

Aggregating and Standardizing Disjointed Integrity Management Data

Robert Floria
Lake Superior Consulting, LLC
130 West Superior Street, Suite 500
Duluth, MN 55802
USA

Alfonso Garcia
Enbridge Employee Canada Inc.
10175-101 Street
Edmonton, Alberta, T5J 0H3
Canada

ABSTRACT

Data management is a critical component of an integrity management plan. Current products and services in the integrity management sector can generate an enormous amount of uncontrolled and disjointed data, housed on multiple platforms. This text examines methods of mapping and linking multiple data sources to achieve optimal data usefulness, while reducing redundant data, through use of spatial and relational techniques. By defining relationships between fixed points, linear values can be generated from calibrated routes. Developing methods to introduce new data, standardized from spatial data, serves to maintain data quality. Recurring data transfer logistics, using relational keys in conjunction with ETL procedures, serve to link databases. Value is achieved on a large-scale using girth welds to automate the process of generating mile post values for point features. Data generated at remote sensors are aggregated from multiple vendors and populated using an exchange governed by universally unique identifiers (UUID).

Key words: database, GIS, remote monitoring, spatial join, linear referencing,

INTRODUCTION

Regulation and business needs of an organization are primary drivers for the decisions made in a pipeline integrity management data plan. How data is acquired and stored can vary greatly as determined by those factors. Two production databases and multiple remote sensor vendor databases are referenced in this text, each with a business case to operate independently.

In this example, pipeline attributes contained within the organization's corrosion control database can weigh heavily on decisions made by users of the pipe book database and those made by users of the corrosion control database. By implementing a system to quickly and accurately query attributes from

both databases, generating reports with the information required to make those business decisions becomes less complicated.

Technologies in remote sensors have an established backbone in the pipeline integrity management industry. Each remote sensor vendor holds proprietary technology and a unique data structure. Over time, advancements in technology, market trends, and product application requirements are likely to result in diversification of vendors. Regardless of vendor loyalties, technological advancements cause older models of hardware to become obsolete and newer technology to take its place. The resulting variations require special attention to ensure reporting is accurate across all vendors and models.

Implementing trusted relational and spatial data management practices are amplified as volumes of data grow. Structured query language (SQL) relational database management systems (RDMS), geographical information systems (GIS), and flat files are the platforms for those practices presented in this text.

EXPERIMENTAL PROCEDURE

Defining Entity Relationships

Defining the data and their relationships is a critical step in the data acquisition process. Figure 1 illustrates the entities used in this example and their relationship parameters. CorrosionControlDatabase is defined as the target database. CorrosionControlFeatureType_A represents a table within CorrosionControlDatabase containing records of corrosion control features along a pipeline. RemoteSensorVendor_A and PipeBookDatabase are defined as source databases. The target feature's attributes are inherited from the source databases. The feature attributes represented in all figures and models have been reduced to include only pertinent fields required for the sample model.

A one-to-one relationship between CorrosionControlDatabase and RemoteSensorVendor_A is governed using a UUID¹. In the sample model, UUIDs are generated at the target database. As remote sensors are installed and activated along the pipeline system's corrosion control features, unique records are created within the RemoteSensorVendor database. The UUIDs generated from the target database are primary keys², populated as foreign keys³ to the corresponding records in the source database.

A one-to-many relationship between CorrosionControlDatabase and PipeBookDatabase is governed by GirthWeldAddress values populated in PipeBookDatabase. The GirthWeldAddress values generated from the source database are primary keys populated as foreign keys to the corresponding records in the target database.

The section "*Spatial Joining*" in this paper explains the possibility for redundant GirthWeldAddress values populated within CorrosionControlDatabase.

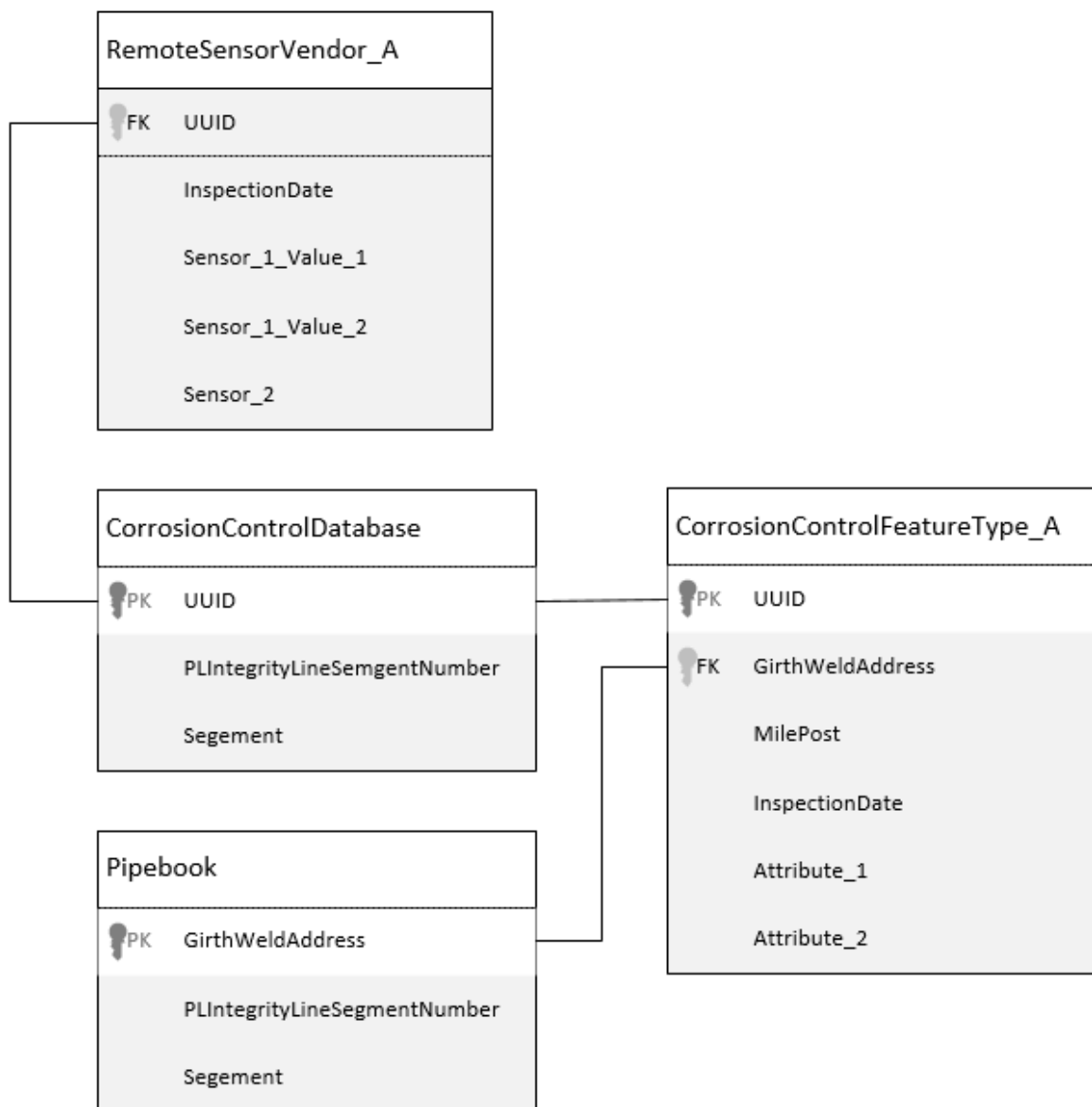


Figure 1: Global Entity Relationship Diagram