Semantic Web Technologies for External Corrosion Detection in Smart Oil Fields
M. Saeed, C. Chelmis, V. Prasanna, University of Southern California; B. Thigpen, R. House, J. Blouin, Chevron U.S.A. Inc.

Abstract

The Oil & Gas industry always seeks to prevent loss of containment (LOC). To prevent such incidents, engineers rely on inputs from various asset databases and software tools to make important safety-related assessments and decisions on a daily basis. One cause of LOC in offshore platforms is external corrosion. The state of corroding assets is extensively monitored and recorded through a variety of data collection mechanisms by various processes and people. Due to heterogeneity of these data sources, providing on-demand access to information with an integrated view can be challenging. A unified view of current data sources is desirable for decision making as it could lead to identification of telltale signatures of LOC events. However, manually cross-referencing and analyzing such data sources is labor intensive. Another challenge is knowledge management, which refers to a systematic way to capture the results of various engineering analyses and automated prediction models. It is beneficial to capture this knowledge for two reasons: (i) auditing, archiving, and training purposes, and (ii) mining of LOC signatures and warning signs for developing machine learning prediction techniques.

We propose the application of semantic web technologies for a holistic and expressive representation of various heterogeneous data sources at scale to deal with information management issues. The key elements of our approach are a reusable asset integrity monitoring ontology and an external corrosion ontology that model various elements from the domain, and a knowledgebase that can serve as a system of record for observed data as well as new knowledge acquired through inferencing and machine learning analytics. We further describe the methodology followed to populate the knowledgebase and how it can be used to convey assessments and alerts to the right people so that actions are taken to address identified risks.

We show that data from multiple sources can be integrated into a central repository serving as a single endpoint for maintaining and retrieving knowledge. We present our integration framework and evaluate the advantages of our approach in terms of expressiveness and ease of information access.

Our metadata and knowledge management platform for external corrosion can be highly beneficial for our strategic vision of LOC early prediction and prevention. The expressiveness afforded by the semantic web stack enables rapid integration and facilitates analysis of multiple data sources related to corrosion detection. The end goal for an enterprise is not just storing and managing lots of data, but to get actionable insights fast. Our proposed solution is a stepping stone towards LOC prevention.

Introduction

Oilfield operations are extensively instrumented and monitored; specialized tools are used by experts to detect, report, analyze and resolve safety related issues. Asset Integrity SMEs make safety-related decisions with inputs from asset databases and available software tools. However, it may be a labor intensive process for a team of humans to analyze the deluge of relevant data and systematically evaluate the risk factors relevant. Figure 1 shows a potential example Asset Integrity workflow diagram. Through metering and surveys conducted on a facility, data in the form of measurements, photographs and textual comments...
may be collected and stored as relational databases and spreadsheets etc. Information spreading across data silos may lead asset integrity SMEs to manually identify appropriate sources, browse through repositories, and create links between these chunks of information. Dedicated software or labor intensive process is used to get the complete picture of the state of the assets for decision making. Once the data is analyzed, potential risks are identified, work orders are issued if repair work is indicated, and schedules for next inspections are generated.

Our basic hypothesis is that loss of containment events have telltale signatures that can be identified early enough from current data sources related to asset integrity. These signatures may not be readily apparent and even when detected it may be difficult to understand their significance for process safety due to the sheer number of assets that have to be monitored. Recent advances in computing such as big data analysis, complex event processing systems, fault diagnosis, and machine learning enable the continuous processing of data in various formats, mining of this data for patterns, and prediction of a potential loss of containment event. The challenge therefore is to automatically integrate all relevant data aspects related to asset integrity in order to present a comprehensive view of the expected safety situation.

In this paper we propose a conceptual framework that facilitates semantic augmentation, integration and analysis of different data sources and we outline the use case of external corrosion detection and addressing any external corrosion. The purpose of our work is to explore semantic technology as a potential promising approach for building extensible Asset Integrity systems. To do this we start by defining the high level asset integrity concepts for asset integrity management and their relationships using a semantic model. We then extend our model to describe the specific use case of external corrosion. Based on our model we integrate heterogeneous inspection data to provide a complete picture of state of corroded entities on a facility. With the integrated semantic repository we can answer queries which may otherwise be time consuming if done manually. The contributions of our work are as follows:

1) We designed a semantic model using the Web Ontology Language (OWL [13]) to uniformly represent the process of data collection and analysis on an oil and gas facility. This high level ontology can be easily extended to zoom in on a specific process such as external corrosion detection.

2) We derived a specialized semantic model for external corrosion data monitoring, analysis and detection.

3) We propose a semantic external corrosion architecture for digital oil fields that facilitates integration of heterogeneous inspection data sources. We show an illustrative use case for external corrosion detection in oil and gas facilities. Overall this demonstrates that semantic modeling can form the backbone of next generation Smart Oil Field applications, facilitating information integration and knowledge representation.

Use Case of External Corrosion Detection

Oil and gas companies have detailed workflow in place to monitor and manage external corrosion. These procedures involve visual inspections, thickness measurements, corrosion level classifications and schedule maintenance work and other specialized system deployed in the operations environment. Such processes generate heterogeneous data which includes but is not limited to asset inspection reports, corrosion coatings readings, non-conformance reports, photographs. Figure 2 shows an example of asset integrity data collected for monitoring and prevention of external corrosion. These datasets have varying formats and update frequencies. Integration of such diverse datasets into a consistent format that is suitable for subsequent use and further analysis is therefore helpful. The task of manually analyzing data in an attempt to identify and diagnose items can
be labor intensive. Also, the task of analyzing factors after an activity can be cumbersome. Multiple inspection teams gather data from all the assets on a facility. This data is in the form of photographs, forms and handwritten notes which later gets entered into the databases. Afterwards, personnel can analyze the data to identify assets with potential risk. Based on the analysis, decisions about inspection scheduling, maintenance jobs and testing are made. This approach is time consuming, labor intensive, not scalable (especially in areas with high number of facilities and assets).

Figure 2: A snapshot of typical asset integrity data (anonymized)

Semantic Web Approaches to Sensors and Data

The inherent property of Semantic Web paradigm makes it easier to give meaning to data and provide the flexibility of linking multiple data sources together. In the domain of Semantic Web, ontologies (or vocabularies) define the concepts and relationships used to describe and represent an area of concern. The role of ontologies in Semantic Web is to facilitate data organization and integration. This integrated data (known as Linked Data) which can be used for reasoning or simply querying is the main strength of the Semantic Web. Most of the applications employing Semantic Web technologies are essentially based on the accessibility and integration of Linked Data at various level of complexities. [1]

Work has been done in the Semantic Web domain in terms of modeling and integrating sensor data from different fields. Semantic Sensor Network Ontology (SSNO) [5] focusses on domain independent sensing application by integrating sensor data (measurements) and sensor specific data (sensing principles, quality etc.). Sensor Cloud Ontology (SCO) [6] extends SSNO by adding provisions for the parameters being sensed as separate entities instead of just metadata, explicitly introduces the concept of time series, creates a Link Open Service (LOS) on top of REST API that provides a SPARQL [16] endpoint for querying integrated sensor data. A framework for encoding domain knowledge in the field of Ecological and Environmental monitoring is presented in [7]. The authors use OWL2 [14] to encode rules and regulations to flag critical sensors and sites. An interface supports both non-expert and expert users in water quality monitoring. Another approach is to build a semantic model on top of relational database by using RDF annotations [8]. A naming convention for URIs is based on the tables name, row and column names etc. However using two technologies in parallel increases complexity. Having everything in a relation database doesn’t allow the expressiveness and dynamics provided by RDF triple stores. Moreover, any application, developed based on such architecture that provides SPARQL endpoint and supports write back can introduce added complexity of managing the relational database and RDF annotations.
In [9] authors propose an extensible model that caters to the information diversity in Smart Grids with provision to integrate new information sources and concepts. It is shown that such model can facilitate dynamic Demand Response (DR) planning for the utilities as it is capable of presenting electric consumption data, weather data, building occupancy data etc. In [10] the authors provide details of a process used to convert the bulk of tabular data (exportable as CSVs) available as Norwegian Petroleum Directorate’s Fact Pages into semantic data. The tabular data in CSV format was first transformed to a relation database and then converted to RDF format. In the next step an OWL ontology was built on top of the data for reasoning. The purpose stated for doing so was lack of standardized tools available for directly converting CSV data into RDF as compared to well defined standards available when converting RDB to RDF. The authors used SQL queries on the RDB database as a baseline to evaluate the query performance over semantic database. An asset integrity and risk management ontology for IT assets is presented in [12] which describes the architecture of different layers and interconnections within an Enterprise network in detail as such information is relevant for IT security risk analysis. Interconnections of different assets is crucial in making informed decisions about Network security.

Some of the works described above sufficiently model sensor data and metadata and provide integrated view of data for decision making purposes. However, there are some unaddressed issues which are summarized below:

**Limited Definition of Data Sources:** Data is no longer being generated by Sensor Networks only. On any industrial facility a large amount of data is man-made, although its volume and velocity differs from sensor data. In such scenarios it will be useful to instead of modeling all data generation entities as sensors, abstract them and focus on the data itself. Information about the entities can be modeled as metadata in the ontology.

**Data Heterogeneity:** In a traditional sense, a sensor observation has so far being modeled keeping in mind the numeric nature of it. However as discussed above, due to new data generation entities, now we have data comprising of images, videos, text (structured and unstructured), drawings etc.

**Data derived from Data:** The works described above sufficiently model sensor data and metadata but limit their focus on providing integrated view of the raw data generating sources to the end user. Due to advent of data analysis techniques such as machine learning, text analytic and image analytic techniques are commonly applied to raw data to detect patterns, learn hidden correlations, finding new facts or extracting data from images. This derived data should be integrated with the raw data for querying and future analysis.

**Proposed Approach**

One of the main advantages of ontology based information integration is its extensibility and reusability which is crucial in environments which have dynamic information streams. Ontologies can be developed in modular fashion to model and represent existing knowledge and can be easily extended to capture new information. In the spirit of Semantic Web, our approach has been to reuse existing ontologies so as to leverage already existing domain knowledge. In this context, we extend basic concepts about modeling data streams from PoEM (Process-Oriented Event Model) [2], an end-to-end event modeling framework which defines industrial processes at a high level. While the emphasis of PoEM is on modeling dynamic data flows leading to detecting events, we believe that any ontology modeling industrial processes should also capture entities (physical objects or static entities) and their inter-relationships. Some of the existing well know relational schemas (PPDM [3]) and ontologies (ISO 15926 [4]) model physical entities such as vessels, pipelines, tanks, motors etc. We use entities from latter to model static assets such as platforms, vessels and pipelines etc. in our ontology. We believe that knowledge about interaction between assets is useful for making informed decisions related to asset integrity.

**Semantic Model of Asset Integrity Data**

We build our semantic model of Asset Integrity data in a modular way. Figure 3 shows the classes and properties defined in our core Asset Integrity Ontology.

We start with the definition of basic concepts and relationships needed to represent key oil field entities such as business unit, production area, facility and equipment using the ISO-15926 [4] ontology from the oil and gas domain. For concepts related to observations and measurements we leverage concepts defined in PoEM [2] ontology.

**System and Facility:** We define the class “System” (class and concept are used interchangeably) by extending “ClassOfOrganization” from ISO-15962 to define a logical collection of other logic or physical objects. A logical organization can be a company, a business unit or a production area. The “hasSystem” self-property signifies that there can be a hierarchical organization of such systems. The concept of facility is defined using “physicalObject” from ISO-15962. A facility represents any grouping of equipment (usually located in close proximity) where certain products are extracted, processed, stored or metered e.g. a rig, platform, refinery etc.
“ObservedEntity” and “CriticalEntity”: The class “ObservedEntity” models all assets (physical objects) using “physicalObject” class from ISO-15962 ontology. These assets have data that may be metered or captured as numeric data, photographs, reports etc. This class can be used to model entire equipment as well as some specific part of the equipment. For example a gas compressor can be modeled as a single entity or as an entity consisting of other entities (vessels, valves etc.) in multiple stages depending on the relevance of data being generated. A specific example from domain of external corrosion detection is the existence of multiple different thickness measuring locations (TMLs) on a single piece of equipment, hence it will be useful to model them as individual data generating entities. The class “ObservedEntity” provides flexibility of modeling the equipment as whole, constituting of multiple sections or of multiple sensing points or measuring locations. The sections themselves can be modeled as whole or as set of individual sensing points. This helps in modeling the separate data generating entities on the same equipment individually but also maintaining the information about their interrelationship. Similarly “connectsTo” property can capture the information about the link between two separate assets e.g. connection of first and second stage separators through a valve. The purpose is to get as much information as possible into the model and use this at the time of data analysis to find correlations between individual data streams. “CriticalEntity” is a sub class of “ObservedEntity” which defines an asset that have an active alert associated with it. This purpose of this concept is to facilitate queries that require a list of all critical assets on a facility.

“Parameter” and “ObservableProperty”: An “ObservedEntity” can have an “ObservableProperty” that we are interested in measuring e.g. pressure, temperature or level or it can has a parameter that is indirectly measured as another “ObservableProperty” e.g. extent of corrosion is measured as thickness in inches. In this case corrosion is the “Parameter” and the “ObservableProperty” is thickness. Based on the definition given in [2] we consider “ObservableProperty” as anything from simple yes/no observation (e.g. Accessibility) to physical and chemical properties requiring sensing techniques.

“Measurement” and Measurement Time Series (“MeasurementTS”): For each “ObservableProperty” we associate a time series of measurements (“MeasurementTS”) which comprises of individual time stamped “Measurement” instances. In case if a single sensor is providing multiple parameters e.g. pressure and temperature, then both measurements can be labelled under separate observable properties for the same “MeasuringLocation” instance for that asset. In our ontology the definition of “Measurement” is not limited to numeric data. We can model a database of images or manually entered comments as a time series as well.

“Threshold”: “Threshold” models set points or acceptable levels for an “ObservableProperty” based on certain defined guidelines e.g. minimum thickness for a pipeline, pressure safety high, low intake pressure for a pump etc.

“AnalysisResult” and Analysis Time Series (“AnalysisTS”): One of the major goals of our ontology is to facilitate data analytics applications for which these particular concepts have been introduced. The “AnalysisResult” is a high level entity which can be used to model any form of results derived from collected data. This analysis can be anything from a simple comparison operation (e.g. set point comparison) to machine learning algorithms. The “AnalysisTS” captures the analysis done
on the same asset over time. This paves way for integrating analyzed data and making it available along with the raw data to enable further analysis in future. For example, an inspector has access to set of photographs of the corroded equipment on a facility. He then manually labels each of them with corrosion level categories (mild, severe or numeric levels). This is an example of manual analysis. These results can then be used, as a baseline, by an image analytics software to automatically perform the same function.

“Alert” and “CriticalAlert”**: “Alert” is an abstract concept based on the definition of Event in PoEM ontology [2]. It is used to model the critical decisions made through analysis. For example, if the thickness of the outer surface of an asset is below minimum threshold then an appropriate alert can be severeCorrosion. Any “ObservedEntity” with a “CriticalAlert” will automatically be labelled as a “CriticalEntity” due to inferencing, making querying simpler.

“WorkOrder” and “CriticalWO”**: Assets can have “WorkOrder” instances associated with them. If the “Asset” has an active “CriticalAlert”, then the “WorkOrder” instances associated with it can be considered “CriticalWO” instances.

“State”: In the ontology we keep the record of “currentState” and “previousState” of an asset. Example of valid states can be working, failed, high alarm or corroded.

**External Corrosion Ontology**

The Asset Integrity model is extended to model all concepts and properties that may be relevant to external corrosion on oil and gas facilities. Figure 4 shows our external corrosion ontology.

“Area” and “Facility”: “Area” refers to a logical grouping of the facilities. It can be a production area or a business unit. Facility refers to an onshore or offshore installment, examples of which are a production facility, storage facility or a refinery etc.

“Equipment”, “Section” and “MeasuringLocation”: These represent the physical assets which are the focus of the external corrosion detection and prevention procedures. An equipment in our ontology refers to an entity which has been given a unique equipment id (or equipment circuit id) within a facility e.g. AB-101 can be a vessel or a tank etc. Each equipment can be of category vessel, tank, pipe, pipeline or a flow line. For visual inspection purposes a large equipment can be divided into sections e.g. a header can be divided into two straight pipe sections referred to as PS-1 and PS-2. For thickness measurements using...
ultrasonic or other techniques, each equipment can have multiple thickness measuring locations (TMLs).

“Parameter” and “ObservableProperty”: The “Parameter” that we are interested in measuring is corrosion. There can be various ways in which facilities monitor corrosion. In our ontology we have specified only three type of data collected, however it can be easily extended to other types. We consider thickness measurements (in inches) from TML points, photographs taken of an equipment or its sections and written assessment (in the form of unstructured text) as shown in Figure 2.

“Thickness” and Thickness Time Series (“ThicknessTS”): For every TML on an equipment we have a time series of thickness values. A method for ontological representation for time series data is described in [12] for four different type of time series. Based on sampling period there are event based and periodic time series and both type of time series can be cumulative or non-cumulative. The thickness measurements form an event-based non-cumulative time series. The authors have proposed rules in OWL to put restrictions on all observations in a time series that they should share same values for common attributes representing metadata i.e., the “FeatureOfInterest” and “observedProperty” should be same e.g. for a time series measuring rainfall, all individual measurements will have data property “observedProperty” set to “rainfall”. In our approach, the hierarchical structure of the ontology allows grouping of the data being collected for a particular entity for the same “observableProperty” in a single time series.

Image and Text Data: Corrosion survey reports usually contain photographs of the corroded surfaces along with manual assessment of state of the assets and recommendations. Our purpose is to make the organized set of images available to users for querying, use them for image analytics and integrate the findings in the semantic repository. Unstructured text data is also available as a text log, which can be used to mine useful keywords to create applications such as tag clouds.

Alerts and States: Based on thresholds comparisons, predictive and image analytics, the state of the corrosion of a particular equipment, section or measuring location can be determined, which can be considered as the “currentState” for the that asset, whereas as the last current state now becomes the “previousState”.

Asset Integrity Knowledgebase

The asset integrity knowledgebase is derived by semantically enriching data from the external corrosion monitoring process (Figure 2). The enrichment is performed based on the concepts and relationships defined in the asset integrity ontology. In semantic repositories knowledge is represented as triples. A triple consists of a subject, predicate and object. The predicate denotes the relationship between the subject and the object. For example <ABC-100, type, Equipment>, <ABC-100, group, Vessel>, <ABC-100, has Part, TML2.0>, <TML2.0, value, 0.5> are triples that define an equipment ABC-100 with equipment group of vessels. ABC-100 has a TML labelled 2.0 which has a thickness measurement of 0.5 inches.

Semantic Query Engine

The semantic query engine is responsible for executing SPARQL [16] queries and communicating the results back to the users. A semantic query may solicit information from multiple data source. The concepts related to asset integrity are obtained from asset integrity ontology and those related to external corrosion are extracted from external corrosion ontology. The capability of linking such data is difficult without a semantic querying mechanism.

The following semantic query returns instances of work orders for equipment for which one or more thickness measurements have fallen within 10% of the threshold for minimum thickness as per the most recent survey for a given facility FacilityA by seamlessly returning data from multiple sources now integrated under a single semantic repository. The query using concepts and relationships defined in both ontologies i.e., asset integrity ontology and external corrosion ontology.

Sample semantic Query:

```
WHERE
{ 
  facilityA eco:hasEquipment ?equipment.
  ?equipment eco:hasPart ?tml
  ?ts aio:contains ?m;
  aio:min ?min.
  ?m aio:date ?lastdate.
  ?m aio:value ?value.

  FILTER (?value < ?min*1.1 && ?sts == "In Progress")
}
```
The above query can easily be modified to retrieve the TMLs which match the given criteria for equipment across facilities. To further demonstrate the capabilities of the proposed system we provide a list of possible query statements that can be useful from the end user’s perspective pertaining to external corrosion, monitoring and analysis. Such statements are converted to semantic query by the semantic query engine. The statements are shown below along with the databases (Figure 2) from which the data will be pulled.

- Find open work orders for equipment of a particular risk category. (Coatings, Images, Work Orders)
- Find TMLs that have been replaced within a certain time frame. (Thickness Readings, Work Orders)
- Show work order on equipment with critical corrosion severity pending due to rope access. (Thickness Readings, Images, Coatings, Notes)
- Show which equipment had a similar issue in the past and what was the solution/recommendation. (Notes, Work Orders)

**Conclusion and Future Work**

In this paper, we presented our work on semantically modelling asset integrity data for external corrosion monitoring and management. Our semantic model is extensible with provision to easily integrate new data sources and domain concepts. To show this we extended the asset integrity ontology to derive the external corrosion ontology. The purpose of creating the ontologies and integrating data was to organize heterogeneous data sources for simplifying on-demand integrated information access & enabling complex analytics to be performed on the integrated knowledgebase. An important reason for choosing semantic web technologies was our vision for an asset integrity management system of record. We believe that the expressiveness afforded by the semantic web stack will prove to be highly beneficial for intuitive representation of asset integrity data. Such representation not only explicitly reveals relationships between facts but can also be used to drive machine learning methods to infer hidden or implicit relationships between entities. It is also intuitive to represent domain knowledge of possible (external) corrosion mechanisms using semantic rules. A rule defines a set of statements as conditions and the result consists of a set of sequences. In our future work we plan to create such rules in order to facilitate rule based inferencing which in turn leads to smart decisions on the runtime by enabling automatic assessment of corrosion critical events during data collection time. We expect our framework to reduce the response time to corrosion detection and optimize the scheduling of preventive maintenance.

**Acknowledgement**

This work is supported by Chevron U.S.A. Inc. under the joint project, Center for Interactive Smart Oilfield Technologies (CiSoft), at the University of Southern California.

**References**

[16] SPARQL Query Language for RDF. http://www.w3.org/TR/rdf-sparql-query/