PeTWIN: An ontology-supported data access for petroleum production digital twin

Mara Abel[1], João Cesar Netto[1], Fabrício Rodrigues[1], Nicolau Santos[1], Rafael Petry[1], Haroldo Rojas de Souza Silva[1],

Institute of Informatics UFRGS (mara.abel, joão.netto, fabricio.rodrigues, nicolau.santos,rafael.petry, haroldo.silva@inf.ufrgs.br),

A digital twin is a system framework tightly attached to a physical production plant conceived for monitoring its operation in real time. This framework integrates data from distinct sources and supports data analytics and predictive evaluation of the petroleum flow and the maintenance schedule. The information that feeds a modern oil-field digital twin integrates temporal series and static data, configuring a complex scenario for data analytics. The data is usually spread across many applications from several service companies that perform specific tasks during operations. These systems exchange data in distinct (sometimes proprietary) formats. Even when accessed in open formats, an integrated digital twin requires uniformization in data meaning, formats, units of measure, the scale of analysis, and interval of time associated with the track of data provenience. The integrated operation center receives such data hand-labeled with its source and meaning and analyses them to support short-term decisions. The PeTWIN project is a 4-year (2020-2024) joint cooperation between University of Oslo (UiO) and Federal University of Rio Grande do Sul (UFRGS), with the participation of Libra Consortium (Brazil), Equinor (Norway), and Shell (Norway), whose goal is defining the best practices for the development of digital twins for petroleum production plants.

One contribution of the UFRGS team to the project was developing a well-founded domain ontology to document the meaning and logical restriction of the assets involved in petroleum production and facility maintenance, which structures the framework for the semantic interoperability of data operated by the digital twin. We proposed the O3PO – *Offshore Petroleum Production Plant Ontology* (Santos et al. 2024), a domain ontology covering a specific part of the petroleum production process. The ontology development started with an extensive survey of existent standards and ontologies in the petroleum industry domain, which included resources such as the ISO 15926 standard, the semantic artifacts from OSDU - Open Subsurface Data Universe Forum platform (OSDU 2020), the standard glossaries of the Professional Petroleum Data Management (PPDM) *What is a Well?* (PPDM 2014) and *What is a facility?* (PPDM 2022), the interoperability standard PRODML (King et al., 2012), and the equipment specification of CFIHOS (IOGP 2020). The analysis of these resources was followed by a sequence of interviews with petroleum industry professionals from subsea, reservoir, flow maintenance, production, and integrated process monitoring areas. The interviews produced a list of digital twin software functional requirements and raised the terminology adopted for system and data labeling that drives the domain ontology usage. O3PO specializes the Basic Formal Ontology (BFO) top ontology, which eases alignment with previously developed BFO-derived domain ontologies, guaranteeing a common conceptual basis for integration. The ontology formalizes the definitions of the relevant entities from the domain, allowing the representation of production plant installation assets and their qualities, relationships, and other associated dependent entities.

The O3PO ontology grounds the semantic framework for the architecture of the developed digital twin, which provides several semantic capabilities to the software applications composing the digital twin. Besides it, the semantic framework also includes an RDF knowledge graph, built based on the referred ontology, composed of hundreds of instances corresponding to entities from several real petroleum production plants. This knowledge graph contemplates wells and their parts (e.g., valves, tubing sections, wellhead, annular spaces), the artifacts that compose the pipelines linking wells to platforms (e.g., manifolds, Christmas trees, risers), and various equipment (e.g., pumps, compressors). On top of that, the graph also includes instances corresponding to key measured properties of different artifacts (e.g., pressure, temperature, flow rate, valve aperture) as well as relationships of interest in the domain (e.g., hydraulic connections, functional partonomy). Completing the framework, we have a set of APIs and software artifacts that connect the ontology and the graph to other applications.

We applied this framework for several tasks, such as retrieving data from historian systems and extracting contextual information about the production plant installation. It was used both in the development of in-house applications and in the adoption of cloud digital twin solutions. One in-house solution, focused on the visual exploration of assets, uses the knowledge embedded in the graph to find the appropriate time-series data in a historian system while also using it to provide more context to the user. With that, the user can select types of entities and/or particular instances of such types and, based on this selection, the application retrieves the data associated with the chosen entities directly from the data storage, including references to time series in the historian system describing the properties of interest (Fig. 1[a]). This solution also integrates asset information in a P&ID (Piping and Instrumentation Diagram) interface to explore data in a visual context representing the structure of the plant (Fig. 1[b]). We have conceived the whole solution as a microservice architecture (Fig. 1[c]), in which each component plays a specific task customized to the environment and software platform. This approach allows the easy adaptation of the solution for other ontology-based tasks and platforms.

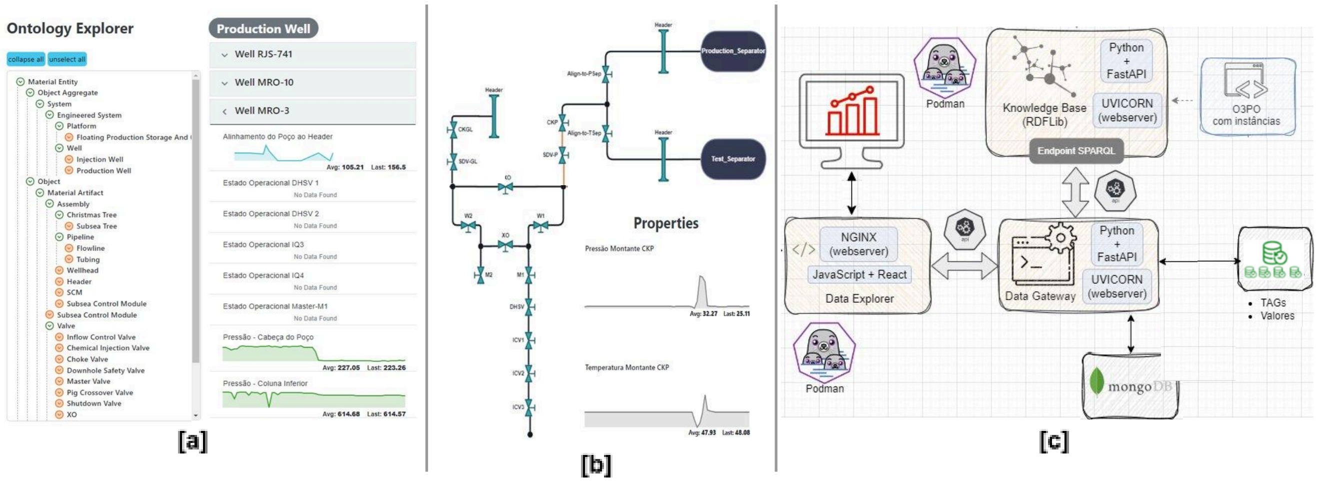

Figure 1: (a) interface of the ontology explorer application (Santos et al., 2023) providing navigation in the ontology taxonomy, exploration of components and properties of selected entities and time-series plots; (b) P&ID interface allowing selection of diagram elements and visualization of associated data; (c) schematic view of digital twin architecture.

Also, the developed semantic framework enables us to map the data from/into third-party data models to enhance the potential of applications developed by other providers, provided that the data domain is totally or partially covered by the O3PO ontology. We have developed applications for data retrieval from cloud-based repositories by integrating our framework with embedded services and data-model cloud providers, such as Amazon AWS and Microsoft Azure, with the semantic capabilities of the resulting applications being limited to the expressivity of the provider's data model.

The PeTWIN project investigated and proposed good practices in digital twin development, especially regarding the semantic integration of data. This was the result of an extensive study of a number of necessities related to the conception of digital twins for the petroleum production domain. We collected the software requirements for the digital twin and constructed business process models, connecting such activities to corporate information systems. We studied current data management platforms (Correia et al., 2023; Correia et al., 2022) and private cloud services (Knebel et al., 2023). We developed machine learning applications for failure diagnosis (Schena, 2023) and production forecasting (Pivetta et al., 2023). Further information about the project can be found at www.petwin.org. The ontology specification and related OWL artifacts are available at https://github.com/orgs/BDI-UFRGS.

Acknowledgement: CAPES Finance Code 001, FINEP, Libra Consortium (Petrobras, Shell Brasil, Total Energies, CNOOC, CNPC).

References

Correia, J. B., Abel, M. and Becker, K. (2023). "Data management in digital twins: a systematic literature review." Knowledge and Information Systems: 1-32.

Correia, J. B., Rodrigues, F. H., Santos, N., Abel, M., and Becker, K. (2022). "Data management in Digital Twins for the Oil and Gas Industry: beyond the OSDU Data Platform." Journal of Information and Data Management **13**(3).

IOGP (2020). Capital Facilities Information Handover Specification (CFIHOS) - Specification Document, International

Association of Oil & Gas Producers.

King, M. J., Ballin, P. R., Bennis, C., Heath, D. E., Hiebert, A. D., McKenzie, W., Rainaud, J.-F., and Schey, J. C. S. P. E. (2012). "Reservoir Modeling: From RESCUE To RESQML." SPE Reservoir Evaluation & Engineering.

Knebel, F. P., Trevisan, R., do Nascimento, G. S., Abel, M., and Wickboldt, J. A. (2023). "A study on cloud and edge computing for the implementation of digital twins in the Oil & Gas industries." Computers & Industrial Engineering **182**: 109363.

OSDU, T. O. G. (2020). The Open Group Guide OSDU™ System Concept, The Open Group.

Pivetta, M. V. L., Simon, A. H., Costa, M. M., Abel, M., and Carbonera, J. L. (2023). A Systematic Evaluation of Machine Learning Approaches for Petroleum Production Forecasting. IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI). Nevada, IEEE.

PPDM (2014). What is a well? Houston, PPDM Association.

PPDM (2022). What is a facility? Houston, PPDM Association.

Santos, N., Rodrigues, F. H., Schmidt, D., Romeu, R., Nascimento, G., and Abel, M. (2024). "O3PO: A Domain Ontology for Offshore Petroleum Production Plants." Expert Systems with Applications **238**(122104).

Santos, N. O., Rivera, J., Petry, R., Rodrigues, F. H., Nascimento, G., Comba, J., and Abel, M. (2023). Ontology Explorer: An Ontology-Based Visual Analytics System for Exploring Time Series Data in Oil and Gas. Formal Ontology in Information Systems. N. Aussenac-Gilles and T. Hahmann. Sherbrooke, Canada IOS.

Schena, R. (2023). A methodology for synthetic generation of failure data for data-driven prognostics and health management modeling for digital twins. Master Dissertation, UFRGS.