CAAP Quarterly Report

06/28/2025

Project Name: A Framework and Integrated Solution of a Dynamic Pipeline Hazard and Risk Data Repository for All Pipelines

Contract Number: 693JK32450004CAAP

Prime University: University of Dayton

Prepared By: Hui Wang, Ph.D., hwang12@udayton.edu, 937-229-2039 (University of Dayton)

Hao wang, Ph.D., hw261@soe.rutgers.edu, 848-445-2874 (Rutgers University)

Lei Wang, Ph.D., wang4li@ucmail.uc.edu, 513-556-7649 (University of Cincinnati)

Homero Castaneda, Ph.D. hcastaneda@tamu.edu, 979-458-9844 (Texas A&M University)

Reporting Period: 04/01/2025 – 06/31/2025

Project Activities for Reporting Period:

Items Completed During this Quarterly Period:

Per the proposal, progress of Task 2, Task 3, and Task 4 are associated with the third quarterly report. The following activities have been completed

Item #	*Task #*	*Activity/Deliverable/Title*
1	2,3,4	3rd Quarterly Report (this report)
2	2	Confirm the availability from public data sources for identified risk factors
3	2	Assess the data completeness and data quality of both public and private databases.
4	2	Knowledge and data-driven analysis (Appendix A)
5	3	Data standardization and harmonization (data template) (Appendix B)

Items in Progress During this Quarterly Period:

Our team has already finished Task 2 and is currently actively working on Task 3 and kicked off Task 4. Based on the progress and technical findings from Task 2, we have confirmed the public data availability and already had a good knowledge of the source data type, quality, and finished data interpretation of pipeline geological and hydrological risk assessment related publicly available data repository. At the end of Task 2, given the cleaned sample data and current existing risk model categories, we have performed knowledge- and data-driven analysis with the domain expert input from the University of Cincinnati team, Rutgers team and Texas A&M team (see Appendix A). In this quarter, our work mainly focusses on Task 3 regarding data validation and update, and data-model compatibility check. To summarize, the following items are on-going with progress to report during this quarter.

Item #	*Task #*	*Activity/Deliverable/Title*
1	3	Data validation and update
2	3	The data-model compatibility
3	4	Develop object data structure

Task #2 Objective:

The main objective of task 2 is to identify, interpret, and assess critical risk factors for the development of the proposed dynamic database.

Summary of work performed:

In Task 2, the research team worked on conducting a comprehensive analysis of the different factors contributing to different hazard induced failure pipeline systems. Following this, we studied popular risk assessment models that incorporate these identified parameters. Additionally, we examined several existing software tools that implement risk assessment models to assess the pipeline integrity. Industry practices commonly utilize proprietary risk assessment software such as PIRAMID by C-FER Technologies, TZOLKIN Risk-Based Inspection (RBI) software by the GIE group, and IMS PEI by Cenosco etc. However, these proprietary solutions are often expensive and lack transparency regarding the risk models employed. To address these limitations, our team adopted a transparent validation approach and utilized historical pipeline incident data from a PHMSA published incident report. Additionally, collaborative teams explored the application of academically developed risk models to this pipeline, with detailed findings presented in Appendix A.

Task #3 Objective:

The main objective of task 3 is to establish a database framework that will integrate the critical factors identified from Task 2 into the database.

Summary of work performed:

The database overall architecture design was completed in the previous quarter. This quarter primarily focused on the integration, standardization, and harmonization of data within the established database framework. Data integration included acquiring general risk factors from Google Earth Engine (GEE), while specific seismic factors such as Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV), Peak Ground Displacement (PGD), and fault rupture displacement were obtained using the USGS-developed software package "*gmprocess*". Employing gmprocess ensured that seismic datasets integrated into the pipeline database adhered to standardized protocols. Standardization efforts encompassed the implementation of a uniform coordinate reference system, consistent measurement units, and a structured database schema connecting milemarker, section, and segment identifiers. The defined schema provides clear standards for data management and interoperability, with comprehensive details presented in Appendix B.

Task #4 Objective:

The aim of this task is to combine the proposed data structure with data model interface.

Summary of work performed:

As the XSD (the data schema of the database element) is finalized, the next step in a Python workflow is to bridge the gap between the schema's formal structure and the native data types the application actually uses like Python in our case. We begin by generating a set of Python classes whose attributes map one-for-one to the elements and types defined in the XSD. With those classes in place, we then invoke an XSD-aware XML parser (using libraries like xsdata or xmlschema) to read each XML report, validate it against the schema, and automatically instantiate the Python classes with the parsed values. As a result, instead of wrestling with raw XML trees or untyped dictionaries, a well-typed Python objects including Coordinate, Event, SoilSurvey, etc.that guarantee every required field is present, every numeric value matches its declared type, and no unauthorized tags slip through. These ready-to-use objects can then be fed directly into downstream analysis workflows, serialized as JSON for downstream APIs, or visualized in interactive dashboards—all with complete confidence that the data faithfully conform to our rigorously harmonized data contract. This task is still ongoing and will be completed next quarter.

Project Financial Activities Incurred during the Reporting Period:

A cost breakdown list of the expenses during this quarter in each of the categories according to the budget proposal is provided below:

Prime Contract Number: 693JK32450004CAAP
Total Contract Value: $774,997.00
Total Funded Value: $774,997.00
Cost-share amount: $116,270

	Current Period Actual	Year To Date Actual	Contract To Date Actual
Salaries & Wages FT senior personnel	$15,400.99	$24,742.64	$24,742.64
Salaries & Wages Graduate Assistant	$6,000.00	$9,000.00	$9,000.00
Salaries & Wages Undergraduate student	$1,000.30	$1,000.30	$1,000.30
Benefit-Faculty/Staff	$3,901.07	$6,267.31	$6,267.31
Student Benefits -GA	$957.60	$1,436.40	$1,436.40
Student Benefits - undergraduate student	$159.65	$159.65	$159.65
Total Labor Cost	$27,419.61	$42,606.30	$42,606.30
Travel cost	$2,323.49	$2,323.49	$2,323.49
Conference Registration	$0.00	$900.00	$900.00
G A Tuition Remission	$3,565.00	$10,670.00	$10,670.00
Total Non-Labor Cost	$5,888.49	$13,893.49	$13,893.49
Total Indirect Cost	$14,931.39	$23,006.91	$23,006.91
Cost-share	-$9,869.33	-$9,869.33	-$9,869.33
Total Federal Expense (UD portion)	$38,370.88	$69,637.37	$69,637.37
Subcontract Federal Cost	$4,922.33	$4,922.33	4922.33

The full-time non-faculty labor hour cost is for the Postdoc research associate Dr. Sreelakshmi Sreeharan. The Graduate Assistantship (GA) support is for the PhD student Kiranmayee Madhusudhan. Kiranmayee is currently on the GA contract with 6 credits tuition remission in the Spring 2025 semester and 3 credits tuition remission in the Summer 2025. Her monthly stipend is $2,000 per month. Starting from Summer 2025, our team has recruited an undergraduate researcher Alexander Chattos starting from May 15th and his summer contract will last to August 15th. As Alex will enter his graduate study, a contract extension will be processed into the fall semester as a graduate assistant. Alex is currently supporting Dr. Sreeharan on Task 4. More details are shown in the project activities narrative.

The PI's research time is consolidated and charged during the summer (starting from 05/15/2025) partially using the U Dayton cost-share fund (66%) and partially using USDOT fund (34%) as planned.

During 04/06/2025 and 04/10/2025, the PI Dr. Wang and PostDoc Dr. Sreelakshmi Sreeharan attended the conference AMPP Annual Conference 2025, Nashville, TN. Dr. Wang organized the session "Digital twin". The conference cost is reflected in the cost breakdown list above.

Project Activities with Cost Share Partners:

The meetings held with the cost share partners during the current quarter are listed below, the summaries provide overviews of key discussions and decisions accomplished.

1) Weekly progress meeting scheduled with Texas A&M team:

Data and time: AMPP conference [04/06/2025-04/10/2025]; 11AM - 12 PM ET on [05/08/2025], [05/22/2025], and [06/12/2025]

Attendees: Hui Wang, Homero Castaneda, Myunghwan Jeong, Sasha George, Sreelakshmi Sreeharan, Kiranmayee Madhusudhan

Agenda discussed:

The meeting focused on refining the understanding of corrosion-related risk factors by differentiating between internal and external corrosion mechanisms and reviewing how they are represented in the risk factor sheets. This was followed by a discussion on potential risk-based models that incorporate these parameters, with the goal of building a system to validate the usage and effectiveness of the pipeline database. The team also clarified the distinctions between various SCADA systems and explored strategies for integrating risk-based decision-making into pipeline management workflows. Access to PHMSA's PIMMA software was reviewed, including its utility for retrieving detailed, location-specific pipeline data. A comparative analysis of corrosion analysis software was conducted, emphasizing common input parameters, output formats, and data structures, while acknowledging proprietary limitations. Additionally, the group discussed how data acquired from industry partners could serve as a foundation for developing a private attribute schema that better reflects real-world operational contexts. The meeting also included discussion on the feasibility of developing a review paper titled *"Risk-Based Corrosion Models: Frameworks, Applications, and Future Directions,"* which would explore existing corrosion risk modeling approaches including deterministic, probabilistic, and hybrid frameworks and highlight the integration of electrochemical factors in modern pipeline risk assessments.

Activities conducted (accomplishments):

An excel workbook was developed to systematically organize and differentiate internal and external corrosion risk factors. These spreadsheets compile key parameters such as pipe material, operating conditions, coating characteristics, environmental exposure, soil chemistry, and soil attributes. The categorized sheets are intended to serve as a foundational reference for both the development of corrosion risk models and the validation of pipeline data inputs. The team also carried out a detailed comparative analysis of two leading industry software platforms, Cenosco's IMS PEI and the GIE Group's TZOLKIN RBI, for pipeline risk assessment. The objective was to identify how each tool handles corrosion-related risk, including the types of input parameters required, risk scoring and weighting methodologies, treatment of uncertainty, and the structure and format of generated reports. In addition, a separate assessment of the input and output formats used in these tools was conducted to evaluate their interoperability with the existing pipeline database framework. Based on this evaluation, the team began drafting the structure of a private attribute schema designed to accommodate operational data shared by industry partners. This schema aims to bridge the gap between real-world field data and the parameters needed for model execution. In parallel, the team made significant progress on the review paper titled *"Risk-Based Corrosion Models: Frameworks, Applications, and Future Directions."* An initial outline was developed, dividing the paper into key sections: Fundamentals of Risk-Based Corrosion Modeling, Risk-Based Corrosion Model Frameworks, Key Input Parameters and Data Sources, Comparative Analysis of Existing Tools and Software, and Future Directions.

2) PHMSA CAAP Progress Meetings with UC and RU:

***Data and time**:* 10:00 AM – 11:00 AM ET [4/4/25], [5/16/25], [6/13/25] [4/11/2025], [4/25/2025], [5/9/2025], [5/23/2025], [6/6/2025], [6/20/2025], [6/26/2025]

***Attendees**:* Hui Wang, Lei Wang, Hao Wang, Yating Yang, Jay Shah, Sreelakshmi Sreeharan, Kiranmayee Madhusudhan

***Agenda discussed**:*

Joint meetings involving the University of Dayton, the University of Cincinnati (UC) and Rutgers University (RU) teams were held to address the interconnected aspects of geohazard and hydrological risk modeling for pipeline integrity assessment. The team discussed the overall progress of the pipeline risk assessment implementations, including database development, presentation preparation, and sourcing data for hydrological and geological factors. Possible pipeline segments were evaluated for testing and validation, and the feasibility of downloading historical incident data from the National Pipeline Mapping System (NPMS) using county filters was explored. Strategies to pinpoint incident locations using operator IDs, counties, and report numbers were also reviewed. The team compared the NPMS public and detailed viewers, noting that the detailed version includes additional attributes. The relevance of "quality" in the context of pipeline materials was discussed, along with the possibility of querying incident records in the public viewer. Challenges related to modeling corrosion damage and flood risk were acknowledged, and a simplified approach focusing on identifying high-risk regions rather than calculating precise probabilities was agreed upon. The team also proposed creating soil erosion and landslide susceptibility maps for validation against known incident sites and studying specific PHMSA reported pipeline failures to map them within the NPMS viewer and highlight high-risk areas.

***Activities conducted (accomplishments)**:*

Both teams agreed to use the PHMSA incident report involving the Gulf South Pipeline Company's Index 381 pipeline in Jackson, Mississippi as a case study for implementing and validating the hydrological and geohazard risk assessment model. The University of Dayton (UD) team focused on accessing and organizing relevant geohazard and hydrological factors shared by the teams. The UD team also evaluated data retrieval mechanisms from the NPMS. Efforts were made to cross-reference incident locations with geological conditions. The teams reviewed the feasibility of replicating previous academic risk models to better align their framework with practical field applications and discussed the integration of conditional logic for site-specific risk factors into the database schema. A major setback encountered during this phase was the limited access to industry-practice risk assessment models, which are proprietary in nature and lack transparency.

<u>3) Review Paper Planning Meeting: Database development</u>

***Data and time**:* 03:00 PM – 04:00 PM ET [6/3/25]

***Attendees**:* Sreelakshmi Sreeharan, Yating Yang, Sasha George, Jay Shah

***Agenda discussed**:*

A meeting was held to discuss the planning and scope of a collaborative review paper focused on pipeline database development. The team explored potential themes to be covered, discussed the overall structure of the paper, and identified possible figures that could effectively illustrate key concepts in a review format. Responsibilities for individual sections were assigned to team members, and a preliminary timeline was established for drafting, internal review, and submission.

***Activities conducted (accomplishments)**:*

As a result of the meeting, the team finalized the central theme of the review paper and agreed on a working title. An initial outline was drafted, dividing the paper into major sections such as hazard identification, risk models, visualization methods and gaps in current practices. A shared document was created to facilitate collaboration, and a preliminary draft is done.

Project Activities with External Partners:

In this quarter, collaborative activities with external partners include the quarterly review meeting on April 22. In this meeting, PI Dr. Wang provided an update on the project's progress, including the completion of Task 1: literature review, the publication of a conference paper, and the ongoing work with three university teams. The postdoc Sreelakshmi presented the design of the database, emphasizing its user-friendliness, efficiency, and compatibility with existing data management systems. Together with industry partners: ROSEN, Integrated Solutions Field Services, and API, the team discussed the challenges of dealing with real-world data, the structure of a root data frame, the segment-wise data frame for each pipeline segment, and the process of connecting and extracting data from various sources.

The UD team lead by the PI Dr. Wang also requested several technical helps and guidance from PHMSA including current geohazard and hydrological risk assessment practices and possible access to the National Pipeline Mapping System (NPMS). The program director Dr. Akter followed up and inquired with field personnel about standard hydrological risk assessment models used by operators. Dr. Akter also discussed with the previous Technical Task Inspector (TTI) Ben Kendrick and circle back to field offices for additional answers. We have collected the compiled information from Ben and forwarded the information to our Rutgers team on hydrological risk assessment data-model interaction. Dr. Akter also suggested the project team to reach out to FEMA for information on infrastructure risk assessment related to pipeline safety. Regarding the NPMS access, Dr. Akter sponsored the UD team on NPMS access and filled out the necessary forms. The UD team has signed the non-disclosure agreement with PHMSA on the Pipeline Information Management Mapping Application (PIMMA) account set up.

Potential Project Risks:

The main technical activities on confirming data availability, database framework design, and prototype development are conducted as expected as the project technical side is moving forward so far, no potential risk on feasibility is noticed at this stage.

Currently, our project supports two postdocs, three PhD students, one master student, and one undergraduate student. The current manpower seems to be good for performing investigations in Task 2 and Task 3. For task 4 we are expecting more intense coding, testing, and iterative refinement. The UD team is expecting recruit one more PhD student in the coming fall semester to expedite the database prototype development. The PI Dr. Wang is actively searching a suitable student to fill in this vacancy in order to avoid possible project delay in Task 4.

On the subcontracts side, as mentioned in our last quarterly report, we observed some delay on the subcontracts set up as the agreements' negotiation and corresponding paperwork process for kicking-off subcontracts costed longer than expected. This delay in this quarter was also reflected on the subcontract invoicing progress to date. The UD post-award team is tracking the subaward expenditure progress and reporting to PI Dr. Wang at a monthly basis.

Future Project Work:

Over the next 30 days, the primary focus will be on finalizing the automation of data ingestion scripts from both Google Earth Engine (GEE) and USGS sources to ensure consistent, scalable, and reproducible extraction of critical risk factors. In parallel, data validation interface routines will be developed based on risk model inputs provided by partner universities to ensure data accuracy and relevance. In the following 60 days, efforts will concentrate on refining specific criteria within the database and implementing conditional activation of site-specific factors. For instance, parameters such as liquefaction-induced settlement are only applicable under certain seismic and geotechnical conditions—specifically when pipelines are located in earthquake-prone regions with sandy soils. Although the XML-based structure accommodates all possible factors, conditional logic interdependencies will be integrated into the data entry framework so that only relevant fields are activated and populated in order to avoid accident wrong input. This will support context-aware, efficient, and accurate data usage. Additionally, role-based access controls will be established to maintain data integrity and security. By the 90-day mark, work will advance toward establishing automated data update mechanisms and developing a web-based visualization interface. This dashboard will enable stakeholders to interact with, analyze, and interpret risk data across pipeline segments, supporting more informed decision-making and predictive risk assessment.

Potential Impacts to Pipeline Safety:

In this quarter, we have presented one paper in AMPP 2025 Annual Conference and chaired one session focusing on pipeline digital twin. We also have one paper accepted for oral presentation at ASCE UESI pipelines 2025 Conference at Tampa, Florida. The published conference papers are listed below:

1) ***Sreeharan, S.***, **Wang, H.***, Castaneda, H., Wang, L., Wang, H., Development of dynamic database for proactive and predictive risk management of underground pipelines: a comprehensive review, ASCE Pipelines Conference 2025, Tampa, KL.
2) **Wang, H.***, ***Sreeharan, S.***, Castaneda, H., Indirect inspection-based Bayesian machine learning model for probabilistic coating status and severity interpretation. In AMPP Annual conference 2025, Nashville, TN.

Appendix A:

Knowledge and Data driven analysis

The UC team worked on implementing a semi-quantitative risk model based on the inform of a specific pipeline failure event that occurred on February 12, 2023, involving Gulf South Pipeline Company's Index 381 pipeline in Jackson, Mississippi. This failure resulted from natural forces, specifically a slow-moving landslide triggered by nearby earth movement related to a man-made dam.

Implementation of risk assessment model

a) Model selection

Based on the nature of the failure and the availability of historical and inspection data, a semi-quantitative, scoring-based risk assessment model was selected. This approach integrates qualitative engineering judgment with quantitative field data and is well-suited for this geohazard-prone pipeline. The principles from DNV-RP-F116 and API RP 581 inspire the framework.

b) Overview of framework

The risk of failure due to geohazards is calculated as follows:

i. Likelihood of Failure (LoF): Represented by hazard susceptibility, pipeline vulnerability, and detection capacity.

ii. Consequence of Failure (CoF): Represented by exposure impacts and environmental/social consequences.

Each factor is scored on a 0-5 scale and weighted according to its relative contribution to overall risk.

c) Assessment categories and indicators

The model is built on five primary risk categories, each of which includes key indicators informed by industry practice and field evidence. In particular, the selected indicators are aligned with recommended geohazard-related parameters from API RP 1187, which provides a comprehensive framework for assessing the integrity threats posed by landslides and other earth movements, as shown in

Table 1.

Table 1. Risk Assessment Categories and Geohazard Indicators (Aligned with API RP 1187)

Category	Description	Indicators Considered
Hazard Susceptibility	Likelihood and scale of geohazard activity near the pipeline	Slope angle and terrain, soil and rock type, groundwater conditions, pore pressure, landslide type and depth, movement rate, triggering mechanisms (e.g., rainfall, excavation), and historical slope failures
Pipeline Vulnerability	Structural and design-related resistance to external loads	Pipe material and grade, girth weld quality, burial depth, pipeline orientation
Monitoring & Detection	Capacity to detect geohazard indicators before failure	Patrol frequency, ILI and IMU strain history, slope monitoring data, InSAR deformation analysis, slope

		inclinometer data
Protective Measures	Measures implemented to mitigate or control geohazard effects	Slope remediation history, prior repairs, surface drainage control
Exposure & Consequence	Potential impact to public safety, infrastructure, or service continuity	Population proximity, service criticality, historical disruption, location class

These indicators help ensure that the model not only reflects the conditions specific to the selected case but also adheres to the broader best practices outlined in API RP 1187 for geohazard-prone pipeline integrity management.

d) Scoring and weighting

Each factor is scored from 0 (negligible risk) to 5 (high risk), based on available evidence from the PHMSA report. The scores are then multiplied by a predefined weight (%) to reflect their importance. 0 = None / negligible, 1 = low; 2-3 = Moderate; 4-5 = High / Severe. As Table 2 shown, the following weights were used:

Table 2. Risk Category Scores and Justifications for the selected pipeline

Category	Score	Justification Summary
Hazard Susceptibility	5	Active landslide, slope instability since 2003
Pipeline Vulnerability	4	Incomplete weld, old pipe, stress damage
Monitoring & Detection	2	Patrols missed hazard, ILI underutilized
Protective Measures	1	Repeated failed remediation
Exposure & Consequence	2	Service disruption, gas release, no injuries

e) Results for scoring

To define the weight of each risk factor, a risk-informed approach was used. The relative importance of each factor was assessed based on its direct contribution to the failure mechanism observed in this case, as well as general practices outlined in standards such as DNV-RP-F116 and API 581. As the table concluded, Hazard susceptibility was assigned the highest weight (30%) due to the primary role of landslide-induced soil movement in triggering failure. Pipeline vulnerability followed with 25%, reflecting the significance of structural weaknesses, specifically, the girth weld defect. Monitoring and detection (20%) were moderately weighted due to their underutilization before the event. Protective measures were given a weight of 15%, as repeated but ineffective remediation attempts were documented. Lastly, exposure and consequence were weighted at 10%, recognizing the moderate service disruption caused by the absence of injury. This weighting scheme ensures the model reflects both technical relevance and incident-specific context.

Table 3 provides the detailed summary.

Table 3. Weighted Risk Scoring Summary

Category	Score (0-5)	Description	Weighted score
Hazard Susceptibility	5	30	30
Pipeline Vulnerability	4	25	20
Monitoring & Detection	2	20	8
Protective Measures	1	15	3
Exposure & Consequence	2	10	4
Total		100	65

f) Risk categorization
Based on cumulative scores, risk levels are categorized as follows. Low Risk: 0–30; Medium Risk: 31–60; High Risk: >60.With a total score of 65, this pipeline segment is classified as High Risk, confirming the segment's vulnerability to geohazard-induced failure.

RU team focused on hydrological risk model implementation and data preparation. The following discussion presents insights from studies that examine pipeline failure in the context of hydrotechnical hazards. Two categories of approaches have been identified: 1) Indirect approach, which focuses on factors that lead to pipeline exposure to hydrodynamic forces, such as soil erosion; 2) Direct approach, which investigates the impact of hydrodynamic forces on already exposed pipelines. Although the primary focus is on gas pipelines, studies involving other types, such as water distribution pipelines, are also included. This broader inclusion helps expand the scope for applying transferable methods to estimate pipeline failure using remote sensing data.

1. Indirect Approach

Shi et al. (2020) used sensor-equipped pipeline data including pipe wall strain, soil pressure, and temperature, combined with weather trends such as precipitation, temperature, and freezing index, to predict pipe deformation. They applied a feature selection method along with a super learning algorithm to correlate various weather trends with different types of deformation. The study found that soil water content has a critical impact on tensile and bending deformations of the pipe, while rainfall patterns were ranked highest for influencing circumferential deformation.

Karthikeyan et al. (2020) examined the influence of twelve parameters, such as pipe age, exposure flow rate, river depth, and pipe material, on the scour depth and critical span length of oil and gas pipelines buried beneath riverbeds. They analyzed pipeline failure data from the PHMSA accident database, focusing on cases involving vortex induced vibrations. Among all parameters, the discharge rate was identified as the most critical factor in predicting scour depth. In comparison, reduced velocity for cross flow oscillation, calculated using Reynolds number, and depth discharge rate were found to be the two most influential factors in predicting critical span length.

Cohen et al. (1979) explored the impact of winter temperatures on frost depth, which can expose buried pipelines. This exposure is particularly important because it may lead to alternating pressure forces due to the frost heaving phenomenon. In regions affected by soil erosion, forecasting frost depth becomes a high-priority task and can be improved with the use of relevant environmental data. A general relationship between the freezing index and the duration of subfreezing temperatures is summarized in eq. 1,

$$X = A + B\sqrt{F} \tag{1}$$

where X is frost depth in meters and F is the freezing index in days.

Silva (2004) investigated the geographical distribution of annual rainfall-induced soil erosivity across Brazil. Using pluviometric records from 1,600 weather stations, the study applied an adaptive soil erosivity equation (eq. 2) to develop detailed soil erosivity maps. This approach enabled the identification of regions with high soil erosivity, specific to different times of the year.

$$A = R.K.L.S.C.P \tag{2}$$

where A represents the rate of soil loss, while R, K, L, S, C, and P are factors corresponding to annual rainfall erosivity, soil erodibility, slope length, slope steepness, cover management, and supporting conservation practices, respectively. The R factor has units of tons per hectare per year (t/ha·yr), K is expressed in megajoules millimeters per hectare per hour per year (MJ·mm/ha·h·yr), and the remaining factors are dimensionless. In addition to this formulation, other studies (Martins et al., 2020; Almedia et al., 2012) have estimated soil erosivity using the Fournier Index, as represented in the following eq. 3,

$$Cc = \frac{M_x^2}{P} \tag{3}$$

where M denotes the monthly precipitation (in mm) for month x, while P represents the total annual precipitation (in mm). Various studies have established correlations between monthly soil erosivity and the Fournier Index, which are represented in eq. 4 and 5,

$$R_x = 3.76\,Cc + 42.77 \tag{4}$$

$$R_x = 36.849(Cc)^2 \tag{5}$$

Winning et al. (2014) used a similar formulation to pinpoint areas near buried pipelines that are prone to high soil erosivity. When local rainfall data were missing, they assumed that 90 percent of the wettest month’s precipitation could occur in a single day as a worst-case scenario. Rainfall intensity and erosivity were then estimated by combining standard intensity values such as 75 mm h^{-1} with climate-based percentages to calculate the rainfall erosivity factor.

2. Direct approach

A direct approach to utilizing remote datasets focuses on pipelines that are already exposed to hydrodynamic forces. These studies typically involve mechanical models that predict localized strains or deformations under hydrological hazard scenarios. Rossi et al. (2022) proposed a vulnerability model to evaluate the probability of pipeline failure during flood events. Several factors were considered, including water elevation, water velocity, hydraulic head, flood frequency, water density, and dynamic pressure. Among these, flood frequency and pipe length were identified as the most critical factors influencing failure probability.

Building on this concept, Caratozzolo et al. (2022) investigated floodwater-induced displacement and floating-related damage to cylindrical infrastructure. They analyzed various NaTech (Natural Hazard Triggering Technological Disasters) scenarios based on floodwater height and velocity. Similarly, Landucci et al. (2014) developed a quantitative vulnerability model for pipeline failure probability, accounting for both the severity of the natural event and the construction characteristics of the infrastructure. Structural integrity loss was evaluated by determining whether the critical pressure (as defined in eq. 6) exceeded the net external pressure acting on the structure.

$$P_{cr} = \frac{2Et}{D}\left\{\frac{1}{(n^2-1)\left[1+\left(1+\left(\frac{2nH}{\pi D}\right)^2\right)^2\right]} + \frac{t^2}{3(1-\nu^2)D^2}\left[n^2-1+\frac{2n^2-1-\nu}{\left(\frac{2nH}{\pi D}\right)^2-1}\right]\right\} \quad (6)$$

where E is the Elastic modulus and ν is the Poisson's ratio, D is the diameter, t is the vessel thickness, H is the height, and n is the integer number defined in eq. 7,

$$n \geq \left(\frac{\pi}{2}\right)\left(\frac{D}{H}\right) \; for \; n \geq 2 \quad (7)$$

The above formulation was validated against literature data confirming the impact of NaTech scenarios.

The Texas A&M team studied two existing software's for their input and output parameters used to determine corrosion related risk.

1. **Cenosco (IMS PEI)**: This is comprehensive software with advanced corrosion risk-based assessment, It supports various deployment options and provides extensive training resources.
    a. **RBI Methodology**: IMS PEI utilizes a semi-quantitative Risk-Based Inspection approach, combining qualitative assessments with quantitative data to evaluate the probability and consequence of failure due to external corrosion. This methodology aids in prioritizing inspection and maintenance activities based on risk levels.
    b. **Corrosion Management Methodology**: The software provides a structured framework to identify and manage various degradation mechanisms, including external corrosion. It however, <u>*doesn't cover internal corrosion factors*</u>
    c. **Advanced Corrosion Predictions**: it incorporates advanced corrosion prediction tools that use historical data and environmental parameters to predict corrosion rates. This predictive capability helps with pipeline inspection scheduling.
    d. **Inspection Planning and Optimization**: By calculating the Remaining Life of equipment based on corrosion rates and minimum allowable thickness, IMS PEI assists in determining optimal inspection intervals.
    e. **Integration with Inspection Data**: This software allows for the integration of inspection findings, such as wall thickness measurements and coating assessments, to refine corrosion rate calculations and risk evaluations. This integration enhances the accuracy of risk assessments and inspection planning.

    The main parameters of this software used for the risk-based model are
    - Soil Characteristics such as soil resistivity and pH levels

- Humidity
- Temperature
- The presence of corrosive agents like chlorides and sulfur dioxide are considered in atmospheric corrosion assessments
- Cathodic Protection Data: The software incorporates data from cathodic protection systems, such as Close Interval Potential Surveys (CIPS).

2. **TZOLKIN**: This was the second software that was studied. This software focuses on RBI aligned with API standards, emphasizing damage mechanism identification and fitness-for-service evaluations.
    a. **RBI Methodology:** This software uses a RBI based methodology to assess the integrity of the pipeline systems. It integrated both the probability and consequence of failure to determine risk levels. This systematic and data driven approach helps optimize the long-term reliability of the pipeline.
    b. **Corrosion Management Methodology**: This software uses historical inspection data, fluid characteristics and environmental conditions such as exposure to different factors to estimated *internal and external corrosion rates*.
    c. **Degradation mechanism identification:** This software identifies the different degradation mechanisms such MIC related corrosion and carbon dioxide related corrosion to determine the pipeline integrity in the specific operating environment.
    d. **Dynamic Data Integration:** This software supports the use of continuous updated data from the new inspection as well as steady monitoring data in the risk-based model. This approach provides a dynamic corrosion prediction which helps with the relevant employment of corrosion management strategies.
    e. **Inspection Planning:** Based on the determined risk, this software generates optimized inspection plans to accurately predict corrosion and employ strategic mechanism for protections.

The main parameters of this software used for the risk-based model are:

- Pipeline design and parameter such as wall thickness, diameter
- Operating pressure and temperature
- Soil conditions such as moisture content, bacteria
- Flow rate and flow regime
- Cathodic protection strategies
- Internal corrosion factors such as pH, bacteria
- Inspection monitoring data

Table 4: Comparing the scope of both software's

	Cenosco	**TZOLKIN**
RBI Methodology	Utilizes Shell's RBI methodology, compliant with API 580 and API 581 standards	Implements API 580 and API 581 standards
Corrosion Management	Offers advanced corrosion predictions and a comprehensive corrosion	Identifies damage mechanisms per API 571; supports corrosion rate

	management framework	calculations and fitness-for-service assessments
Inspection Planning	Facilitates inspection scheduling with dynamic forms and integrates RBI results with inspection data	Provides risk-based inspection planning with optimized intervals based on risk assessments

Appendix B:

<u>Data standardization and harmonization</u>

The developed pipeline database integrates diverse data types essential for comprehensive pipeline integrity assessment, encompassing geohazard, hydrological, and electrochemical factors. The compilation and harmonization of these varied data streams require meticulous standardization and systematic integration. Primary data sources include Google Earth Engine (GEE) and the United States Geological Survey (USGS). Google Earth Engine serves as a robust repository for publicly accessible satellite-derived geospatial datasets, effectively covering a wide array of risk factors, such as soil moisture, land subsidence, precipitation, flooding indices, and surface runoff characteristics. However, GEE does not adequately encompass detailed seismic and geophysical event-specific data, which are critical components for precise pipeline risk analyses.

To address this gap, publicly available USGS databases are integrated to supplement the pipeline database, providing authoritative seismic event data and detailed geophysical parameters. USGS datasets include highly accurate seismic parameters like Peak Ground Acceleration (PGA), Peak Ground Velocity (PGV), Peak Ground Displacement (PGD), and spectral accelerations. The *"gmprocess"* software toolkit, a specialized Python package developed by the USGS, is utilized specifically to process seismic ground-motion data systematically. It automates the retrieval, processing, and standardization of seismic records from authoritative sources. Through *gmprocess*, critical parameters such as PGA, PGV, PGD, and spectral acceleration are computed accurately and consistently. The integration of *gmprocess* ensures that the seismic datasets incorporated into the pipeline database meet standardized protocols, thereby significantly enhancing the quality and reliability of subsequent integrity analyses and predictive modeling.

Our pipeline integrity system begins by ingesting a broad spectrum of data satellite-derived geospatial layers from Google Earth Engine (GEE), authoritative seismic event and ground-motion records from the USGS, and high-resolution field survey measurements. To bring these heterogeneous sources into a single coherent framework, we enforce a unified spatial reference: every coordinate is reprojected to EPSG:3857 (Web Mercator) immediately upon ingestion. This reprojection guarantees that when we compute distances or overlay soil-moisture rasters with pipeline alignments, all data align perfectly, with no distortion or axis-ordering ambiguities.

A standardized database schema was established using uniform identifiers, including milemarker ID, section ID, and segment ID, facilitating seamless integration and retrieval. By defining a comprehensive XML Schema (using XSD file) for our combined geospatial and seismic report, we create a single source of truth for every piece of data. Temporal harmonization is equally rigorous. Both our public and private XML schemas begin with a Dates section capturing an ISO-8601 RetrievalDate timestamp. In the public data schema, we also record QueryStartDate and QueryEndDate to document the exact window used for GEE reducers and USGS event searches. Embedding these timestamps gives every downstream user or automated audit process a precise provenance chain, so that analysts can replicate the exact data pull or understand the freshness of the flood-frequency or seismic-hazard inputs.

To facilitate seamless integration across data domains, we standardized on a minimal identifier backbone: SectionID, SegmentID, and MileageID. These three string fields appear at the start of every Coordinate record in both our “public attributes” and “private attributes” schemas. By enforcing identical naming, typing, and ordering, we eliminate any risk of downstream joins failing due to mismatched key names or unexpected ordering. Whether you’re correlating peak-ground-acceleration values with soil resistivity logs or combining NDVI trends with integrity survey results, you always have the same three fields in the same order to align on.

Quantitative measurements are normalized into SI-based units at the point of data capture or immediately upon retrieval. Elevations and depths are stored in meters (m), velocities in m/s, accelerations in m/s^2, soil resistivity in ohm-meters ($\Omega \cdot m$), and volumetric moisture or coarse-fraction percentages in Vol % etc. We apply these conversions automatically GEE’s pixel values are multiplied or divided to match our unit convention. USGS ShakeMap grid values (which may originally be in g, gravity acceleration units or raw counts) are transformed into m/s^2 or m. This removes any semantic ambiguity and ensures that every field in every record can be directly compared or mathematically combined without on-the-fly unit conversions.

The backbone of our validation and harmonization is the pair of XML Schemas (XSDs) we authored. Each schema declares every data element exactly once no synonyms, no legacy aliases assigning it a precise XSD type (xs:decimal, xs:boolean, xs:string, or xs:dateTime). We embed xs:documentation annotations on every quantitative element to specify its unit (“m/s^2” for PGA, “kPa” for soil strength etc). Although standard XML validators don’t enforce unit correctness, these annotations serve as machine-readable metadata that downstream converters or schema-aware applications can parse to verify, convert, or annotate units automatically, greatly reducing risk of misinterpretation as unit and type confusion is a common reason of mistakes in scientific computing for engineering risk assessment.

Structural consistency is enforced through strict sequencing and cardinality. In the public schema, each Coordinate must contain identifiers, then GeoData bands (elevation, slope, NDVI, etc.), followed by a nested Events section. In the private schema, the same identifiers are followed by Crossing, then nested HighVoltageLine and PreviousFailure blocks, and finally SoilSurvey and IntegritySurvey groups. By declaring these child sequences with xs:sequence and leveraging minOccurs="0" for optional fields (e.g., liquefaction indices or distance to high-voltage lines), we allow variability only where appropriate, while guaranteeing that any XML parser can rely on a fixed tree shape.

Finally, every XML instance document includes the xsi:noNamespaceSchemaLocation attribute pointing to its XSD, binding the document to the schema. This binding ensures that any deviation missing required elements, extra tags, typographical errors will be caught by a standard XML validator before the data ever reach analytical pipelines or GIS import routines. The result is a harmonized, self-validating data exchange framework: all producers “speak” the same schema, all consumers import against the same contract, and every data field from flood-inundation depths to soil corrosivity flags is unambiguously defined, typed, and annotated. The root structure of both schemas is illustrated in Figure 1, showing how each document begins with a Dates block followed by a sequence of Coordinate entries. For each <Coordinate>, the first three child elements SectionID, SegmentID, and MileageID serve as the canonical key triple that unifies records across all data sources.

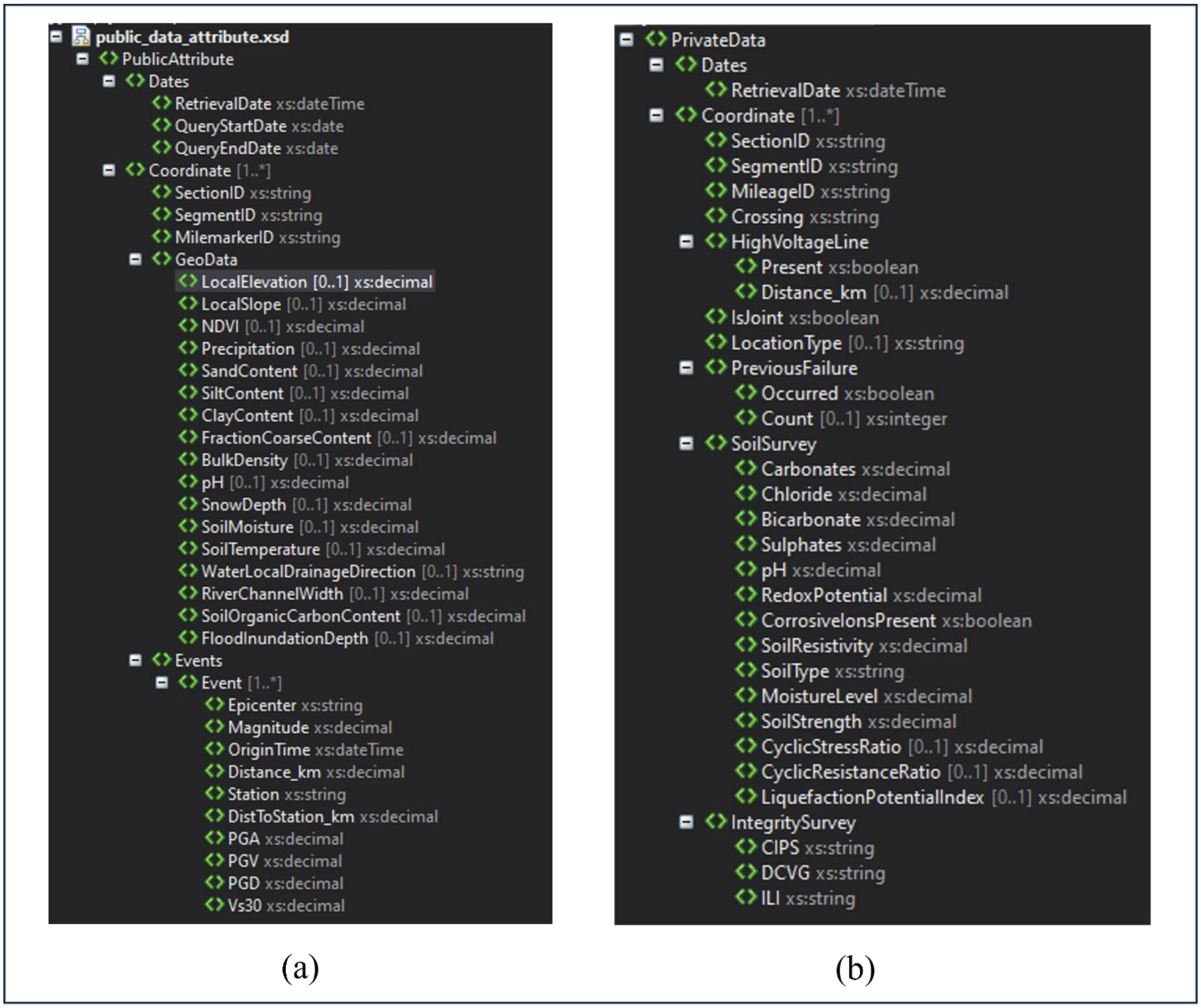

Figure 1: (a) Public attributes schema (.xsd) and (b) Private attribute schema(.xsd)

References

Jing, Hongyuan & Hao, Jianbin & Han, Bing & Wu, Zhangzhong. (2009). Semi-Quantitative Assessment Technology of Geohazard Risk for Pipeline. 1542-1551. 10.1061/41073(361)162.

Almeida, CO, Amorim, RS, Eltz, FL, Couto, EG, & Jordani, SA (2012). Rainfall erosivity in municipalities of Mato Grosso: seasonal distribution and correlations with rainfall data. *Brazilian Journal of Agricultural and Environmental Engineering*, *16*, 142-152.

Caratozzolo, V., Misuri, A., & Cozzani, V. (2022). A generalized equipment vulnerability model for the quantitative risk assessment of horizontal vessels involved in Natech scenarios triggered by floods. *Reliability Engineering & System Safety*, *223*, 108504.

Cohen, A., & Fielding, M. B. (1979). Predicting frost depth; Protecting underground pipelines. *Journal-American Water Works Association*, *71*(2), 113-116.

da Silva, A. M. (2004). Rainfall erosivity map for Brazil. *Catena*, *57*(3), 251-259.

Karthikeyan, A., Mirza, S., Hur, B., Pearlstein, G., & Ledbetter, R. (2020). Quantifying variable importance in predicting critical span length and scour depth for failure of onshore river crossing pipelines using ann. *Journal of Marine Science and Engineering*, *8*(11), 840.

Landucci, G., Necci, A., Tugnoli, A., Antonioni, G., & Cozzani, V. (2014). NaTech scenarios caused by flooding: Evaluation of accident frequency by the use of fragility models. *Chemical Engineering Transactions*, *36*, 427-432.