

Integrity Data Management and Integration

API BULLETIN 1178
SECOND EDITION, NOVEMBER 2024



Special Notes

API publications necessarily address problems of a general nature. With respect to particular circumstances, local, state, and federal laws and regulations should be reviewed. The use of API publications is voluntary. In some cases, third parties or authorities having jurisdiction may choose to incorporate API standards by reference and may mandate compliance.

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 bulletin should not rely exclusively on the information contained in this bulletin. 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 the Institute to ensure the accuracy and reliability of the data contained in them; however, the Institute 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, 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.

Any manufacturer marking equipment or materials in conformance with the marking requirements of an API standard is solely responsible for complying with all the applicable requirements of that standard. API does not represent, warrant, or guarantee that such products do in fact conform to the applicable API standard.

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, 200 Massachusetts Avenue, NW, Suite 1100, Washington, DC 20001.

Copyright © 2024 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.

Suggested revisions are invited and should be submitted to the Standards Department, API, 200 Massachusetts Avenue NW, Suite 1100, Washington, DC 20001, standards@api.org.

Contents

1	Scope.....	1
2	Normative References	1
3	Acronyms and Abbreviations	1
4	Benefits to an Enterprise Data Management System.....	3
5	Data Quality Oversight	3
5.1	General	3
5.2	Objectives	4
5.3	Data Governance.....	5
5.4	Data Quality Determination	6
6	Transforming Subject Matter Expert Knowledge into Data	7
7	Data Models	7
8	Global Positioning System Coordinates	8
8.1	General	8
8.2	Coordinate Nomenclature	8
8.3	Datum Selection	8
8.4	Accuracy	9
8.5	Base Station Elevation	9
9	Alignment for the Purpose of Pipeline Integrity	9
9.1	General	9
9.2	Linear Referencing.....	9
9.3	Weld Alignment	10
9.4	Centerline.....	10
9.5	Absolute Referencing	10
9.6	Axial Position and Extent	10
9.7	Circumferential Position.....	11
10	Sources of Measurement Error	11
10.1	Spatial Error.....	11
10.2	Severity Error	11
10.3	Confidence Interval of the Error	12
11	Management of Change.....	12
11.1	General	12
11.2	New Data Relate to Old Data	12
11.3	Accessing Historical Data.....	12
11.4	Rerouting (Centerline Swap).....	13
11.5	Moving Fit Points	13
11.6	Data Reconciliation.....	13
12	In-line Inspection Life Cycle	14
12.1	General	14
12.2	General Reporting Requirements.....	14
12.3	Data Quality Assessment and Preliminary Report	14
12.4	Immediate Responses	15
12.5	In-line Inspection Final Report Format	15
12.6	Quality Assurance of Final Report before Acceptance.....	15
12.7	Anomaly Assessment.....	16
12.8	Excavation Program	18
12.9	Provide Correlation Results to In-line Inspection Vendor	19
12.10	Program Closeout and Establishment of Reassessment Intervals	19
13	Execution of Digs and Field Data Collection	19

13.1	Pre-dig Information	19
13.2	In-ditch Data Collection	22
14	As-built Asset Integration	26
14.1	General Data Requirements	26
14.2	Data Collection	27
14.3	Virtual Pipeline Creation.....	28
14.4	Data Storage	28
14.5	Continuity of Linear Referencing Schema	28
14.6	Baseline In-line Inspection.....	28
14.7	Baseline Indirect Assessments	29
15	Over-the-line Surveys (Indirect Assessments)	29
15.1	General	29
15.2	Alignment.....	29
16	Operational Data	31
17	Performance Metrics.....	31
Annex A (informative) Data Integration and Interpretation Report		32
Annex B (informative) Pipeline Construction Data Collection Requirements.....		54
Figures		
1	Schematic of Relevant Excavation Site Information	24
2	Examples of Common Feature Locations	30
Tables		
1	Typical Location and Correlation Work Package Contents.....	21
2	Guidelines for Photographs.....	25
A.1	Various Data Fields and Description of the Intended Contents.....	32
A.2	Threat Matrix and Interpretive/Quality Control Methodologies.....	33

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 Standard 1163, *In-Line Inspection Systems Qualification*

API Recommended Practice 1176, *Recommended Practice for Assessment and Management of Cracking in Pipelines*

API Recommended Practice 1183, *Recommended Practice for Assessment and Management of Pipeline Dents*

NACE SP0102-2017 ¹, *In-Line Inspection of Pipelines*

3 Acronyms and Abbreviations

AC alternating current

ACVG alternating current voltage gradient

¹ NACE International (now Association for Materials Protection and Performance), 15835 Park Ten Place, Houston, Texas 77084, www.amp.org.

AGM	above-ground marker
CIPS	close interval potential survey
CIS	close interval survey
CP	cathodic protection
CPS	cyber-physical systems
DA	direct assessment
DCVG	direct current voltage gradient
DMA	discrete metal loss anomaly
DOC	depth of cover
ERF	estimated repair factor
FBE	fusion-bonded epoxy
GIS	geographic information system
GPS	global positioning system
HCA	high consequence area
HDD	horizontal directional drill
ILI	in-line inspection
IMU	inertial mapping unit
MAOP	maximum allowable operating pressure
MFL	magnetic flux leakage
MIC	microbiologically influenced corrosion
ML	metal loss
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
QC	quality control
ROW	right-of-way
SCC	stress corrosion cracking
SME	subject matter expert
SMYS	specified minimum yield strength
TDC	top dead center
UPDM	Utility and Pipeline Data Model
UT	ultrasonic testing
UTM	Universal Transverse Mercator
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 datasets provides an operator with several advantages, including the following.

- Improved auditing and traceability: When spreadsheets are created, the logic and judgment that are applied while an individual is manipulating data are 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: 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 principle. 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. A common example where source data sometimes may not be fit for purpose is risk assessment. To support the analysis associated with risk assessment, a range of numerical values are typically assigned to the inputs. However, the source of these data in many cases was not structured with this use in mind. Specific considerations would include the delineation of different pieces of information into different fields or attributes (database schema design) and defining allowable content (attribute domain). A common domain constraint is the use of code lists or controlled vocabulary to avoid saying the same thing in innumerable different ways due to spelling or the use of synonyms. Care should be taken to not overly constrain data capture so as to mask new, valid and potentially critical data. The option of “other” with an associated metadata field could support the ongoing maintenance of the domain such that it can align with evolving processes.

Similarly, values that should be numeric in calculations should be stored as numeric rather than a string inside the database. In addition, long string values that are a concatenation of distinct values should be avoided, opting instead to store these in distinct fields.

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 are protected and preserved in accordance with business, legal, and policy requirements;
- 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 dimensions² of quality.

- Accuracy: The data represents reality.
- Completeness: All needed data are available.

² Standard data quality classification—TDWI (The Data Warehousing Institute).

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

5.2.4 Strategies and Policies

Developing policies and guidelines helps create a standard set of procedures for ensuring that quality data effectively flow 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, including 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 and GIS.

5.3 Data Governance

5.3.1 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.

The core elements of governance include the following:

- defining decision rights;
- designating responsibilities;
- assigning accountabilities;
- establishing policies;
- defining processes;
- 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. 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 support good decisions). Similarly, bad data are worse than no data because they may drive misinformed, nonconservative decisions.

5.3.2 Data Governance Roles

5.3.2.1 General

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.

5.3.2.2 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 5.3.2.3) 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;
- determining the desired accuracy or degree of data quality.

5.3.2.3 Data Stewards

Data stewards ensure that data governance initiatives are successful by overseeing data inventories and workflows. 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 their 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 procedure definitions;
- data issues and conflict management.

5.3.2.4 Data Custodians

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

5.4 Data Quality Determination

5.4.1 General

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, 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 determinations 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;
- review of post-monitoring reports.

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

5.4.2 Communication Strategy

5.4.2.1 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 5.4.2.2 and 5.4.2.3.

5.4.2.2 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.

5.4.2.3 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 Subject Matter Expert 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 predefined interval that 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.

7 Data Models

Data modeling is used for assigning an overarching design to guide the database schema used for the storage of data. The intent of a data model is greater data quality and consistency, while enabling the processes that are running against the data. As such, certain data models are optimized to support certain applications and vice versa.

There are varying approaches to data modeling, including, but not limited to, the following, which 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 are utilized. The PODS data model leverages a third normal form data modeling technique, which is optimized for transactional systems where many transactional datasets are inserted and updated.
- The Utility and Pipeline Data Model (UPDM) is designed primarily for the management of spatial data. Similar to PODS, this model provides a template for the management of natural gas and hazardous liquid pipe system data, including provision for gathering, transmission, and distribution systems. UPDM is extensible, allowing customized schemas to address the corporate, local, or national regulatory requirements or needs of the operator.
- Object-relational data modeling.
- Dimensional modeling is optimized for simplified retrieval of data for the purposes of reporting and analysis. To improve analytical performance, consider recasting relationally modeled data into a dimensional model for this purpose.

A consideration for data models is the spatial component. There are commonly employed relational databases that meet the GIS requirements for the management and analysis of spatial data.

Depending on how the integrity datasets are to be used, these various modeling techniques can be evaluated to determine the best fit and extended as required.

8 Global Positioning System 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 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 1 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 website, as these coordinates are typically in the North American Datum of 1927 (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 two to three 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 survey 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 in this section 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 two-dimensional (tight chain) or three-dimensional (slack chain). The linear measure is typically a measure from a known point of reference, but the difference between the two-dimensional (which ignores the pipe profile and elevation changes) and three-dimensional 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. Examples of data elements that are not an attribute of the pipeline may include items such as test stations (point), zinc ribbon (polyline), or parcel data (polygon). 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 reported either 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 regard to bidirectional flow lines). Analytics on orientation is typically easier with degrees solely based on its data format (i.e. integer from 0 to 360).

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, nondestructive 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 they are from the odometer location of a feature or from the inertial mapping data that define 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 regard 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 dataset can generate significantly different measurement error ranges.

11 Management of Change

11.1 General

Management of change 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 management of change may merit additional consideration.

11.2 New Data Relate 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 [ML]) 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.
 - 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 supersede ILI data, as they are 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.
 - Accurate, in-ditch sized features that remain on the line (recoated or under a sleeve) may become a correlation dataset for the next ILI run.
 - Applicable records shall be maintained to substantiate in-ditch measurements and associated data.

The composite dataset should be redefined in consideration of new data to support timely and informed interpretation by others in the organization or in the future. The superseded datasets are still maintained for reference; the challenge is ensuring that they are kept current with regard to their positioning along the centerline (i.e. updating old ILI alignments).

11.3 Accessing Historical Data

After a cutout is performed, timely access to historical pipe attribution and defect populations may be relevant to support the investigation of cause-and-effect relationships and inferring the extrapolation of

these attributes beyond the extent of cutouts, as applicable. At its simplest, this is the management of the failure history so as to avoid recurrence of the same conditions. Descriptions of repairs, pipe replacements, and inspections create a knowledge bank that can accurately summarize the events with specific information to aid future assessments and facilitate optimal decisions.

The mechanics of how this is achieved can vary significantly. Any underlying assumptions, constraints, or distortions should be clearly understood by all stakeholders who could access these data.

11.4 Rerouting (Centerline Swap)

A particularly challenging change to manage is the replacement of the entire centerline, or significant portion thereof, with a more accurate centerline trajectory. This scenario can arise when new IMU data from an ILI are more accurate than vintage centerline data within a GIS. Given the level of effort necessary to replace the centerline, the improvements and resulting benefits of a new centerline should be carefully weighed against alternatives. A more common scenario is the localized modification to the centerline.

If the source data are all in one frame of reference (e.g. GPS), translation to the new centerline can be managed in a uniform approach. Where the source data exist in several frames of reference (e.g. GPS, stationing, and weld alignment), greater effort and care may be required. If the source frame of reference is drawn upon to remove location inaccuracies that may have been induced through originally fitting the data against the previous, less accurate centerline, additional attention is required to ensure that the relative position of the various data remains aligned.

11.5 Moving Fit Points

Another common scenario is the potential repositioning of a reference point after data have been positioned relative to the point. An example would be where an intended benchmark location was surveyed before an ILI run. In this instance, during the course of the ILI run, the benchmark was repositioned onto the other side of the road due to interference from an overhead powerline at its original location. The updated location was not effectively communicated, and the ILI was imported and aligned within the corporate data management system based on the original position. If ILI data have already been leveraged in an integrity workflow, associations and dependencies may have been created against these ILI records. An update process to proportionately reposition all the impacted ILI records based on the revised location of the benchmark would be required. Such a “refit” of the data can be challenging given certain applications or data architectures. Where the new ILI data have been used to reposition previous ILI data through a weld match, the edits are even more difficult where the linkage across the ILI runs is not maintained to support cascading the update.

The process of responding to modified location data for fit points applies to any control point type (including valves, branch connections, etc.) that may have been incorrectly surveyed in the past. A variant of the preceding process is when the weld exposed in an excavation is not where it is anticipated to be based on information from the corporate dataset. The challenge is how to use the accurate location of the weld to correct the corporate dataset. This generally would not be resolved by simply moving the one weld, as it would corrupt the dataset by significantly distorting the adjacent joint lengths. Assuming the error accumulated over a distance, as is usually the case, the error would have to be proportionately applied up to the preexisting bounding fit points.

11.6 Data Reconciliation

When a differential analysis is conducted, such as through a data reconciliation exercise, the potential exists to identify disparities in the data. A process of resolving the hierarchy or precedence of the different datasets relative to each other is required to resolve which data are the “data of record” that flow into the subsequent integrity management analysis and decisions.

During this resolution process, an operator should consider which datasets may contain false positives, true negatives, improper identification, or outdated information. For example, magnetic ILI tools typically have high detection rates for casings. If, however, a casing is not reported by an ILI tool, an operator should not assume that the casing does not exist and remove it by default from other databases during a data reconciliation process because non-detects are common.

12 In-line Inspection Life Cycle

12.1 General

One of the core workflows within pipeline integrity is the management of the life cycle of an ILI run, extending through its integration, analysis, remediation of actionable anomalies, correlation with excavation results, and subsequent reinspection.

12.2 General Reporting Requirements

A successful ILI run depends on clear and well-documented reporting requirements that detail the underlying assumptions and clarify the expectations of the deliverables. Central to this is the pipeline questionnaire, populated by the operator. A number of questionnaire attributes are essential for determining the fitness of purpose of the ILI features and are not otherwise discernable through the ILI data directly.

Delivery timing is highly dependent upon the length of the pipeline surveyed, the number and type of features detected on that pipeline, and the type and amount of analysis or calculations performed on the data prior to delivery (e.g. anomaly burst pressure calculations, effective area profiles, overlay with other ILI tool runs, or other integrity data, etc.).

12.3 Data Quality Assessment and Preliminary Report

The data quality assessment report is issued shortly after the tool is received. This report provides the data used to determine if a rerun is required and may be available before the tool and crews demobilize from the area. These reports generally focus on the extent of coverage for the inspection in terms of percent of linear distance or pipe surface area for which the detection and sizing specifications can be met. Contributing factors could include speed excursions, debris, sensor failure, or electronics failure. Data integration considerations arise when the data are degraded but potentially usable. If data are degraded but potentially usable, then ILI datasets can be spliced together, or new metadata should be carried forward to account for location-specific uncertainties. Although these processes are not applied until later in the ILI life cycle, the decision to leverage them to accept a less-than-ideal run is made early.

The preliminary reporting of anomalies that require an immediate response can be challenging. There may be limited data to accurately locate the anomalies reported. Even where inertial mapping services have been run with the ILI, the data may not be available for early reporting. In order to leverage weld matching to position the preliminary features, where a historical weld tally exists (previous ILI or as-built survey), a complete weld tally should be requested as part of the preliminary reporting. Identification of welds by an ILI vendor is largely an automated process, and this deliverable may be available shortly after the run.

The following approaches can be taken regarding preliminary reporting.

- The entire run is assessed for defects beyond a certain depth or interaction with another feature (e.g. corrosion in dent) where the analysis can be quickly performed.
- Full reporting is requested, but for only a small portion of the line. Particularly beneficial when conducting crack reinspections, this approach lets the operator quickly develop an informed program based on the most probable sites, assuming an unmitigated growth such that correlation excavations could possibly be used in the refinement of the final report.

Additional critical features may be identified as the vendor works through full analysis and review processes (presuming the operator has instructed the vendor to provide such notifications). A protocol for reporting critical features should be developed because these features may merit a highly aggressive, limited time frame response.

12.4 Immediate Responses

12.4.1 Pressure Reduction

Pressure limiting defects present the opportunity of instituting a pressure reduction to mitigate the associated risk where their severity would merit such a response. This requires an ILI vendor or the pipeline operator to quickly calculate the burst pressure of the suspect anomalies. In some cases, such as complex anomalies where interaction or clustering of adjacent defects is applied, this level of calculation may not always be possible in a preliminary report.

Resolving the appropriate level of pressure reduction should include a review of the recent pressure history at the location of interest.

12.4.2 Leak Detection

Where the reported depth of the feature indicates a potential for a through-wall defect, the operator may assess the location for the possible presence of a leak. The specifics of the response and the required timeliness and accuracy of the feature location would depend on the product being shipped in the pipeline and regulatory obligations, but could range from over-the-line gas detection and instrumented aerial surveillance to shutting in and sectionalizing the line with pressure monitoring.

12.5 In-line Inspection Final Report Format

The ILI final report will vary by vendor, tool technology, and operator requirements. Normal components of the ILI final report include the following:

- the hard copy binder;
- the digital version of the hard copy content;
- the pipe or features tally in a digital and queryable format;
- the data/image files associated with the vendor's client viewer software.

The number of copies and the format of the first three deliverables are influenced by the operator's specification, but the underlying challenge is the management of change across these various representations of the data when the analysis is changed or updated. Different approaches can address this challenge depending on the circumstances and the operator's preferences. The approach should be documented and clearly understood, so that if versions become inconsistent all stakeholders understand which version is the trusted source.

12.6 Quality Assurance of Final Report before Acceptance

12.6.1 Initial Review

This first quality assurance process is focused on the ILI data as a stand-alone deliverable and will consist of a review of the data to ensure that the ILI tool vendor has met the general reporting requirements specified by the operator.

12.6.2 Consideration of Historical Remediation

Before resolving the anomalies for considerations as actionable from the response criteria, historical records regarding past repairs and remediation should be considered to avoid inadvertently, and unnecessarily, excavating an anomaly.

After the aforementioned reviews are complete, any errors or omissions, if present, should be resolved before accepting the final report.

12.7 Anomaly Assessment

12.7.1 General

Although the specifics vary by jurisdiction and operator, the response criteria for anomalies are a combination of the risk posed by the features and an appropriate and timely investigation of the anomaly. Upon excavation, separate criteria would be applied to the in-ditch assessment to determine if a repair is required.

12.7.2 Establish Date of Discovery

Date of discovery occurs when an operator has sufficient information to confirm the reported anomaly from the ILI data. Depending on the notification protocols agreed upon, the date of discovery can range from prior to a preliminary report to sometime after receipt of the final ILI report. Due dates for actionable anomalies are also established at this point.

12.7.3 Review for Interaction with Risk Receptors

The potential consequences of a release are central to the integrity management of a pipeline. A prevalent example of this is the high consequence area (HCA) designation in the United States. For liquid lines, this includes both direct interaction (i.e. within the HCA) and indirect interaction (i.e. overland drainage, spill plume, and downstream transport). The extent of a pipeline interacting with HCAs is not an attribute of the pipe. If the centerline (i.e. assumed location of the pipeline) or features on the centerline are moved significantly due to improved spatial alignment, then the interaction with HCAs should be redetermined as part of the management of change process.

12.7.4 Integration of Other Data Sources

12.7.4.1 Datasets

Operators should consider datasets that capture the preexisting anomaly population, pipe properties, and environmental data. The data in this context are primarily leveraged to determine susceptibility to the various threats, determine current fitness for purpose of assessment results, and resolve growth or time-dependent degradation to inform reassessment intervals. Depending on the threats being actively managed and how they are manifesting, there are numerous attributes that may be relevant in integrity management. A key consideration in the data management strategy is flexibility and agility regarding the storage, integration, and presentation of new relevant datasets as they are identified. An alignment of welds should be performed between datasets containing welds, such as ILI and excavation data. Once welds are integrated and aligned, the anomalies (such as dents, ML, or cracks) should be correlated and reviewed.

12.7.4.2 Interpretive and Analytic Techniques

Once the datasets are integrated and spatially normalized, multiple interpretative and differential analytic techniques can be applied to further assess the severity and potential causes of the reported anomalies. Possible considerations include, but are not limited to, the following.