New ways of using standards for semantic interoperability towards integration of data and models in industry

Jean-Charles Leclerc[1], Yves Keraron[2], Claude Fauconnet[3], Nicolas Chauvat[4], Martin Zelm[5]

[1] TotalEnergies, Tour Coupole, 2 place Jean Millier, 92400 Courbevoie, France

[2] ISADEUS, 20 avenue des muses, 44470 Carquefou, France

[3] SousLeSens, claude.fauconnet@gmail.com

[4] Logilab, 104 Bd Auguste Blanqui, 75013 Paris, France

[5] Interop-VLab, 21 rue Montoyer, 1000 Brussels, Belgium

Corresponding author: jean-charles.leclerc@totalenergies.com

Abstract.

Recent European H2020 projects and clusters, Joint Industrial Projects in industry and advanced standards of Standardization Development Organizations converge towards new ways of using standards to enable integration of data and applications, thus enabling new ways of working all along the products' and plants' lifecycle and ecosystem.

In this paper, we will describe innovative means developed by TotalEnergies to address the lack of interoperability of data produced all along the lifecycle of an asset. These means result in a TotalEnergies Semantic Framework, which aims at interfacing according to generic principles source data with reference standards to provide internal and external users with data they can compute in their own applications to support their processes.

Keywords: Enterprise innovation and standardization, digitization methodology, semantic interoperability, data modelling, digital twin, semantic web, linked data.

1 Introduction

Industrial companies have a huge amount of data which cannot be retrieved and used easily because they are locked in, being not computable, stored on paper, or in proprietary formats. TotalEnergies has experimented a standards-based approach to solve these issues in an innovative way.

New types of standards are necessary for industries to succeed in their digital transition. Efforts are being carried out in ISO/IEC to provide computable standards with model-based standards, e.g., ISO 15926 series [1], and SMART standards [2]. European coordination projects support integration of innovation and standards for industry of the future, e.g., the projects as Ontocommons [3] or StandICT [4]. Professional organizations or clusters in different industrial ecosystems are publishing roadmaps and conduct Joint Industrial Projects to develop the framework for needed standards development and use at the digital era.

The conceptual framework we draft in this paper is driven by a Systems Engineering mindset, the methodology is based on ontologies to make data interpretable by all stakeholders of the ecosystem thanks to the linking to standards, and opensource tools are using semantic web technological standards for the implementation.

This paper is organized as follows: In section 2, we will describe the industrial context and motivations, in section 3, we will present the necessary shared conceptual framework to organize and manage information on an industrial asset, in section 4, we describe the TotalEnergies semantic framework, supported by the development of a set of open-source tools based on W3C standards to match the methodological process requirements, in last section, we will open the discussion on social challenges and conclude.

2 Context

2.1 Innovation and standardization

Standardization and innovation apparently seem contradictory. Efforts have been carried out to raise awareness among researchers and innovators on the importance of a standardization strategy in the development of game-changing products and services.

As shown on figure 1, innovation and standardization are in dynamic relationship with industry competitiveness, and we need to consider the impact of the digital technological environment on these dynamics.

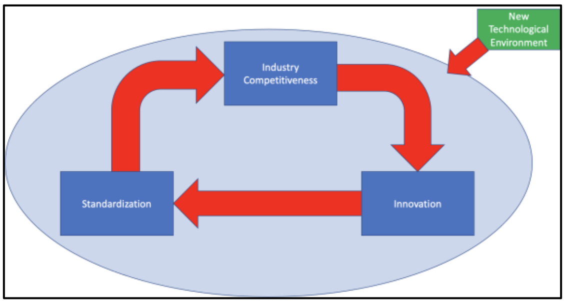

Figure 1: Standardization – Industry competitiveness – Innovation in a digital environment

Data and models' standardization stresses this problematic when advanced information and Communication Technologies, as Semantic Web standards and technologies, offer new opportunities for re-modelling industrial ecosystems to improve their economic balance between fulfilled needs and used resources. This approach of combining knowledge and technology, contributes to build a more sustainable and inclusive future and to address major societal challenges as climate change.

The approach will follow the concrete industrial Oil and Gas ecosystem, based on the principles, which made the success of the Web. The initial proposal of Tim Berners Lee [5] deserves a careful reading. It is striking how the analysis of the issue related to the availability of information in an ecosystem and the Tim Berners Lee's proposition fits to other ecosystems than CERN as industrial ones. The solution consists of using hypertext and insists on the separation of the data storage systems from the data display systems and on the interface between them.

In the Oil and Gas ecosystem, and beyond, in process industries, standards and the linking to standards play a key role in this interfacing to leverage the value of existing data and to support the use and creation of knowledge for decision making in the collective processes of industrial engineering.

2.2 Industrial context

The current way of working is often one dimensional and fragmented; actors struggle to traverse different viewpoints such as operations, maintenance, engineering and aligning References Data Libraries from various standards and normalization groups is still challenging in terms of harmonization.

In real life, we meet complexity and need to work with several hierarchical data structures that serve the needs of different perspectives to represent in a precise manner our Industrial assets all along the lifecycle.

Industry of the Future (Digital Factory, IEC CDD[1], AAS[2], of RAMI 4.0[3], …) needs a holistic plan and a methodological common framework to accelerate alignment of an Industrial References Data Foundation. We need to agree and to develop a governance model to enable Digital Transformation. This implies structuring and digitalizing information and standards in the same way to avoid dispersion of initiatives.

It is being considered that the coupling of ISO 15926-14 [6] with ISO/IEC 81346 [7] series standards, offers a reference backbone adapted to structure and link correctly together the data of multi-energies assets.

3 The necessary shared conceptual framework

We have already addressed the need for a shared conceptual framework for data interoperability in the domain of maintenance [8].

This conceptual framework shall fit to the formal ontology of the industrial engineering, that means that of systems engineering, as a foundation to other domain ontologies [9].

This approach is focused on the composability of different parts of a complex system, seen as a dynamic whole, open to legacy and future models, based on existing, in the making or on future standards.

This is the best guarantee to bring sustainable solutions to long-standing issues of lack of interoperability and to the impossibility to exploit the entire actionable legacy knowledge of a company.

---

[1] IEC CDD: International Electrotechnical Commission Common Data Dictionary
[2] AAS: Asset Administration Shell
[3] RAMI 4.0: Reference Architectural Model for Industry 4.0

3.1 ISO/IEC 81346 and Systems Engineering

Our approach is based on existing cross domain proven standards as ISO/IEC 81346, Reference Designation System [7], which introduces the concept of aspect, one key to have a common approach to integrate different perspectives.

ISO/IEC 81346 brings very helpful principles to build a complex system; that means building from heterogeneous parts and bringing them together in a common framework.

It fits with a systems' engineering approach and is the pillar of the asset Information Management Framework (IMF) [10] of the READI JIP project [11], which aims at digitalizing the requirements of standards applicable to the Oil and Gas exploration sector in Norway.

ISO/IEC 81346 offers the necessary and sufficient framework to support a systems' engineering methodology, to provide linked structures for the management of all the information produced in the lifecycle of an asset and to specify the data models and standards to be used in different contexts.

This Framework opens the door to benefits brought by early verification of the conformity of design with requirements through early trustworthy digital twins, which will be further completed and used in the operation phases with connection to the streams of data from sensors of the physical observable asset, associated with data analytics.

Figure 2 shows the 3 primary aspects of an object according to ISO/IEC 81346-1; these aspects are supported by the ISO/CD TR 15926-14, Data model adapted to OWL 2 Direct semantics [6]. Data standards are used for specific contexts, e.g. ISO 14224 for reliability and maintenance data.

ISO/IEC 81346-1 also brings principles of modularity and configuration management enabling the reuse of modules and the follow-up of the modifications during the lifecycle, a critical topic of configuration management. It benefits from dozens of years of application in various industries, as power generation, and supports the integration of existing designation systems. This flexibility and scalability make the standard easier to implement in industry and able to integrate existing set of data.

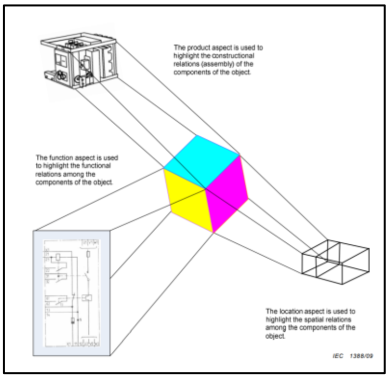

Figure 2: The 3 main aspects of an object – Extract from ISO/IEC 81346-1[7]

ISO/IEC 81346-10 [12] in its new version to be published in 2022 specifies the reference designation systems for objects of Energy Production Systems.

ISO 15926-14 reconciles the needs of the whole ecosystem and supports its coherence because the applications of the recommendations of the W3C supports a self-regulating, self-developing and self-reinforcing information system with common constraints.

3.2 Application in the context of Oil and Gas pilot

Contributions to normalization bodies groups such as IOGP[4]-DISC[5], JIP33, JIP36 CFIHOS[6] are unique opportunities to cross several knowledge domains and to experience together a methodology and its applicability by developing a learning non-competitive space to experiment on real asset data together.

We conducted in IOGP (2020-2021) a Low Voltage Motor Full Digital-Twin Pilot of standard modelling and digitalization in partnership with the JIP[7] READI[8]. It makes possible structuring fragments of information step by step, in such a way that each step provides immediate value without the need for a huge upfront investment.

Finally, we do believe that a common language across our disciplines and métiers is of key importance; this standard language creates unambiguous understanding among all stakeholders internally and externally. This semantization layer, as described in section 4, is key to get adequate data quality, in a trustable and reliable way, for data analytics and further reuse in other applications.

4 TotalEnergies approach of a standard Semantic Framework: a methodology supported by SousLeSensVocables Tools

4.1 TotalEnergies Semantic Framework (TSF) foundations

Relational models and databases hardly manage complex and evolving data and faces complexity efficiency challenges when trying to interconnect data because of the "spaghetti plate" effect. The semantic web approach is the most convenient way to implement standards through ontologies:

- capability to manage complexity with the concept of aspects
- agility to evolve compared to the rigidity of SQL models

Implementing the TSF process implies that the standards used are expressed using so-called controlled vocabularies that ensure disambiguated semantics both for human communication and machine processing. For controlled vocabularies we use W3C Standards such SKOS [13] for thesaurus and OWL [14] for ontologies both having a graph structure.

---

[4] IOGP: International Oil and Gas Producers
[5] DISC: Digitalization and Information Standards Subcommittee
[6] CFIHOS: Capital Facilities Hand Over Specification
[7] JIP: Joint Industry Project
[8] READI: REquirement Asset Digital lifecycle Information

In addition, semantic web technologies define the Linked Data principles that provide for each atomic resource a unique resource identifier URI and eventually an URL, unique resource locator, which can be also an hyperlink to a web site that hosts the resource identity card.

The TSF methodology and tools are designed to align enterprise data on standards using semantic web technologies and are guided by several principles:

- transparency and auditability of all treatments (no black box effect)
- usable by business experts in a collaborative process
- iterative and cumulative process to deliver both short term and patrimonial value

Standards are historically expressed in pdf format including tables and text. For years now some pioneer standards have been natively expressed or translated in OWL.
TSF uses ontologies natively expressed in OWL by their editor or when necessary build OWL ontology from native format csv or even pdf arrays: SousLeSensVocables [15] manages mainly graphs and URI's at all levels.

4.2 The TotalEnergies Semantic Framework Processes supported by SousLeSensVocables (SLSV)

The most valuable outcomes of TSF process consist in producing standardized knowledge graphs: foundations for digital twins and business domains ontologies. This provides a framework for both data and knowledges governance as showed in figure 3.

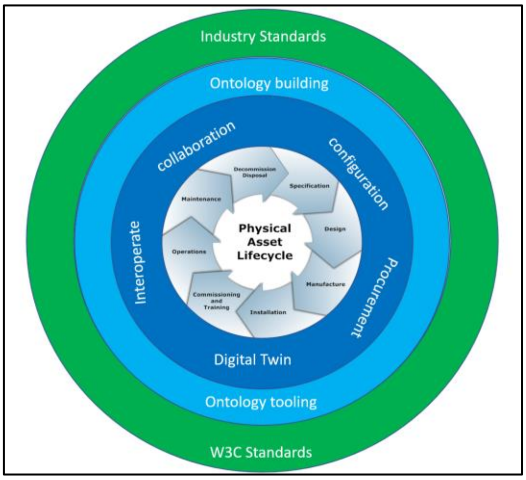

Figure 3 – TSF framework to manage assets life cycle stages with Standards, Ontologies and Graph management tools

SLSV is a set of experimental tools we developed and continuously improved to implement this process. These tools can be used separately to explore each step of the process as well as to orchestrate generation of standardized knowledge graphs. The implementation of TSF processes faces different challenges as described in figure 4.

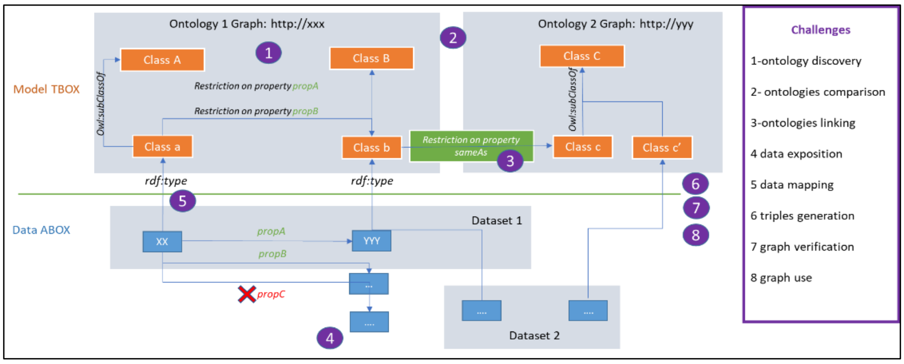

Figure 4– TSF process' challenges to transform traditional datasets into knowledge graphs

4.3 Standards TBOX discovery and comparison based on label similarity

To reach TSF goals and choose the most convenient classes among all available standards one needs to explore and compare them in their different aspects. This requires an interactive and visual environment.

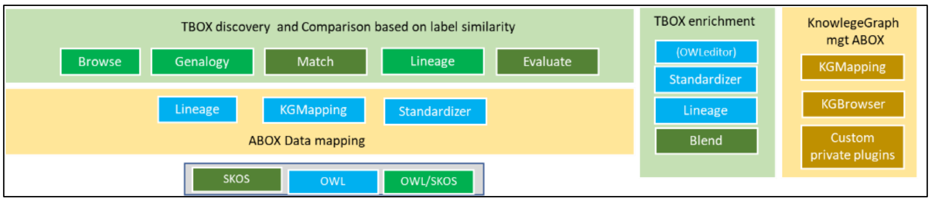

Figure 5– SousLeSensVocables tools and their functional perimeter

In SousLeSensVocables toolbox as shown in figure 5, this ontologies exploration task is mainly made possible using Lineage Tool. It allows users to navigate graphically inside the ontology's taxonomies, open close node branches, search and compare classes and properties by their labels, explore linked nodes and edit all metadata. Those tools allow also to create and modify nodes and relations in a controlled process. These technical functions help to evaluate the:

- completeness and precision of concepts (called here classes)
- quality of taxonomies organizing classes in hierarchical trees of concept
- richness of predefined properties describing the intrinsic semantic links between classes
- richness of metadata associated with classes: definitions, synonyms, links to other standards…

4.4 ABOX data mapping to Standard TBOX

To convert tabular data into knowledge graphs and setup the enterprise ontology models, the first step is to map the words contained in the data to classes of the standard using labels. Theoretically the content of a table column should be semantically consistent thus mapping the whole column of a standard class should be enough. Unfortunately, it is not always the case because the semantic and the content of relational data models is much less precise and ambiguous compared to the fine and smart semantics of the standards.

So, for table columns containing terms, it is necessary to analyze the semantic context of each value while mapping them to relevant standard classes.

This task is achieved in SLSV using the Standardizer tool coupled to KGMappings tool schematized in figure 6. the Standardizer tool extracts distinct values from table columns containing terms (not numbers) and tries to align them with the labels (names) of Classes from several standards. For terms without exact match the Standardizer tool has a search engine to find fuzzy matches in the standards or even search capability to perform a manual mapping.

It should be noticed that the effort to finalize vocabulary mapping will be drastically reduced for future data integration belonging to the same business domain.

The smartest standard ontologies contain not only class taxonomies but also properties and constraints that link classes together by design. Using these predefined links between reference classes (technically domains, ranges, and restrictions) it is possible to infer many triples of the knowledge graph just by using the transitive link between data typed to classes, themselves being linked in the ontology TBOX [16] with some properties. This mechanism ensures the consistency of the knowledge graph and allows a processable quality control between requirements as defined in the standards and the reality of data.

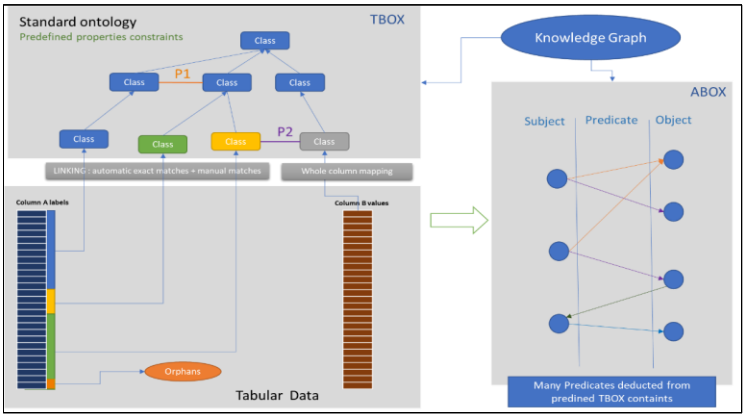

Figure 6 – Using transitivity between data, classes, and ontology constraints to generate efficient and consistent knowledge graphs

4.5 Knowledge graph construction and Management

Once the mapping is achieved, it is time to generate the knowledge graph. This task is done automatically by the KGgenerator tool that takes the data, the mappings and the dictionaries and produces sets of triples, concrete form of the knowledge graph. SLSV has also prototypes the KGbrowser tool to allow navigation through and perform complex queries in the knowledge graphs combining graph traversal algorithm and automatically generated SPARQL queries, as presented in the figure 7 below.

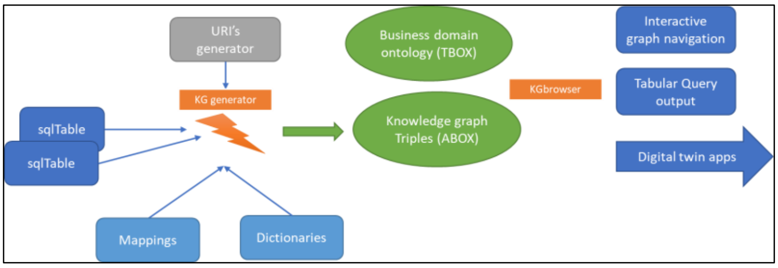

Figure 7 – From tabular data to standardized knowledge graph generation and exploitation

5 Discussion and conclusion

The development and use of the TotalEnergies Semantic Framework are an innovation, which faces different challenges. Technical ones should be overcome through projects proving the value of the methodology and of the chosen technology, its scalability for the big volume of data of a real asset. We observe that the main remaining challenge is a social one. Innovation needs support, trust, protection from skepticism to grow progressively and safely. Thus, the social aspect appears to be key to motivate teams to carry out the needed efforts required for a successful digital transition.

Our main motivation for developing methodology and tools managing together industry standards, semantic web, and data, is that it is probably the most efficient and promising way to link complex knowledge and data for existing assets. This shapes future information systems like digital twins in a context of interoperability and to support patrimony information structuration and knowledge capture in a coherent way.

This change of mindset supports a progressive participation of the stakeholders for connecting our data and taxonomies to higher level conceptual common nodes. We believe that this is the key to remove bottlenecks in our fragmented information system. A new generation of tools adapted to industrial requirements in terms of functionalities and performances are being developed and are interoperable thanks to their standards compliance.

This approach needs further co-development, cooperation of the internal and external actors through collective intelligence processes to bring coherence and resilience to the industrial ecosystems.

References

1. ISO 15926-1:2004, Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 1: Overview and fundamental principles.
2. What's next in Standards and Standards Publishing at ISO and IEC: https://www.typefi.com/standards-symposium-2021/whats-next-in-standards-publishing-iso-iec/ Consulted in December 2021
3. Ontocommons, Ontology-Driven Data Documentation for Industry Commons: https://ontocommons.eu
4. StandICT: https://www.standict.eu
5. Information Management: A proposal, The original document of Tim Berners Lee; March 1989: https://www.w3.org/History/1989/proposal.html
6. ISO/CD TR 15926-14: Industrial automation systems and integration - Integration of life-cycle data for process plants including oil and gas production facilities - Part 14: Data model adapted for OWL2 Direct Semantics
7. ISO/IEC 81346-1: 2009 - Industrial systems, installations and equipment and industrial products - Structuring principles and reference designations - Part 1: Basic rules
8. Maintenance terminology standards: some issues and the need of a shared framework for interoperability, Yves KERARON et Antoine DESPUJOLS, I-ESA 2020, Interoperability for Maintenance Workshop, November 2021
9. Systems Engineering as the foundation for industrial domain ontologies, Jinzhi Lu, Yves Keraron, David Cameron, Barry Smith, Dimitris Kiritsis. SemWeb.Pro 2021, December 2021
10. READI JIP - IMF, Asset Information Model Framework: https://readi-jip.org/asset-information-modelling-framework
11. READI JIP - Shaping the Future of Digital Requirements and Information Flow in the Oil and Gas Value Chain: https://readi-jip.org
12. ISO/TS 81346-10:2015 - Industrial systems, installations and equipment and industrial products - Structuring principles and reference designation — Part 10: Power plants
13. W3C SKOS – Simple Knowledge Organization System: https://www.w3.org/TR/skos-reference/
14. W3C OWL 2 – Web Ontology Language: https://www.w3.org/TR/owl2-primer/
15. SousLeSensVocables under MIT license: https://github.com/souslesens/souslesensVocables
16. TBox and ABox reasoning in expressive description logics: https://www.aaai.org/Papers/Workshops/1996/WS-96-05/WS96-05-004.pdf