EDRC User Case 2: towards understanding and application

(February 2016, version prepared by Geiza M. Hamazaki da Silva and Livia Ruback, Tecgraf Institute/PUC-Rio)

1. Introduction

This document describes the adoption of ISO15926 standard for oil companies' data integration, showing the main advantages through a real case study. Section 2 brings the main concepts and technologies related to the oil data integration. Section 3 shows the architecture of the integration process proposed. Section 4 presents a case study, developed by the EDRC (Capturing Equipment Data Requirements Using ISO 15926 and Assessing Conformance)[1], about a request for quotation for pressure transmitters. Finally, Section 5 brings some conclusions and future work.

2. Basic concepts and technologies

The main concept related to the integration problem is the *Interoperability*. Interoperability is usually associated with two applications being able to "work together" (whatever that means) by virtue of each, independently, following an outside standard. In the end they may be able to work together just as well as two integrated applications, but because the "working together" was achieved by each implementing an outside standard we call it "interoperability" [1].

Another important concept in this context is *Linked Data*, which is a set of principles and technologies that take advantage of infrastructure of the Web to enable data sharing and reuse on a massive scale [2]. One of these principles argues in favor of using a simple graph-based data model that has been designed for use in the context of the Web, called *RDF* (Resource Description Framework)[2]. The RDF model allows data be structured in triples (subject, predicate and object) that are stored in a specific triples-oriented data server called *triplestore*. The language to express queries across di-

[1] http://www.fiatech.org/operations-maintenance/231-group-assets/1155-capturing-equipment-data-requirements-using-iso-15926-and-assessing-conformance-edrc
[2] http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/Overview.html

verse RDF data sources is known as SPARQL[3] and allows querying required and optional graph patterns along with their conjunctions and disjunctions.

Another concept important to mention is *Ontology*. According to Gruber [3], an ontology is "an explicit specification of a conceptualization". An ontology, basically, comprises a set of classes representing some domain and together with the properties, which qualifies the classes' instances. There is no clear division between "vocabularies" and "ontologies". According to W3C [3], the trend is to use the term "ontology" for more complex, formal set of terms, whereas "vocabulary" is used when such formalism is not strictly necessary.

The *ISO 15926* ("Lifecycle Integration of Process Plant Data Including Oil and Gas Production Facilities"), also known as the Oil and Gas Ontology, is a standard for exchange and reuse complex plant and project information. In other words, the motivation behind the creation of the standard was to mitigate the current high costs of rekeying and reformatting information to move it from one proprietary system to another [1].

ISO 15926 is divided into some parts, which can be seen as "parts of human speech". The most relevant in this context are: (i) Part 2, the *data model*, is the core of ISO 15926 [1] and is equivalent to the rules of grammar; (ii) Part 4, the *reference data*, is equivalent to the dictionary [1]. When two people learn the same grammar rules using the same dictionary, they can communicate freely. Similarly, two machines exchanging information using Parts 2 and 4 can therefore communicate freely and without ambiguity [1].

Besides the core of the ISO 15926, Part 7, in turn, comprises the so-called templates, the equivalent to a phrase book that allows users to construct meaningful sentences. As well as Parts 2 and Part 4, the templates are represented with ontologies and serialized using the RDF model.

POSC Caesar Association (PCA)[4] is a non-profit global-standardization member organization that shall promote the development of open specifications to be used as standards for enabling the interoperability of data, software and related matters. PCA initiated ISO 15926 and is committed to its maintenance and enhancement.

All these concepts are essential for understanding the architecture proposed and the examples of the next sections.

[3] http://www.w3.org/TR/sparql11-overview/
[4] https://www.posccaesar.org/

3. The architecture

Figure 1 shows the general architecture for dealing with data interoperability in the oil/gas scenario. It comprises oil and gas companies sharing or consuming data from each other or from others.

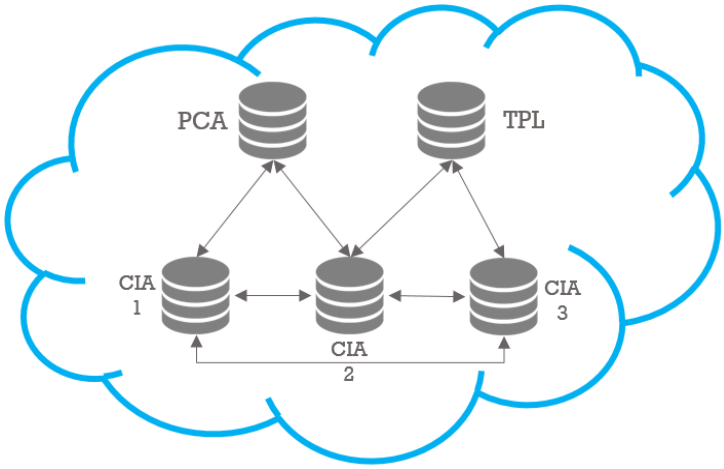

Figure 1: Interoperability process

PCA is the dataset published by the POSC Caesar Association[5]. TPL is the dataset with ontology comprising the definitions of ISO 15926-7 templates, as described in Section 2. The others datasets are oil and gas companies dataset published in the RDF format following the ISO ontology, for instance the Emerson Process Management[6] as presented at the EDRC User Case2 (section 4).

Figure 1 represents the integration among all the datasets, which means that all datasets inside the ring can communicate and exchange information with each other. Imagine that Company 2 needs to buy some equipment and ask for a Request For Quotation. Company 1 (in the example, Emerson) have the pressure transmitters that match the Company 2 needs. Since both companies datasets are able to exchange information, Company 2 can easily access Company 1 data.

With the datasets integrated, one can perform distributed queries over all of them. Next sections brings some examples of queries.

[5] http://data.posccaesar.org/rdl/
[6] http://www2.emersonprocess.com/

4. The pressure transmitter case study

The case study described in following was extracted from the EDRC Use Case 2 and refers about RFQ (Request For Quotation) information for pressure transmitters[7] and a reply to it. Recalling the example described in last section, imagine that the equipment Company 2 (see Figure 1) needs is a certain pressure transmitter. In this context, some questions that integrate all the datasets can arise.

In the following, some examples (for more details, see the technical document by Tecgraf, "UC2_SPARQL_Examples.pdf").

Question 1: *Who the element that the Request for Quotes refer about?*

Question 2: *What are the parts of the Pressure Equipment?*

Question 3: *What are the Calibration Ranges for the Pressure Equipment ordered?*

Question 4: *What are the Normal Operating Pressure and Temperature of the Pressure Transmitter requested?*

Question 5: *What are the cost (currency, cost type and value) of the equipment ordered?*

As an example, we can see the following SPARQL query that answer question 5.

[7] http://www2.emersonprocess.com/en-us/brands/rosemount/pressure/pressure-transmitters/pages/index.aspx

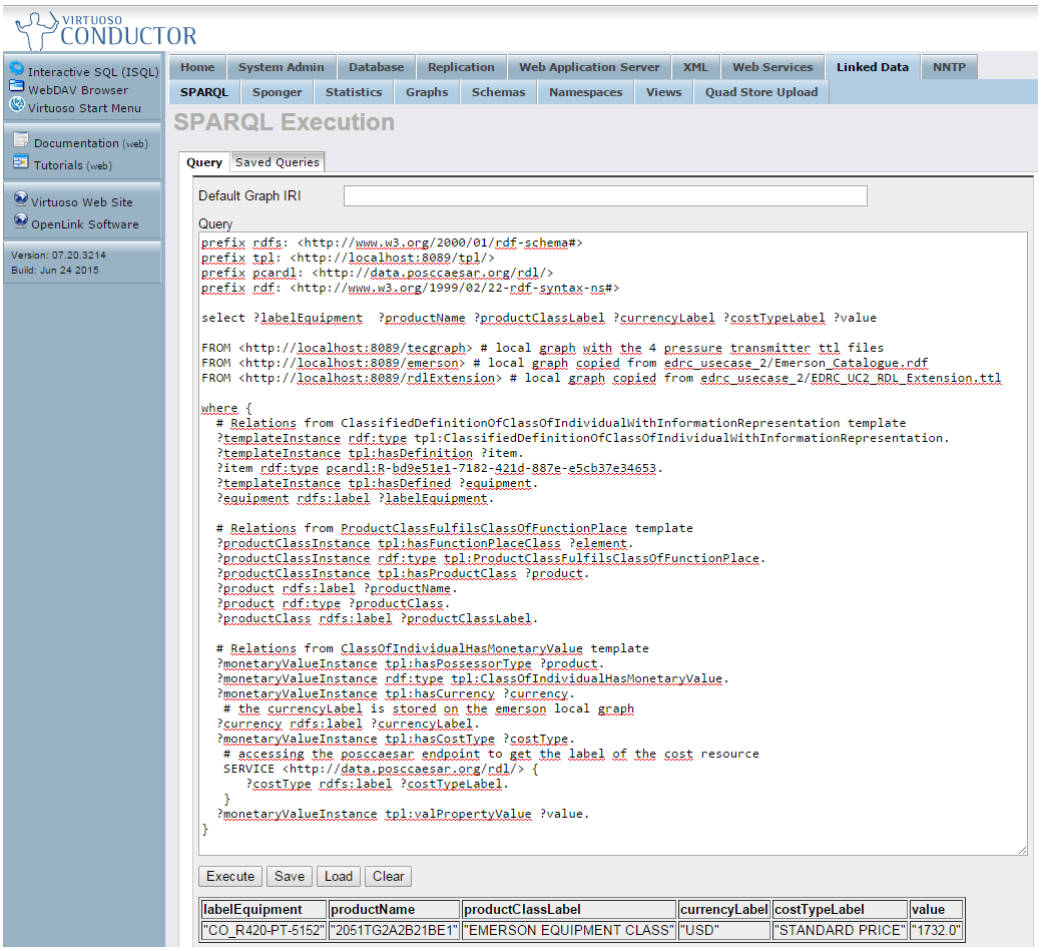

Figure 2: Result of the SPARQL query execution

Figure 2 shows the result of the Query 5 being executed over Virtuoso triplestore[8]. As we can notice, without the ring, i.e., the interoperability layer that allows the companies exchange structured data, these queries would not be answer – at least not so easily. Not so easily because often the companies publish the data in heterogeneous format (excel spreadsheets, pdf documents, html files). In these cases, a complex mapping between the models, the formats and the instances is necessary before putting the data together. Figure 3 shows part of the graph used for the SPARQL query shown in Figure 2. It contains three templates (tpl:ProductClassFulfilsClassOfFunctionPlace, tpl:ClassifiedDefinitionOfClassOfIndividualWithInformationRepresentation and tpl:ClassOfInidividualHasMonetaryValue) with its respective instances (see more in the technical document "UC2_SPARQL_Examples.pdf")

[8] http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/

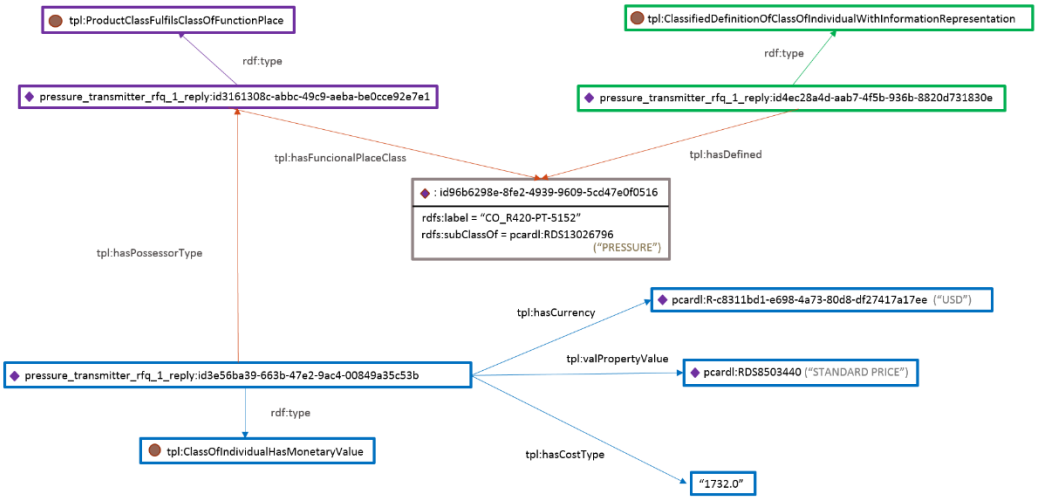

Figure 3: Graph related to the SPARQL query

If the mapping is previous done for all companies, following the same patterns the data combination becomes an easy task.

5. Conclusions and future work

This document described an example of how interoperability can be applied in the context of ISO15926 standard and oil company's cases. We introduced the main concepts related to this scenario, such as Interoperability, Linked Data, ISO 15926, among others. We showed a general architecture of the intended integration scenario and some questions that can be answered in this integrated scenario and the SPARQL query for one of these questions, together with the query execution result.

This document is intend to be less technical. For more technical issues, please check the technical document "UC2_SPARQL_Examples.pdf" and " Report EDRC Use Case 2.pdf"

6. References

[1] Bentley – An introduction to ISO15926. Available at http://iringtoday.com/wordpress/wp-content/uploads/2011/12/iso-intro-ver1.pdf. Acessed: 13th January 2016

[2] Tom Heath and Christian Bizer (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

[3] Thomas R. Gruber. 1993. A translation approach to portable ontology specifications. Knowl. Acquis. 5, 2 (June 1993), 199-220. DOI=http://dx.doi.org/10.1006/knac.1993.1008 Acessed: 13th January 2016

[4] Ontologies - W3C - http://www.w3.org/standards/semanticweb/ontology . Acessed: 13th January 2016

[5] Fiatech - http://www.fiatech.org/desing/projetcs/589-iringtools-interfacing-project. Acessed: 13th January 2016