



**XIX Simpósio Brasileiro de  
Sistemas de Informação**

**INSIDE: an Ontology-based Data Integration  
System Applied to the Oil and Gas Sector**

Júlio G. Campos, Vitor P. de Almeida, Elvismary M. de Armas, Geiza M.  
Hamazaki da Silva, Eduardo T. Corseuil, Fernando R. Gonzalez

UNIRIO, PUC-Rio, Petrobras

# Introdução

---

- Apesar dos avanços feitos nas últimas décadas, a integração de dados continua sendo um grande desafio enfrentado pelas organizações na era da informação.
- Os problemas mais difíceis de integração de dados são causados por dados com **semântica heterogênea**, que impedem que diferentes sistemas de informação consigam compartilhar informações entre si com entendimento comum do significado. (ZIEGLER, 2004)

# Problema

---

As empresas do setor de óleo e gás se deparam com vários problemas para integrar dados:

- Como lidar com grandes volumes de dados?
- Como diferenciar dados que possuem mais de um significado?
- Que padrão para representação e troca de dados deve ser adotado?
- É possível desenvolver soluções de código aberto para lidar com semântica, ciclo de vida, interoperabilidade e escalabilidade?
- Como diminuir a dependência tecnológica de softwares comerciais?

# Solução

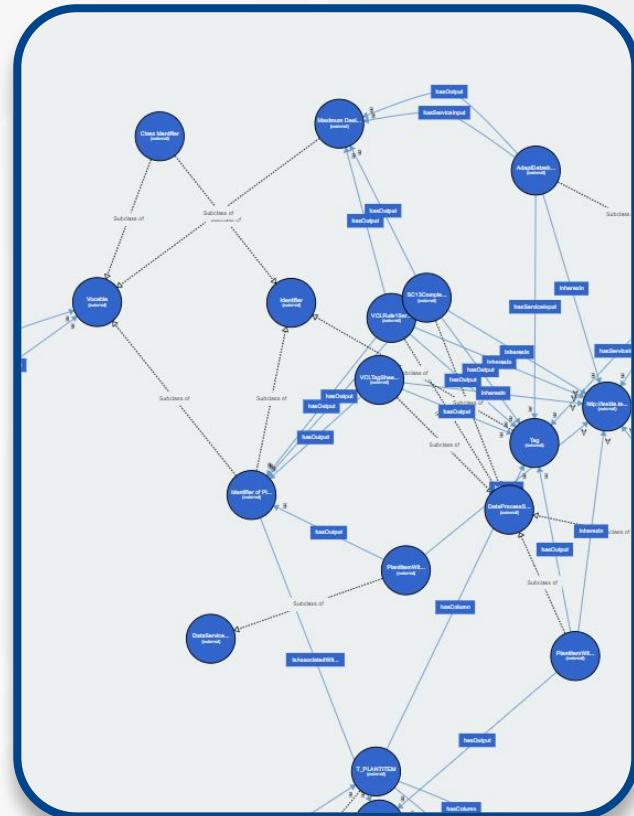
---

INSIDE - INteroperabilidade Semântica para Integração de Dados de Engenharia  
- um sistema de integração de dados orientado a **ontologia** que tem como princípios básicos:

- Integração de bases de dados estruturadas e heterogêneas;
- Representação formal de conhecimento para incorporar significado aos dados e alcançar a interoperabilidade semântica entre sistemas de informação;
- Utilização de tecnologias de código aberto e escaláveis.

# O que é uma ontologia?

- É uma **especificação formal explícita** de uma conceitualização compartilhada. Os conceitos representados são símbolos que podem ser compreendidos por humanos e decodificados por máquinas (GRUBER, 1993).
  - A **OWL** é uma linguagem de ontologia Web criada para adicionar significado a dados. Permite definir relações entre conceitos e especificar regras lógicas para raciocinar sobre eles.



# Metodologia

---



## Revisão do Estado da Arte

Revisão da literatura sobre arquiteturas de sistemas para integração de dados entre 2015-2020. A maioria é baseada em ontologia.



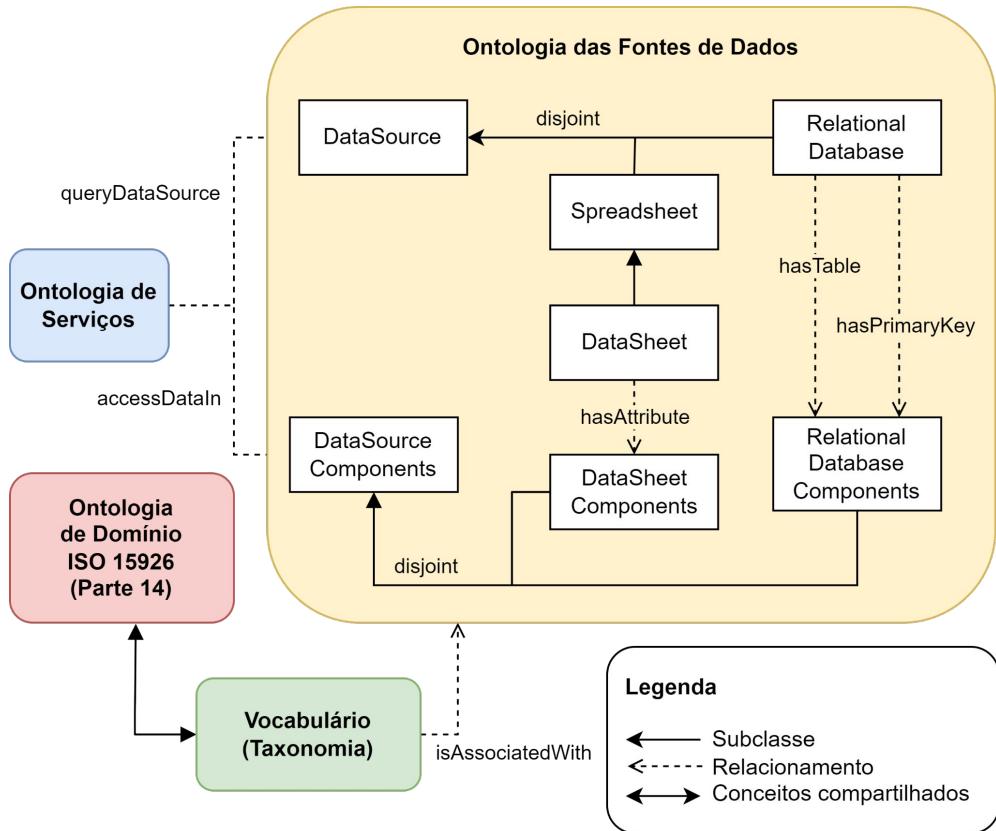
## Representação de Conhecimento

Elaboração de três ontologias: domínio, fontes de dados e serviços de dados.



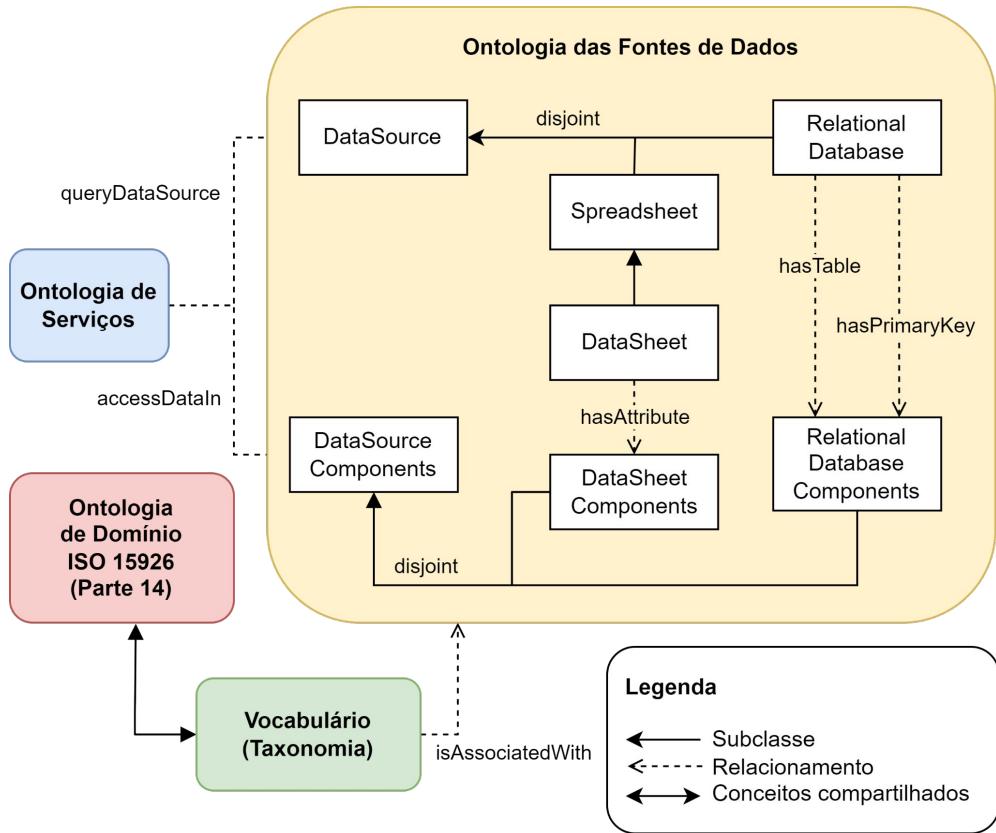
## Arquitetura e Desenvolvimento

Elaboração de uma arquitetura de sistema e desenvolvimento com tecnologias de código aberto e escaláveis.

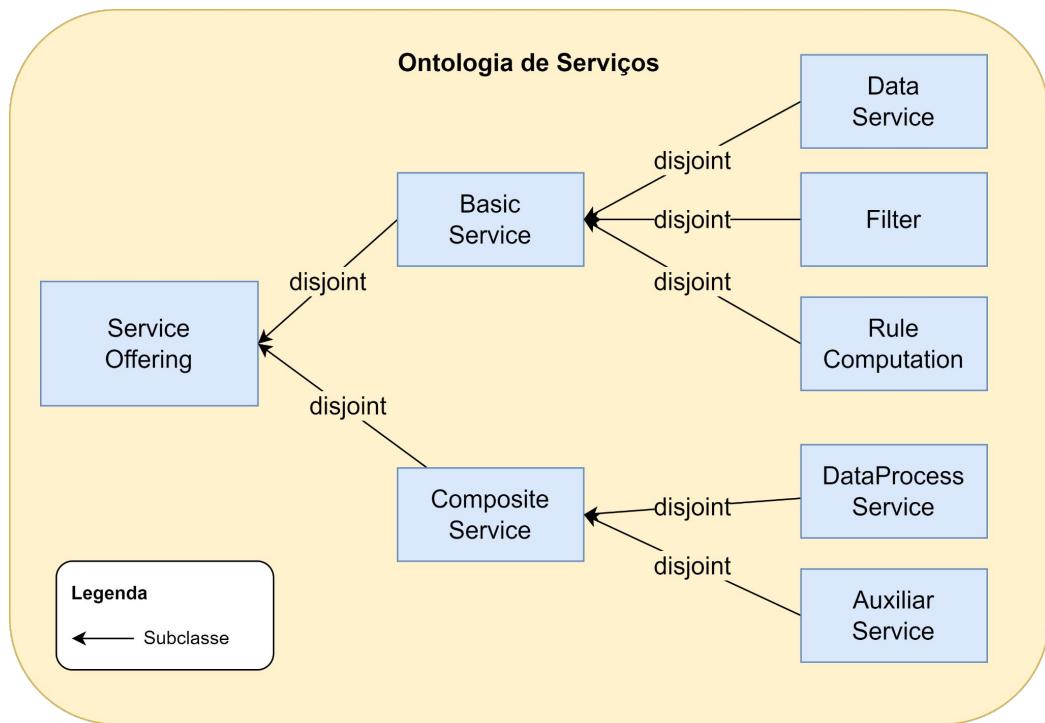


**Ontologia de domínio:** conceitos sobre fase de projeto de comissionamento de plantas industriais. Baseada na norma ISO 15926 (parte 14), padrão para representação de informações industriais incluindo instalações de óleo e gás.

**Ontologia das fontes de dados:** descreve e representa tabelas de banco de dados relacionais e folhas de dados, documentos técnicos em formato de planilhas.



**Vocabulário:** reúne conceitos que estão presentes na ontologia de domínio, atribuindo significado semântico a eles, além de termos usados no modelo conceitual das bases de dados ou no modelo do negócio.



## Ontologia de serviços de dados:

Baseada na UFO-S, ontologia de fundamentação para representar serviços em diversos domínios de conhecimento.

### Tipos de serviços:

Básico: realiza consulta apenas em uma base de dados. Inserido manualmente por um especialista em ontologia.

Composto: criado pelo usuário a partir da composição de serviços básicos ou outros serviços compostos.

# Arquitetura de Sistema

- Primeira arquitetura baseada na especificação OBDA (Ontology Based Data Access).
- Projetada para caso de estudo sobre componentes de elementos de uma planta industrial de uma empresa de óleo e gás.
- Trabalho publicado na iiWAS2021, Conferência Internacional sobre Integração de Informação e Inteligência Web (DE ARMAS et al., 2021).



ORGANIZAÇÃO  
INSTITUTO FEDERAL  
Alagoas

REALIZAÇÃO  
SBC  
Sociedade Brasileira  
de Computação

## Hybrid Architecture to Achieve Semantic Interoperability for Engineering Oil and Gas Industry Process

Elvimiry Molina de Armas  
Instituto Tegraf - Pontifícia  
Universidade Católica do Rio de  
Janeiro (PUC-RIO)  
Rio de Janeiro, RJ, Brasil  
elmodm@tegraf.puc-rio.br

Geiza Maria Hamazaki da Silva  
Universidade Federal do Estado Do  
Rio De Janeiro (UNIRIO)  
Rio de Janeiro, RJ, Brasil  
geiza.hamazaki@unirio.br

Eduardo Thadeu Leite Corseuil  
Instituto Tegraf - Pontifícia  
Universidade Católica do Rio de  
Janeiro (PUC-RIO)  
Rio de Janeiro, RJ, Brasil  
thadeu@tegraf.puc-rio.br

Denyson Tomaz de Lima  
Petrobras S.A. Rio de Janeiro  
Rio de Janeiro, RJ, Brasil  
denyson@petrobras.com.br

Rodrigo G. Guzmão Caiado

Instituto Tegraf - Pontifícia  
Universidade Católica do Rio de  
Janeiro (PUC-RIO)  
Rio de Janeiro, RJ, Brasil  
rodrigocaiado@tegraf.puc-rio.br

Fernando Rodrigues Gonçalvez  
Petrobras S.A. Rio de Janeiro  
Rio de Janeiro, RJ, Brasil  
fernando.goncalvez@petrobras.com.br

Vitor Pinheiro de Almeida  
Instituto Tegraf - Pontifícia  
Universidade Católica do Rio de  
Janeiro (PUC-RIO)  
Rio de Janeiro, RJ, Brasil  
valmeds@tegraf.puc-rio.br

Hugo Neves  
Instituto Tegraf - Pontifícia  
Universidade Católica do Rio de  
Janeiro (PUC-RIO)  
Rio de Janeiro, RJ, Brasil  
juliuscampone@tegraf.puc-rio.br

Geiza Maria Hamazaki da Silva, Rodrigo G. Guzmão Caiado, Hugo Neves, Fernando Rodrigues Gonçalvez, Denyson Tomaz de Lima, and the task Rodriguez Gonçalvez, 2021, Hybrid Architecture to Achieve Semantic Interoperability for Engineering Oil and Gas Industry Process. In the 2nd International Conference on Web Services and Applications for Business Intelligence (IWAS2021), November 29 December , L2B U.Sa, Ankara, AÇMA, NY, USA, 7 pages. <https://doi.org/10.31949/978-6154-0077-02>

KEYWORDS

Semantic interoperability, Architecture, Polystore, Micro-services, Oil and Gas industry

ACM Reference Format:

Elvimiry Molina de Armas, Vitor Pinheiro de Almeida, juan-pedro campone, Geiza Maria Hamazaki da Silva, Rodrigo G. Guzmão Caiado, Hugo Neves, Fernando Rodrigues Gonçalvez, Denyson Tomaz de Lima, and the task Rodriguez Gonçalvez, 2021, Hybrid Architecture to Achieve Semantic Interoperability for Engineering Oil and Gas Industry Process. In the 2nd International Conference on Web Services and Applications for Business Intelligence (IWAS2021), November 29 December , L2B U.Sa, Ankara, AÇMA, NY, USA, 7 pages. <https://doi.org/10.31949/978-6154-0077-02>

## 1 INTRODUCTION

Making information available in an essential objective for the scientific and industrial community. However, it is necessary to overcome challenges related to the integration of data from different sources. In large companies, it is common to use different software to achieve a common goal. The company's ability to search for specific data and identify inconsistencies between the data already produced is crucial and can bring operational and financial benefits. As such, when data are stored in different formats, these tasks become unfeasible without data integration.

In the context of the oil and gas industry, the picture does not look better. From land exploration to deep water, the oil and gas industry is over 80 years old in Brazil. As a result, much information is scattered across various systems, which have different models and structures. Among the challenges related to interoperability in this industry, it is possible to find dynamic data growing in time with variable detail levels and a high data volume that needs scalable

# Arquitetura de Sistema

- Arquitetura atual, orientada à ontologia.
- Extensão do estudo de caso anterior, para verificar quais equipamentos do tipo vaso de pressão se enquadram na Norma Regulamentadora No. 13 (NR-13)
- Trabalho publicado na ICEIS 2023, Conferência Internacional sobre Sistemas de Informação Empresarial (ALMEIDA et al., 2023)



ORGANIZAÇÃO  
INSTITUTO FEDERAL  
Alagoas

REALIZAÇÃO  
SBC  
Sociedade Brasileira  
de Computação

## INSIDE: Semantic Interoperability in Engineering Data Integration

Vitor Pinheiro de Almeida<sup>1</sup>, Júlio Gonçalves Campos<sup>2</sup>, Elvismary Molina de Armas<sup>3</sup>, Geiza Maria Humazaki da Silva<sup>2</sup>, Hugo Neves<sup>1</sup>, Eduardo Thadeu Leite Corseuil<sup>1</sup>, Fernando Rodrigues Gonzalez<sup>2</sup>

<sup>1</sup>Instituto Tecnológico Pontifícia Universidade Católica do Rio de Janeiro (PUC-RIO), Brazil

<sup>2</sup>Universidade Federal Do Estado Do Rio De Janeiro (UNIRIO), Brazil

<sup>3</sup>Petrobras S.A., Rio de Janeiro, Brazil

Keywords: Data Integration, Semantic Web, Information Systems, Semantic Interoperability.

### Abstract:

One of the critical problems in the industry is data integration. Data is generated continuously and on a large scale and is present using different formats, software, and terminologies. Integrating multiple databases belonging to different information systems can provide a unified data view and a better understanding of the data. With that in mind, this paper present INSIDE, a system that enables Semantic Interoperability for Engineering Data Integration. INSIDE represents queries to one or multiple databases through the concept of data services, where each service is defined using an ontology. Data services can conceptually represent the commitment and claims between service providers (databases) and service customers (users). The service lifecycle the process of querying, integrating, and delivering data. The contributions of this paper are the following: (1) Use of formal mechanisms for semantic data representation with the connection with the intratransactional community; (2) A conceptual model for a distributed system based on ontologies for querying and manipulating data from multiple data sources; (3) An implementation of this model, called INSIDE, developed on top of Apache Spark; and (4) An experimental evaluation of the service composition strategy of INSIDE for data integration of multiple data sources using a real-world scenario.

SCIE INSIDE: Semantic Interoperability in Engineering Data Integration

## 1 INTRODUCTION

Data is the most valuable asset for any organization. However, as much as it is needed for making important business decisions, most organizations still lack a coherent and efficient approach to data integration. In the context of the oil and gas industry, the scenario is similar (Ate Gülla, 2008). Data integration in this area presents problems such as: (1) Data heterogeneity; (2) Dynamic data with varying levels of detail because engineering systems are constantly changing and have a lifecycle of more than 40 years; (2) Need for scalable solutions for big data. An approach to

- <https://ieeic.org/0000-0002-6594-9541>
- <https://ieeic.org/0000-0002-5923-3836>
- <https://ieeic.org/0000-0002-5923-3894>
- <https://ieeic.org/0000-0002-7554-2413>
- <https://ieeic.org/0000-0002-6834-2654>
- <https://ieeic.org/0000-0002-7543-4140>
- <https://ieeic.org/0000-0002-7440-9509>

data integration for oil and gas needs to be scalable, fast, and reliable; handling data heterogeneity and produce results in acceptable times; (3) Lack of knowledge of the system's native query language to explore data from databases. This requirement addresses the importance of providing an interface to interact with the integrated data for users who do not know how to formulate database queries; and (4) Assistance of domain specific data representation. This, in most cases, requires the knowledge of a domain expert. Thus, for an oil and gas data integration approach, it is necessary to provide a way to make it easier for the user to map schemas.

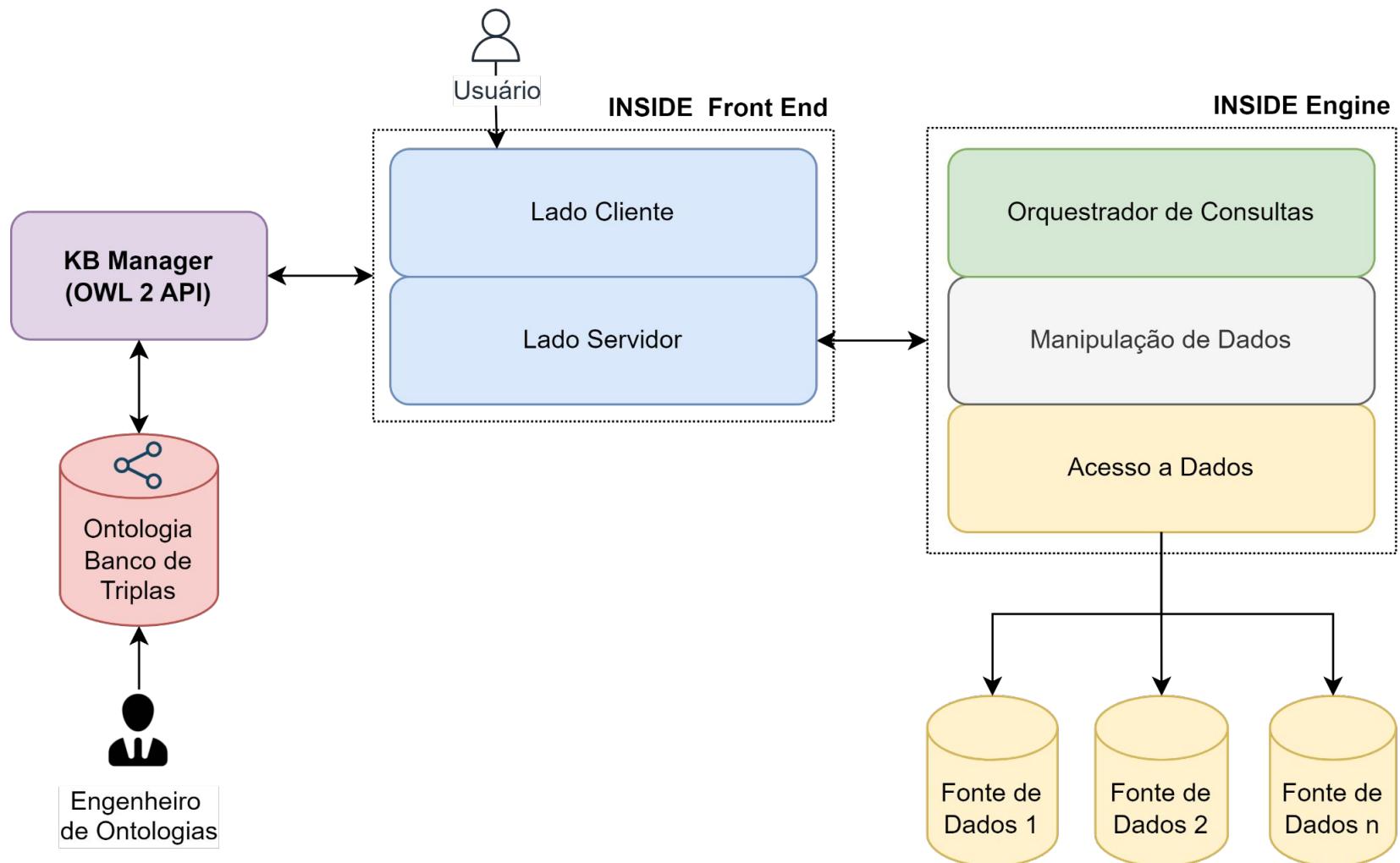
With that in mind, the Semantic Interoperability for Engineering Data Integration (INSIDE) project is presented. INSIDE enables the representation of queries for one or multiple databases through the concept of data services, where each service is defined using an ontology. Data services can conceptually represent the commitments and claims between service providers (databases) and service customers (users).

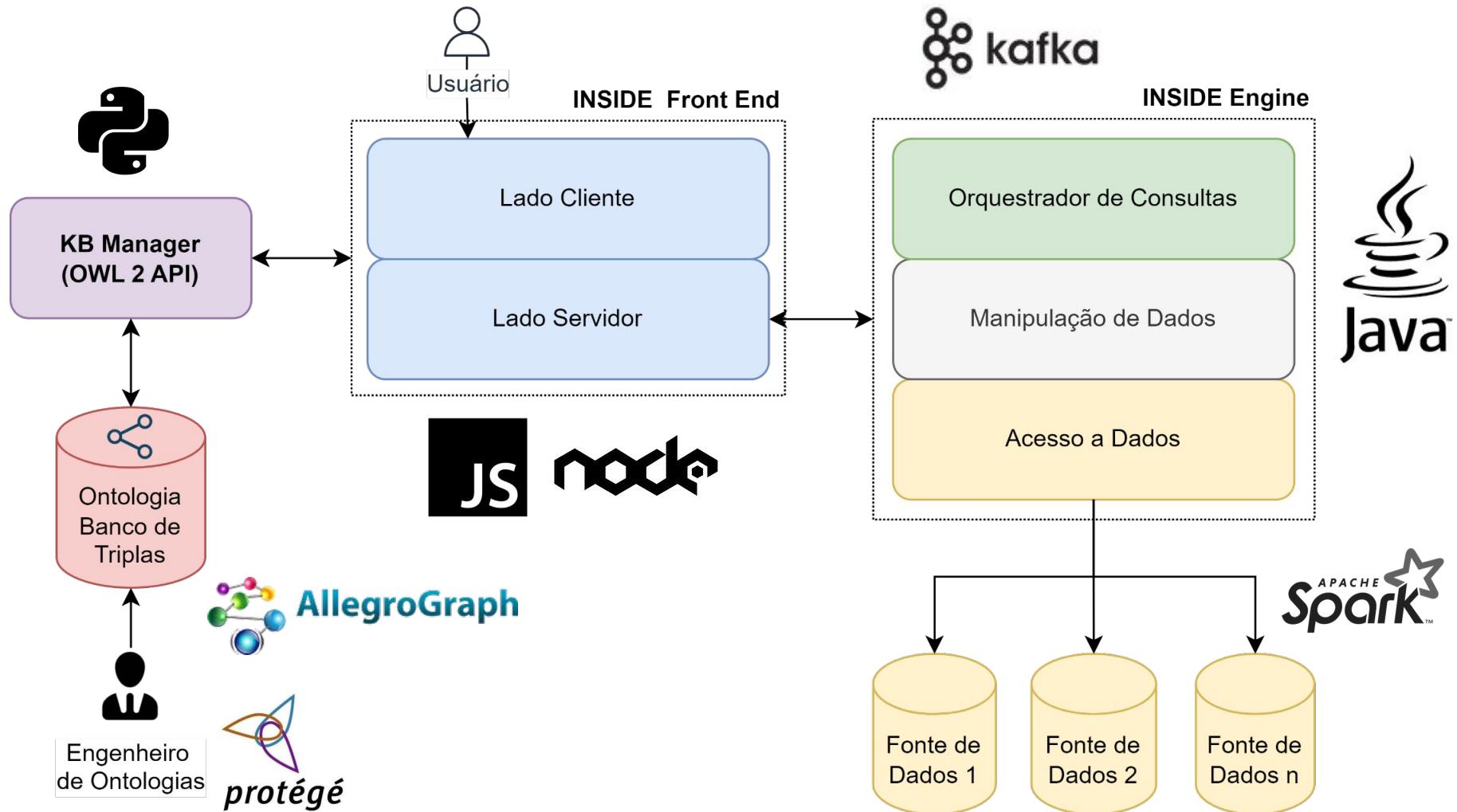
107

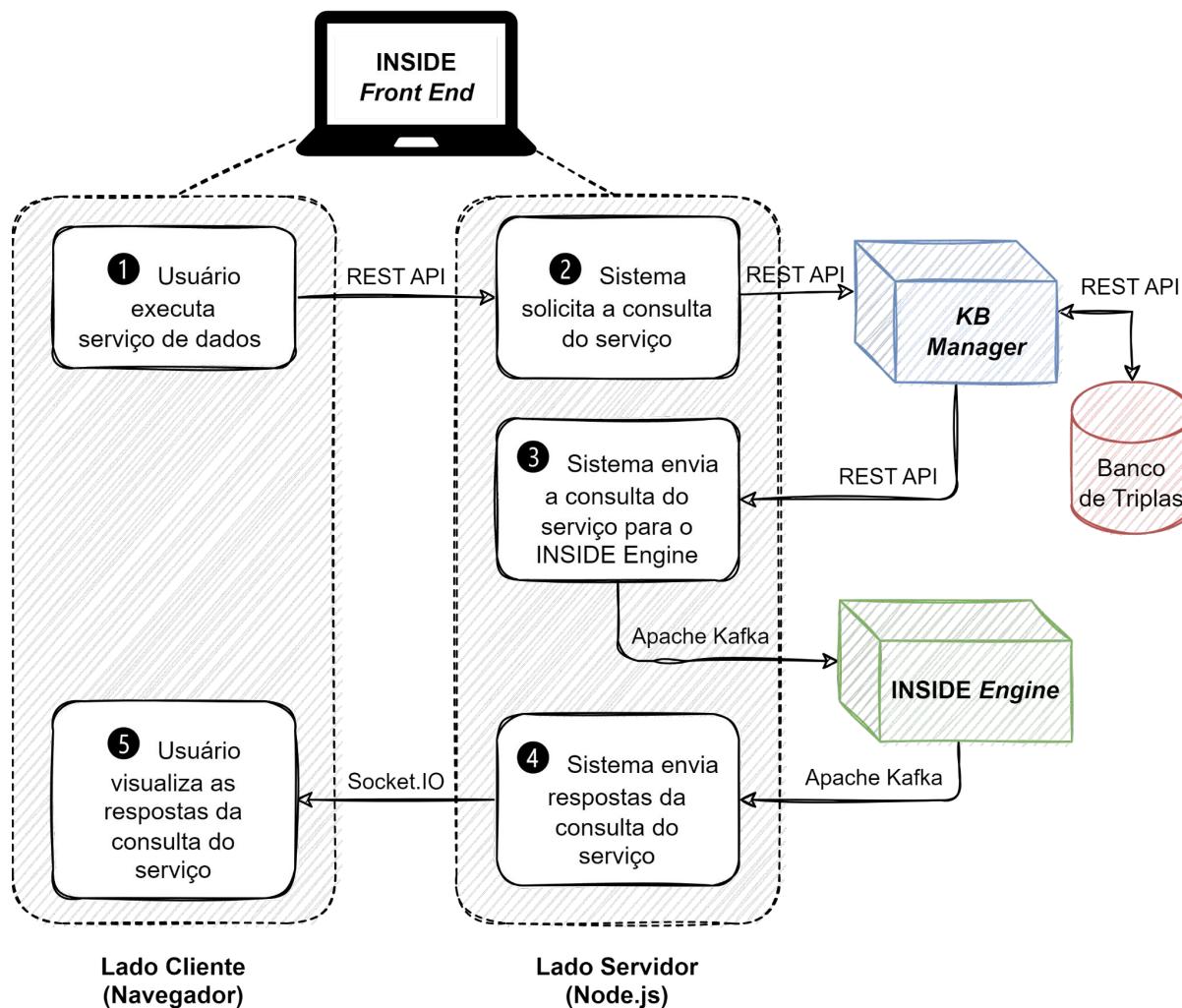
Almeida, V., Campos, E., Molina de Armas, F., Humazaki da Silva, G., Neves, H., Corseuil, E. and Gonzalez, F.  
INSIDE: Semantic Interoperability in Engineering Data Integration.  
In: Proceedings of the International Conference on Enterprise Information Systems (ICEIS 2023). Volume 1, pages 107-114.  
ISBN: 978-989-9740-68-4. DOI: 10.5120/iceis2023-0002

sbsi2023.ifal.edu.br

@ sbsibrasil







## Dashboard

## Data Dictionary

ISO 15926



37943 terms

Consult Terms



37995 terms

INSIDE



52 terms

## Data Services

Run Services



31 available services

Create Composite Service



Generate new services

## CFIROS

Run CFIROS Services



8 available services

CFIROS Validator



Validate an Excel file

INSIDE

Semantic Interoperability in  
Engineering Data Integration

## Data Services

Enter a term

Clean

1

Cases types list

**IRI:** CaseTypesList.**Definition:** This service returns the case classes with their codes and names.**Outputs:** (4)

- o Identifier of Plant Item
- o Operation Case Type Identifier
- o Project Case Class Code Identifier
- o Project Case Type Text



Run service

2

Cases class code list

**IRI:** CasesClassCodesList.**Definition:** This service returns the codes that correspond to case classes.**Outputs:** (1)

3

Cases types code list

**IRI:** CasesTypesCodesList.**Definition:** This service returns the codes that correspond to the types of cases.**Outputs:** (1)

4

SubSystems from the Model Item

**IRI:** CommissioningSubSystemCodeList.**Definition:** This service returns the SubSystems (SSOPs) from the Model Item value.**Outputs:** (3)**IRI:** CommissioningSubSystemEnumerationIdList.

```
{
    "_isGroupQuery": false,
    "_renameDataServiceAttributes": null,
    "_renameDataService": null,
    "_clientQueryID": "1581ccb4918cbd70ce3c019c19b7ecf5",
    "_clientQueryName": "This service returns the plant items with operating pressure information with their \"operation\" parameter. It joins all the fields from the T_CASE table and the fields from the T_PLANTITEM table that have the same ID. The filterT_CASE.CASETYPE== 2 (\"operation\") is applied.",
    "_dataServices": [
        {
            "type": "DataService_RelationalDB",
            "properties": {
                "_query": "select distinct PI.SP_ID as PlantItemIdentifier, CA_OPER.CASETYPE as OperationCaseTypeIdentifier, CA_OPER.CASECLASS as OperationCaseClassIdentifier, CA_OPER.SP_ID as PlantItemIdentifier_1 from P74TSPID.T_PLANTITEM PI left join P74TSPID.T_CASE CA_OPER on CA_OPER.SP_MODELITEMID = PI.SP_ID and CA_OPER.CASETYPE = 2",
                "_stringToFilter": null,
                "_attributeToApplyFilter": null,
                "_expression": "",
                "_columns": null,
                "_relationalBD_info": {
                    "_connectionURL": "jdbc:oracle:thin:@orclcae.tecgraf.puc-rio.br:1521:ORCL",
                    "_userBD": "proteus4",
                    "_passwordBD": "proteus4",
                    "_driverClass": "oracle.jdbc.driver.OracleDriver",
                    "_driverFormat": "jdbc",
                    "_databaseFileName": null,
                    "_databaseType": null
                },
                "_serviceID": "PlantItemWithOperatingPressureInformation",
                "_clientQueryID": "1581ccb4918cbd70ce3c019c19b7ecf5"
            }
        }
    ],
    "_serviceIdToGetResult": "PlantItemWithOperatingPressureInformation"
}
}
```

Actions

Endpoint	Function	Methods
/	index()	OPTIONS / GET / HEAD /
/aliasequivalent	aliasEquivalent()	OPTIONS / GET / HEAD /
/baseiri	baseiri()	OPTIONS / GET / HEAD /
/cfhosservicelist	cfhosservicelist()	OPTIONS / POST /
/cfhosservices	cfhosservices()	OPTIONS / GET / HEAD /
/classes	classes()	OPTIONS / GET / HEAD /
/columnlabels	columnlabels()	OPTIONS / GET / HEAD /
/compositeservice	compositeservice()	OPTIONS / GET / HEAD /
/compositeservices	compositeservices()	OPTIONS / GET / HEAD /
/counttriplesdictionary	counttriplesdictionary()	OPTIONS / GET / HEAD /
/counttriplesontologydictionary	counttriplesontologydictionary()	OPTIONS / GET / HEAD /
/counttriplesofcfhosservices	counttriplesofcfhosservices()	OPTIONS / GET / HEAD /
/counttripleservices	counttripleservices()	OPTIONS / GET / HEAD /
/counttriplesontology	counttriplesontology()	OPTIONS / GET / HEAD /
/dataservice	dataservice()	OPTIONS / GET / HEAD /
/dataservices	dataservices()	OPTIONS / GET / HEAD /
/dictionary	dictionary()	OPTIONS / GET / HEAD /
/dictionaryiso	dictionaryiso()	OPTIONS / GET / HEAD /
/dictionaryontology	dictionaryontology()	OPTIONS / GET / HEAD /
/dynamicservice	dynamicservice()	OPTIONS / GET / HEAD /
/dynamicservices	dynamicservices()	OPTIONS / GET / HEAD /
/filters	filters()	OPTIONS / GET / HEAD /
/insideiri	insideiri()	OPTIONS / GET / HEAD /
/isa	isa()	OPTIONS / GET / HEAD /
/page/sparql	pageSparql()	OPTIONS / GET / HEAD /
/prefixes	prefixes()	OPTIONS / GET / HEAD /
/rulecomputations	rulecomputations()	OPTIONS / GET / HEAD /
/savecompositeservice	savecompositeservice()	OPTIONS / POST /
/sparql	sparql()	OPTIONS / GET / HEAD /
/spreadsheetservice	spreadsheetservice()	OPTIONS / GET / HEAD /
/spreadsheetservices	spreadsheetservices()	OPTIONS / GET / HEAD /
/uploadontology	uploadontology()	OPTIONS / GET / HEAD /
/vdprojects	vdprojects()	OPTIONS / GET / HEAD /
/vcservice	vcservice()	OPTIONS / GET / HEAD /
/vcservices	vcservices()	OPTIONS / GET / HEAD /

 Data Services

29 - Composite Service 13

**Buntime:** 0:46 seconds. **Results:** 334.

Exp

**Close**

### Outputs: (3)

5

Enumerates identifiers for the commissioning subsystem

**IRI:** CommissioningSubSystemEnumerationIdList

**Definition:** This service returns the enumeration identifiers that correspond to the commissioning subsystem.

### Outputs: (1)

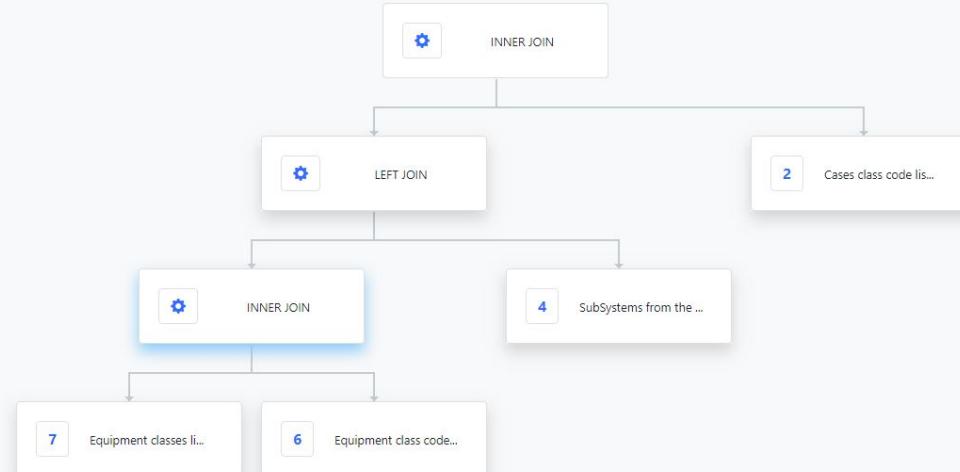
## New Composite Service

Joins Services

Enter a term  Clean

- 1 Cases types list
- 2 Cases class code list
- 3 Cases types code list
- 4 SubSystems from the Model Item
- 5 Enumerates identifiers for the commissioning subsystem
- 6 Equipment class codes list
- 7 Equipment classes list with codes and names
- 8 Equipment classes with code and name from equipment code
- 9 Equipment type codes list
- 10 Equipment types list with codes and names

Clean Save



Virtual Assistant

Unable to save composite service. The join highlighted in blue does not have two children or has not been configured.

## New Composite Service

Joins

Services

Enter a term

Clean

### INNER JOIN

Intersection with all data.

1 Cases types list

2 Cases class code list

3 Cases types code list

4 SubSystems from the Model Item

5 Enumerates identifiers for the commissioning subsystem

6 Equipment class codes list

7 Equipment classes list with codes and names

8 Equipment classes with code and name from equipment code

9 Equipment type codes list

#### Left Side

3 - Cases types code list

Project Case Class Code Identifier

#### Right Side

Intersection with all left data

Project Case Class Code Identifier

#### Filters

Select a filter

There are no filters available

#### Rule Computations

Generates Type Column Vessel

**Definition:** This service generates the type column with the value: "V/ASOSDEPRESSAO".

Close

Save

Clean

Save

# Conclusão

---

O sistema INSIDE contribui para atender necessidades inerentes do setor de óleo e gás:

- Permite que usuários realizem e criem consultas a partir de um vocabulário conhecido, sem necessidade de conhecimento de linguagens de consultas de bases de dados (SQL) e ontologias (SPARQL);
- Possibilita a virtualização de dados, integrando dados sem movê-los de seus repositórios de origem;
- Possibilita a representação formal de conhecimento de especialistas, permitindo adicionar significado a dados, criar relações e raciocinar sobre eles;

# Conclusão

---

- Compartilhar dados com significado comum com outros sistemas de informação permitindo alcançar a interoperabilidade semântica;
- Uso de tecnologias de código aberto e escaláveis para suportar processamento de grandes volumes de dados de diferentes fontes.

Atualmente, estuda-se como o sistema INSIDE pode ser utilizado para extrair e transformar dados de acordo com padrões internacionais para transferência de informações entre empresas do setor de óleo e gás como o CFIHOS (Capital Facilities Information HandOver Specification).

# Referências

---

ALMEIDA, V. et al. INSIDE: Semantic Interoperability in Engineering Data Integration. In: **Proceedings of the 25th International Conference on Enterprise Information Systems**. Volume 1: ICEIS; 2023. p. 107–114.

CAMPOS, Júlio et al.. Estado da arte sobre arquiteturas de sistemas para integração de dados. In: **RIO OIL & GAS EXPO AND CONFERENCE**. IBP, 2020. n. 414.

DE ARMAS, Elvismary Molina et al. Hybrid Architecture to Achieve Semantic Interoperability for Engineering Oil and Gas Industry Process. In: **The 23rd International Conference on Information Integration and Web Intelligence**. 2021. p. 176–182.

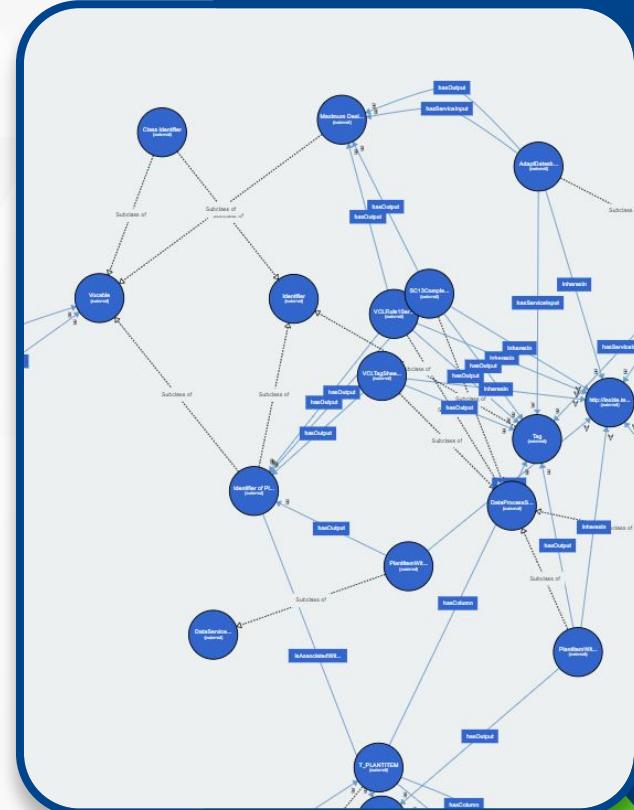
GRUBER, Thomas R. A translation approach to portable ontology specifications. **Knowledge acquisition**, v. 5, n. 2, p. 199–220, 1993.

ZIEGLER, Patrick; DITTRICH, Klaus R. Three decades of data integration—All problems solved?. In: **Building the Information Society: IFIP 18th World Computer Congress Topical Sessions 22–27 August 2004 Toulouse, France**. Springer US, 2004. p. 3–12.

# CONTATOS



juliogcampos@tecgraf.puc-rio.br





XIX Simpósio Brasileiro de  
Sistemas de Informação

# Obrigado!

**INSIDE: an Ontology-based Data Integration  
System Applied to the Oil and Gas Sector**

Júlio G. Campos, Vitor P. de Almeida, Elvismary M. de Armas,  
Geiza M. Hamazaki da Silva, Eduardo T. Corseuil, Fernando R. Gonzalez