Management and Integration

API BULLETIN 1178
FIRST EDITION, NOVEMBER 2017



AMERICAN PETROLEUM INSTITUTE

Special Notes

American Petroleum Institute (API) publications necessarily address problems of a general nature. With respect to particular circumstances, local, state, and federal laws and regulations should be reviewed.

Neither API nor any of API's employees, subcontractors, consultants, committees, or other assignees make any warranty or representation, either express or implied, with respect to the accuracy, completeness, or usefulness of the information contained herein or assume any liability or responsibility for any use, or the results of such use, of any information or process disclosed in this publication. Neither API nor any of API's employees, subcontractors, consultants, or other assignees represent that use of this publication would not infringe upon privately owned rights.

Classified areas may vary depending on the location, conditions, equipment, and substances involved in any given situation. Users of this standard should consult with the appropriate authorities having jurisdiction.

Users of this standard should not rely exclusively on the information contained in this standard. Sound business, scientific, engineering, and safety judgment should be used in employing the information contained herein. API is not undertaking to meet the duties of employers, service providers, or suppliers to warn and properly train and equip their employees, and others exposed, concerning health and safety risks and precautions, nor undertaking their obligations to comply with authorities having jurisdiction.

Information concerning safety and health risks and proper precautions with respect to particular materials and conditions should be obtained from the employer, the service provider or supplier of that material, or the safety datasheet.

API publications may be used by anyone desiring to do so. Every effort has been made by API to assure the accuracy and reliability of the data contained in them; however, the API makes no representation, warranty, or guarantee in connection with this publication and hereby expressly disclaims any liability or responsibility for loss or damage resulting from its use or for the violation of any authorities having jurisdiction with which this publication may conflict.

API publications are published to facilitate the broad availability of proven and sound engineering and operating practices. These publications are not intended to obviate the need for applying sound engineering judgment regarding when and where these publications should be utilized. The formulation and publication of API publications is not intended in any way to inhibit anyone from using any other practices.

All rights reserved. No part of this work may be reproduced, translated, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission from the publisher. Contact the Publisher, API Publishing Services, 1220 L Street, NW, Washington, DC 20005.

Copyright © 2017 American Petroleum Institute

Foreword

Nothing contained in any API publication is to be construed as granting any right, by implication or otherwise, for the manufacture, sale, or use of any method, apparatus, or product covered by letters patent. Neither should anything contained in the publication be construed as insuring anyone against liability for infringement of letters patent.

The verbal forms used to express the provisions in this document are as follows.

Shall: As used in a standard, “shall” denotes a minimum requirement in order to conform to the standard.

Should: As used in a standard, “should” denotes a recommendation or that which is advised but not required in order to conform to the standard.

May: As used in a standard, “may” denotes a course of action permissible within the limits of a standard.

Can: As used in a standard, “can” denotes a statement of possibility or capability.

This document was produced under API standardization procedures that ensure appropriate notification and participation in the developmental process and is designated as an API standard. Questions concerning the interpretation of the content of this publication or comments and questions concerning the procedures under which this publication was developed should be directed in writing to the Director of Standards, American Petroleum Institute, 1220 L Street, NW, Washington, DC 20005. Requests for permission to reproduce or translate all or any part of the material published herein should also be addressed to the director.

Generally, API standards are reviewed and revised, reaffirmed, or withdrawn at least every five years. A one-time extension of up to two years may be added to this review cycle. Status of the publication can be ascertained from the API Standards Department, telephone (202) 682-8000. A catalog of API publications and materials is published annually by API, 1220 L Street, NW, Washington, DC 20005.

Suggested revisions are invited and should be submitted to the Standards Department, API, 1220 L Street, NW, Washington, DC 20005, standards@api.org.

Contents

	Page
1	Scope 1
2	Normative References 1
3	Abbreviations 1
4	Benefits to an Enterprise Data Management System 3
5	Data Quality Oversight 4
5.1	General 4
5.2	Objectives 4
5.3	Strategies and Policies 5
5.4	Data Governance 5
5.5	Data Quality Assessment 7
6	Transforming SME Knowledge into Data 7
7	Data Models 8
8	GPS Coordinates 8
8.1	General 8
8.2	Coordinate Nomenclature 9
8.3	Datum Selection 9
8.4	Accuracy 9
8.5	Base Station Elevation 9
9	Alignment for the Purpose of Pipeline Integrity 10
9.1	General 10
9.2	Linear Referencing 10
9.3	Weld Alignment 10
9.4	Centerline 10
9.5	Absolute Referencing 10
9.6	Axial Position and Extent 11
9.7	Circumferential Position 11
10	Sources of Measurement Error 11
10.1	Spatial Error 11
10.2	Severity Error 12
10.3	Confidence Interval of the Error 12
11	Management of Change 12
11.1	General 12
11.2	New Data Relates to Old Data 12
11.3	Accessing Historical Data 13
11.4	Rerouting (Centerline Swap) 13
11.5	Moving Fit Points 13
11.6	Data Reconciliation 14
12	ILI Life Cycle 14
12.1	General 14
12.2	General Reporting Requirements 14
12.3	Data Quality Letter and Preliminary Report 14
12.4	Immediate Responses 15
12.5	ILI Final Report Format 15

Contents

	Page
12.6	Quality Assurance of Final Report before Acceptance 16
12.7	Anomaly Assessment. 16
12.8	Excavation Program 19
12.9	Provide Correlation Results to ILI Vendor 20
12.10	Program Closeout and Establishment of Reassessment Intervals 20
13	Execution of Digs/Field Data Collection. 21
13.1	Pre-dig Information. 21
13.2	In-ditch Data Collection 24
14	As-built Asset Integration 30
14.1	General Data Requirements. 30
14.2	GPS Survey 31
14.3	Data Collection 31
14.4	Virtual Pipeline Creation 32
14.5	Data Storage 32
14.6	Continuity of Linear Referencing Schema. 33
14.7	Baseline In-line Inspection 33
14.8	Baseline Indirect Assessments. 34
15	Over-the-Line Surveys (Indirect Assessments) 34
15.1	General. 34
15.2	Alignment. 34
16	Operational Data 35
17	Reporting and Data Mining 36
17.1	General. 36
17.2	Data Integration/Analysis. 36
17.3	Data Integration and Analysis Output 37
17.4	Integrity Metrics 37
	Annex A (informative) Data Integration and Interpretation Report 38
	Annex B (informative) Representative Data Listing. 50
	Bibliography 53
Figures	
1	Schematic of Relevant Excavation Site Information 26
2	Examples of Common Feature Locations 35
Tables	
1	Typical Location and Correlation Work Package Contents. 23
2	Guidelines for Photographs 27
A.1	Threat Matrix and Interpretive/QC Methodologies 40

Integrity Data Management and Integration

1 Scope

This bulletin provides a compendium of methodologies and considerations for integrating the underlying data used to support integrity management. Any one approach, let alone the entirety of the document, may not be appropriate or applicable in all circumstances. The document reviews possible approaches for consideration by operators in the context of their specific circumstances.

The primary focus of this bulletin is the methodologies and processes used to spatially integrate and normalize the data to support the application of comparative techniques used in interpreting integrity data, with particular emphasis on in-line inspection (ILI) data. The document begins with a discussion of general data-quality processes, goals, and considerations such that data quality approaches can be considered in the context of the data integration processes.

An impediment to informed integrity decisions is the inability to efficiently review a broad spectrum of data in a format that has been normalized and spatially aligned. With the variations in organizational structures, integrity management programs, and technologies used across the pipeline sector, individual operators design data integration procedures that are customized to their organizational structure, processes, and pipeline systems.

Properly managed and integrated data support agile analytics to integrate new data as they become available and to recognize coincident events and patterns. The data source may be from within an organization, or may be external to the company, as in the case of representative data based on industry experience or manufacturing processes. The intent is to empower operators to efficiently analyze and integrate threat- and integrity-related data to support their integrity management programs.

2 Normative References

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document applies (including any addenda/errata).

[API RP 1160](#), *Managing System Integrity for Hazardous Liquid Pipelines*

[API RP 1163](#), *In-Line Inspection Systems Qualification*

[API RP 1173](#), *Pipeline Safety Management Systems*

[API RP 1176](#), *Recommended Practice for Assessment and Management of Cracking in Pipelines*

3 Abbreviations

AC	alternating current
ACVG	alternating current voltage gradient
BSEE	Bureau of Safety and Environmental Enforcement
CFR	Code of Federal Regulations
CIS	close interval survey

CP	cathodic protection
DA	direct assessment
DCVG	direct current voltage gradient
DMA	discrete metal loss anomaly
DOC	depth of cover
ECDA	external corrosion direct assessment
ERF	estimated repair factor
EXT	external
FPR	failed pressure ratio
GIS	geographic information system
GPS	global positioning system
HCA	high consequence area
HDD	horizontal directional drill
ILI	in-line inspection
IMU	inertial mapping unit
INT	internal
IT	information technology
MAOP	maximum allowable operating pressure
MFL	magnetic flux leakage
MIC	microbiologically influenced corrosion
ML	metal loss
MOC	management of change
MOP	maximum operating pressure
MPI	magnetic particle inspection
MTR	mill test report
NAD27	North American Datum of 1927
NAD83	North American Datum of 1983
NDE	nondestructive examination

OD	outside diameter
POD	probability of detection
PODS	Pipeline Open Data Standard
POI	probability of identification
ROW	right-of-way
RPR	rupture pressure ratio
RTK	real time kinematic
SCC	stress corrosion cracking
SME	subject matter expert
SMYS	specified minimum yield strength
TDC	top dead center
TPD	third party damage
TQM	total quality management
UT	ultrasonic testing
WB	wrinkle bend
WGS84	World Geodetic System 1984

4 Benefits to an Enterprise Data Management System

Managing pipeline integrity data historically involved the rather manual process of populating data within spreadsheets or disparate databases. Transitioning to an enterprise database to manage large pipeline integrity data sets provides an operator with several advantages, including the following:

- Improved auditing and traceability: When spreadsheets are created, the logic and judgment that is applied while an individual is manipulating data is not captured, or easily understood. In most cases, this logic exists only in the mind of the individual who created the spreadsheet, which may result in compliance risk.
- Improved tracking of data corrections: Propagating corrections to data errors across multiple dependent spreadsheets, or back to the original data sources, is difficult and may potentially introduce further errors.
- Improved safeguards against human error: Human errors, such as versioning errors and corruption errors, can compromise the integrity of data entry. Databases and their associated graphical interfaces facilitate the implementation of quality rules and constraints that mitigate the potential for human error.
- Improved resource utilization: Databases may provide improved efficiency over data management that uses disparate spreadsheets.

- Improved data security: Server-based data may be more difficult to access and propagate than individual files, which can be easily transferred to local or portable drives.
- Improved scalability: Spreadsheets and small-scale databases may have size limitations that enterprise databases do not have.

5 Data Quality Oversight

5.1 General

The aspects of data quality listed in 5.2.3 are examples of elements an operator may consider when developing a data management system, but the extent to which they are relevant varies depending on asset complexity and organizational structure. As with any system, continuous improvement is a core principal. The system in this context is often referred to as a geographic information system (GIS) or database, but it may be a compilation of applications and databases, with a map-based visualization being just one aspect of the solution.

Data management planning considerations include the identification of key objectives to be achieved, as well as the strategies and policies that assist in achieving those objectives.

5.2 Objectives

5.2.1 Core Objective

The core objective of a data management system is to achieve the highest degree of data quality possible for the intended purpose, while also doing the following:

- promoting the efficient use of resources;
- providing easier access to critical information for qualified employees;
- ensuring that data is protected and preserved in accordance with business, legal, and policy requirements; and
- communicating with other systems using a common frame of reference for broader analysis capabilities.

5.2.2 Data Quality Criteria

Stakeholders should clearly define what is meant by data quality to make the data fit for purpose in supporting the intended processes. The data should meet the following criteria:

- Be defect free: Data should conform to the dimensions of data quality (outlined below), as applicable.
- Conform to specifications: Data should conform to specified metadata requirements such as type, length, value, precision, and units of measure.

5.2.3 Dimensions of Data Quality

In addition, a data quality definition could consider the following dimensions¹ of quality:

- Accuracy: The data represents reality.
- Completeness: All needed data is available.
- Consistency: The data is free of internal conflicts.
- Precision: The data is as exact as is needed.
- Granularity: The data is kept and presented at the right level of detail to meet the needs.
- Timeliness: The data is as current as needed and is retained until no longer needed.
- Integrity: The data is structurally sound. This connectivity is frequently referred to as topology within the geomatics community.

5.3 Strategies and Policies

Developing policies and guidelines helps create a standard set of procedures for ensuring that quality data effectively flows to and between all stakeholders.

Although data quality is a mandate for all employees, the formation of an identified team can be helpful in developing and maintaining the policies and guidelines that support standardized data quality management practices, including the following:

- Data quality audits: Data audits should occur on a periodic basis. Audits should focus on the most vulnerable/critical data. Lessons learned from data quality issues should be integrated into future audit processes.
- Data remediation processes: As audits occur, remediation prioritization should be according to the level of risk.
- Technology: Technology standards for governing data and information form part of the core system design. Data changes, such as to metadata, should be governed and executed per company procedures. While spreadsheets can be used, as the breadth of the data and the organization increases it may be easier to manage data quality and data integration by utilizing database technology or GIS, or both.

5.4 Data Governance

5.4.1 General

Data governance is the discipline of creating the vision, strategy, and structures needed to deliver an integrated and coordinated approach to improving trust in data. Effective data governance ensures that data can be trusted, and that there are processes to ensure traceability and accountability for any adverse events that result from poor data management practices. The purpose in establishing this structure is to maintain a clear focus on data quality and integration initiatives within departments and projects.

¹ Standard data quality classification – TDWI (The Data Warehousing Institute)

The core elements of governance include the following:

- defining decision rights,
- designating responsibilities,
- assigning accountabilities,
- establishing policies,
- defining processes, and
- managing performance.

Data governance should start small and grow systematically. As better predictions and forecasting are achieved, inspections and assessments can be completed where they are most needed. Consequently, the financial burden of reacting to unforeseen incidents may also be reduced. The quality of integrity decisions is directly related to the quality of the data input into the many analyses performed on the integrity data (i.e. good data supports good decisions). Similarly, bad data is worse than no data, as it may drive misinformed, non-conservative decisions.

5.4.2 Data Governance Roles

The distribution of data governance across the following proposed roles can be beneficial in clearly delineating responsibilities in larger and more geographical dispersed organizations, but their applicability will vary depending on the operator's specific circumstances.

5.4.2.1 Data Owners

Data owners are individuals who are ultimately accountable for data governance implementation and execution. Specifically, they are accountable for the data quality and governance elements that the data stewards (see 5.4.2.2) are executing, and for understanding the organizational dependencies of the data. Data owners should have a working knowledge of regulations, policies, and laws regarding the data. Data owners are typically responsible for the following:

- establishing escalation points for data governance issues;
- setting the strategy for data quality, privacy, and security; and
- determining the desired accuracy or degree of data quality.

5.4.2.2 Data Stewards

Data stewards ensure data governance initiatives are successful by overseeing data inventories and work flows. Data stewards should be proficient in data management processes, and it would be beneficial for them to have a working knowledge of pipeline integrity data and its use. Data stewards are usually responsible for execution, which would include the following:

- data definitions, usage, and quality measures;
- metadata collection and recording;
- data quality review and verification;
- data policies and procedures definitions; and

- data issues and conflicts management.

5.4.2.3 Data Custodians

Data custodians work with the data on a day-to-day basis and have the technical knowledge to profile and secure data.

5.5 Data Quality Assessment

Assessing the quality of the data should become part of a continuous improvement process. This effort is in line with the principles of total quality management (TQM), which state that the quality of information and processes is the responsibility of all those who create or consume the information or processes. Data quality assessments should include the following actions:

- identification of data quality measurements,
- identification of where and when to monitor data quality,
- implementation of monitoring processes,
- completion of a baseline assessment of the data quality and intent of usage, and
- review of post-monitoring reports.

An assessment effort should be structured and prioritized based on risk.

5.5.1 Communication Strategy

5.5.1.1 General

Continuous improvements are facilitated through a communication strategy that establishes routines for when and how important communications should be conducted and could include the aspects detailed in 5.5.1.2 and 5.5.1.3.

5.5.1.2 Data-Related Alerts and Events

Create a distribution strategy to ensure that all stakeholders have access to review and update information regarding ongoing data issues.

5.5.1.3 Data Quality Measures and Metrics

Measure data quality on established metadata created for each asset. Metrics could reflect the proportion of errors by asset on each metadata element. As with any reporting of metrics, they should be prioritized and limited in order to retain relevance.

6 Transforming SME Knowledge into Data

A significant challenge for data integration is capturing and quantifying subject matter expert (SME) knowledge. This knowledge is often lost when the SME leaves their current role within the company. Given the amount of information SMEs typically process, consideration should be given to scheduling interviews with them on a pre-defined interval can help capture the information.

Determining how to turn SME knowledge into measurable data points requires careful planning. In many cases SMEs may work in different operating areas where culture, attitude, and performance may be

defined and measured differently. Subjectivity can make comparing opinions across SMEs and operating divisions difficult. To minimize subjectivity, parameters and consistent data-gathering processes should be well defined. For example, rather than asking about the current condition of the right-of-way (ROW), operators should ask about the adequacy of the ROW clearing program in place, and should define adequacy in accordance with the company's ROW clearing policy and develop tools to facilitate location-specific knowledge.

During interviews, the applicable stakeholders should review with SMEs the pipeline alignment sheets with a photographic background for potential areas of concern. The interview should focus on issues that are known to exist on the pipeline, rather than the exact locations, which can be determined at a later date. After the interview is complete, the applicable stakeholders should determine the required pipeline integrity decisions based on the pre-defined parameters and the actions that are triggered by these parameters.

7 Data Models

Data modeling is used to assign a structure for the storage of integrity data. When designed correctly, the data model ensures that the stored data sets have structural integrity, which assists in ensuring a high degree of data quality.

There are varying approaches to data modeling, such as the following, that can be considered when designing a data model for pipeline integrity:

- The Pipeline Open Data Standard (PODS) is an example of an industry-standard data model. There are benefits to adopting this referential data model, but due consideration should be given to how the data is utilized. The PODS data model leverages a third normal form data modeling technique, which is optimized for transactional systems where many transactional data sets are inserted and updated.
- Object-relational data modeling.
- Dimensional modeling is optimized for simplified retrieval of data for the purposes of reporting and analysis. In order to improve analytical performance, consider recasting relationally modeled data into a dimensional model for this purpose.

Depending on how the integrity data sets are to be used, these various modeling techniques can be evaluated to determine the best fit.

8 GPS Coordinates

8.1 General

The measurement error for global positioning system (GPS) coordinates does not accumulate; each point is discretely resolved independent of previous points. The commonality between GPS coordinates is with the satellites used for resolving the coordinates and the correction factors applied to them, whether through real time kinematic (RTK) radio transmission or post processing.

The material in this section builds upon the 2008 API white paper, "Uniform Standards for GPS Submittals for One Call Purposes" (prepared in support of Common Ground Alliance Best Practices). One of the issues discussed within that white paper is the potential for introducing error into GPS coordinate locations. Reliance on GPS readings without a clear understanding of important settings and nomenclature options can lead to a false sense of security and introduce new and significant errors into the data. The four main sources of error are detailed in 8.2 through 8.5.

8.2 Coordinate Nomenclature

Coordinates stated in degrees and decimal minutes look similar to coordinates stated in decimal degrees. For example, the coordinates N 29 50.30 W 95 50.50 and N 29.5030 W 95.5050 are actually more than 30 miles apart. Unfortunately, many users of GPS instruments are unaware that the format of GPS readings is a critical aspect, and it is easy to confuse one format with another, or to not understand what format a GPS unit is using.

8.3 Datum Selection

A separate but related issue is the selection of “datum” (a theoretical model of the earth’s surface against which the GPS coordinates are referenced). Each datum is different, reflecting different models of the earth’s surface, ovality, etc. There are more than 100 datums that can be selected for most GPS units, although few people are familiar with what they mean or how to select the correct one, simply using whatever default datum the unit was set to when purchased. The same GPS coordinates will identify different points on the earth when using different datums. The difference is usually rather small (dozens of feet), but this is a source of error that could be significant for integrity purposes.

These potential errors could be eliminated by imposing uniform guidelines for GPS data submittal. To maintain consistency with the National Pipeline Mapping System, it was recommended that the Common Ground Alliance adopt the North American Datum of 1983 (NAD83) and use decimal degrees (with a minimum of six decimal place digits) as the standard GPS nomenclature for one-call use. The World Geodetic System 1984 (WGS84) datum is within one meter of the NAD83 datum throughout North America and should also be accepted. This same approach would also serve North America pipeline operators for their broader integrity program. Caution should be used when utilizing offshore coordinates, such as those provided on the Bureau of Safety and Environmental Enforcement (BSEE) website, as these coordinates are typically in NAD27.

8.4 Accuracy

The accuracy of GPS readings varies due to a number of factors, including the number of satellites being tracked, the precision of the GPS unit being used, etc. The accuracy information (i.e. ± 29 ft, ± 58 ft, etc.) can be obtained from typical GPS units, but in many cases this is only an estimate. Different processes require varying levels of accuracy; the important consideration is that an operator understands the measurement such that they can make a determination as to its adequacy. In the absence of a uniform level of accuracy, the measurement error should be maintained as an attribute of the data to which it pertains. Any error in the GPS coordinate propagates where the coordinate is used to align other data; accordingly, benchmark surveys for inertial mapping unit (IMU) ILI runs benefit from a high degree of accuracy. The vertical measurement error in the elevation is generally 2 to 3 $\times$ that of the horizontal measurement error for GPS coordinates.

The error associated with GPS coordinates is compounded where two separate GPS points are used to derive the relative length between these points, such as resolving the length of a pipe joint. The better practice is to acquire the coordinate for one weld and then manually measure (tight chain) the length of the joint.

8.5 Base Station Elevation

A common error in the GPS survey of a pipeline centerline is that the elevation of the base station is incorrectly entered. This usually manifests as a vertical discontinuity where surveyed data collected on different days are immediately adjacent to each other. An example of this would be at road bores where the mainline survey was completed separately from the bore installation and tie-in.

9 Alignment for the Purpose of Pipeline Integrity

9.1 General

The approaches discussed here target integrity management objectives, of which a map-based visualization is but one aspect.

9.2 Linear Referencing

Pipelines, like railways and highways, can leverage a one-dimensional frame of reference given their linear structure. Knowing the path of the pipeline, any point on the pipeline can be uniquely defined based on a measure along the line. The genesis of this linear referencing is during the construction process (construction stationing), where a common and readily apparent measure of reporting progress and the location of events along the line is required. This stationing is typically associated with construction spreads; this frame of reference is commonly replaced by as-built stationing when the line enters service. The as-built stationing is relative to a local physical structure such as a pump or compressor station, but it may include regional or governmental boundaries. This redefinition of the spatial frame of reference creates significant challenges in accurately and efficiently utilizing original construction records in operations.

An important consideration is whether the linear reference is 2D (tight chain) or 3D (slack chain). The linear measure is typically a measure from a known point of reference, but the difference between the 2D and 3D measures quickly accumulates as the distance from the reference increases.

9.3 Weld Alignment

Using the weld tally, where present, as another framework for spatial alignment provides the benefit of a fit point (location of common alignment) where the relative error can be re-zeroed every 40 ft to 80 ft along the line. All data elements should be referenced to the nearest weld aside from any other location reference, such that a highly accurate weld alignment can be leveraged at a later date even if a weld tally does not currently exist (e.g. unpiggable lines on which a direct assessment [DA] is being conducted).

9.4 Centerline

The centerline of a pipeline is typically derived in one of the following three ways:

- line-of-sight survey,
- GPS survey, or
- inertial mapping.

A line-of-sight survey is an accumulation of relative measures whereas a GPS survey is a collection of absolute coordinates.

Inertial mapping uses a combination of inertial and odometer measurements to map the pipeline in between known GPS points at regularly spaced benchmarks. The spatial error increases based on the distance from the benchmarks. As the inertial mapping is dependent on the odometer measurements, any problem with the odometer measurements impacts the delivery of the inertial centerline.

9.5 Absolute Referencing

Data elements that are not an attribute of the pipeline commonly have a location that is defined by a GPS coordinate or polygon. The GPS points can be orthogonally projected against the centerline to resolve their position along the pipeline. In the case of a polygon, an intersection or union can be performed

against the centerline to resolve the extent of interaction. In both cases, if either the centerline or external data element is moved, their relative position may need to be reassessed.

9.6 Axial Position and Extent

Regardless of the method of reference (stationing, weld offset, or GPS), attributes with a linear extent along the line are generally referenced in one of three ways:

- center and length,
- leading edge and length, or
- start and end.

The operator should understand the format used for the data deliverables received, and whether they have to be transformed to configure with the operator's data management tools, workflows, and analytics.

Another variant is where the feature can be either reported as a point or with a linear extent; valves are a common example. It is important to remain consistent in the method of reporting each type of feature (e.g. valves are all point features, sleeves are all linear features, etc.) as either point or linear features for data accuracy purposes.

9.7 Circumferential Position

Where a feature on the pipe is not fully encircling, its location around the circumference is typically referenced in one of two ways:

- center and width, or
- top edge and width.

In either case, the position can be provided as an "o'clock" orientation, or as degrees from top dead center (TDC); in both cases the convention is looking in the direction of flow (or increasing stationing in regards to bidirectional flow lines).

A special consideration for a top-edge reference is that on the back side of the pipe this reference becomes the bottom edge of the feature when standing in the ditch. This referencing convention is more prone to circumferential correlation error between in-ditch, non-destructive examination (NDE) and ILI data. Special consideration should be given to the format associated with o'clock values to ensure that this data structure is properly handled within a spreadsheet or database field (i.e. the use of a colon can create unintended issues in how the value is stored).

10 Sources of Measurement Error

10.1 Spatial Error

Spatial error can exist in a range of data inputs. ILI data, whether it is from the odometer location of a feature or from the inertial mapping data that defines the centerline, will be used here to exemplify the issue and its associated considerations. The accuracy, or conversely the measurement error, is influenced by the underlying measurement process and potentially by the relative distance from a control point as well. The spatial error may be managed by increasing the control point density (reduced), applying appropriate buffers (accounted for), or flagging the isolated occurrences where the potential for an abnormally high level of error exists (noted).

10.2 Severity Error

An ILI tool performance specification sheet shall state the tool tolerance and associated confidence interval for the depth, width, and length of an anomaly, typically based on its morphology or shape. In regards to corrosion, this is commonly expressed in terms of the Pipeline Operators Forum dimension classes (i.e. general, pitting, axial grooving, axial slotting, circumferential grooving, circumferential slotting, and pinhole). Additional limitations such as seamless pipe, girth welds, heat affected zone, bends, and fittings, can drive the measurement error beyond that stated in the specification sheet. To account for ILI tool performance error, the operator may incorporate the tool tolerance within their acceptance criteria. Increased confidence levels can be achieved through increased tolerances. Where the increased error cannot be quantified, the record can be flagged to support qualitative interpretation of the anomaly in consideration of the additional uncertainty.

10.3 Confidence Interval of the Error

Measure error is typically derived from a statistical analysis of the measurements. As such, the stated error is based on an assumed or stated level of confidence or reliability. Applying different levels of confidence to the same data set can generate significantly different measurement error ranges.

11 Management of Change

11.1 General

Management of change (MOC) is typically well defined within an operator's management system as it relates to operational controls and procedures. In regard to integrity data, the application of MOC may merit additional consideration.

11.2 New Data Relates to Old Data

The process of controlling how new data relate to historical data, in terms of representing the current threats and fitness for purpose of the pipeline, becomes increasingly important and complex as layers of data increase. The challenge is exemplified in the case of multiple ILI runs at different points in time, using different technologies, and performed on the same section of the pipeline.

Scenarios that should be addressed through documented decisions or processes and data management tools include the following:

- a) A new ILI run that completely supersedes a previous run, as it identifies the same features but with better quality.
- b) Additional features types relative to the previous run (e.g. cracks vs. metal loss) from a new ILI that would be added to the composite ILI dataset, but redundant features (e.g. weld and valves) would be filtered out to reduce duplication.
- c) A new ILI run with missing or degraded data (e.g. experienced battery failure during the run) so that only a portion of the data supersedes the previous run.
 - 1) Pending a rerun, if applicable, the historical data could be grown to create an inferred population of features to fill the gap.
- d) Excavation data that supersedes ILI data, as it is presumed to be more direct with less measurement error. This may not be the case with historical data, internal features, or crack features depending on the in-ditch assessment techniques.