# Raising the Role of Vocabulary Hubs for Semantic Data Interoperability in Dataspaces

3rd Workshop on Semantic Interoperability in Data Spaces

Robert David, Petar Ivanov, Vladimir Alexiev

2024-10-01

ontotext

SEMANTIC WEB COMPANY

UNDERPIN
DATA SPACE FOR MANUFACTURING

# Introduction

- The European data economy depends on the availability of data
- Technological foundation + services and tools to make use of it
- Different industries and communities
- Machine-learning systems need high-quality training data
  - predictive maintenance

Data space standards, like IDS, provide **technical foundation** and **data sovereignty** principles.

# Motivation

- Interoperability is key for making use of shared data
- Syntactic data exchange already solved, I.e. defined data formats and exchange standards
- Semantic interoperability still an open challenge
- Starting point: IDS RAM and IM defines semantic metadata
- we need to extend it to cover **semantic interoperability for data**

# Approach

- Semantic Interoperability via a **Semantic Layer** for Dataspaces
  - based on IDS Vocabulary Hub
  - based on knowledge graph (KG) technologies
  - implemented by **GraphDB + PoolParty** products
- Services + data for semantic interoperability
- data discovery & harmonisation
- Semantic metadata + data for discovery and integration

# Use Cases

Two running projects:

**DataBri-X**

- energy community simulation and prediction
- legal knowledge graph + document analysis

**UNDERPIN**

- refinery predictive maintenance
- wind farms predictive maintenance

# Approach

Semantic Layer supports structured and unstructured data interlinked in the KG.

**Structured**

- Ontology-Based Data Access
- RDB to RDF Mapping Language
- CSV on the Web (CSVW)


**Unstructured**

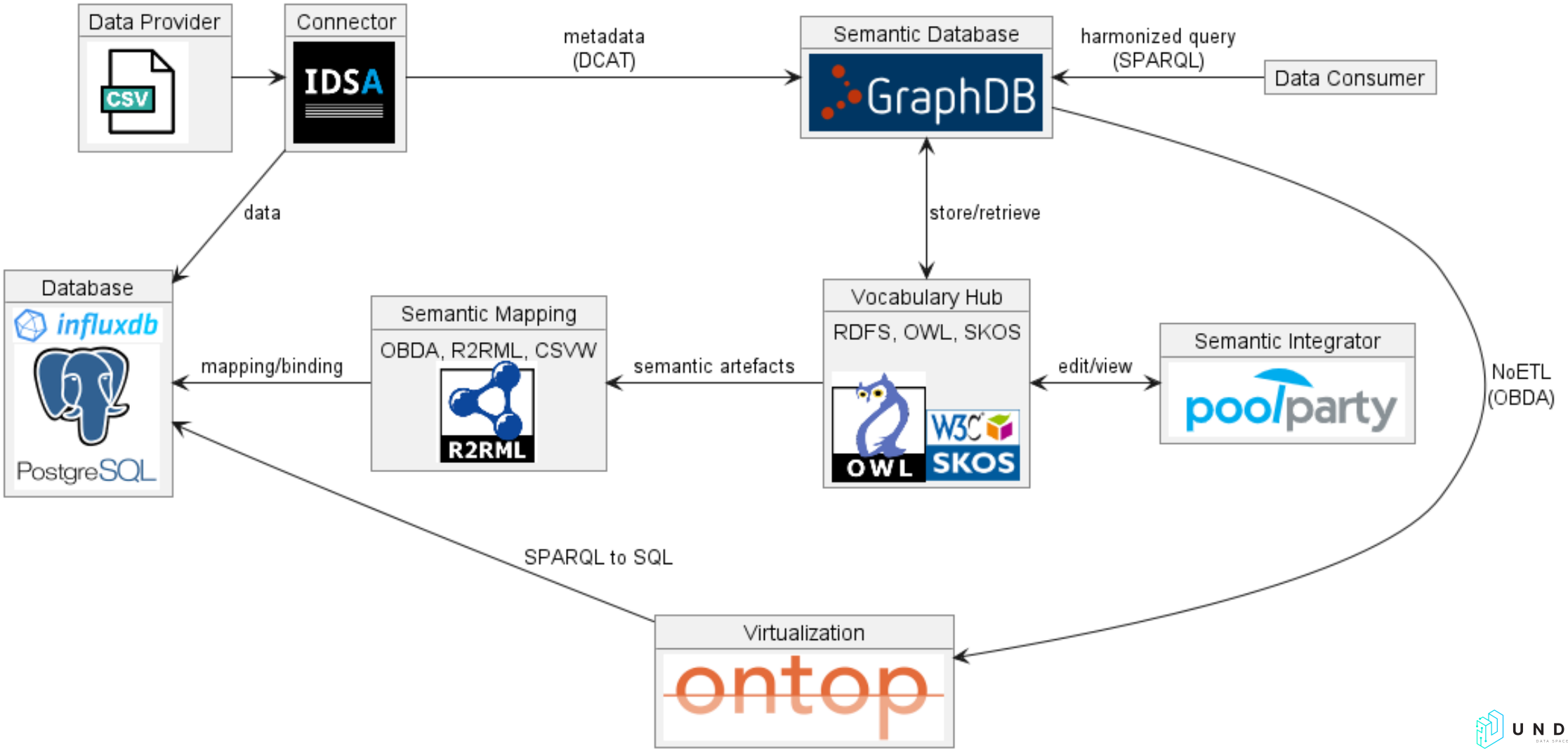
- Semantic document annotations via KG entities

# Approach

Semantic Layer supports inference services.

- Reasoning based on W3C recommendations OWL & SHACL
- Inference tagging expands semantic annotations in documents
- Vocabulary crosswalks connect different similar vocabularies
- Metadata inference to determine specific vocabularies for domains

# Integration Architecture



# Data Consumer Benefits

- Harmonized and semantically integrated data from disparate providers
- Richer metadata descriptions
- Incoming data bound dynamically to semantic descriptions
- Richer discoverability / easier data querying
- Vocabulary crosswalks: expand queries and annotations by traversing the interlinked graph

# Summary and Future Work

A Semantic Layer enchances the Vocabulary Hub’s role by interlinking structured and unstructured data from multiple providers and formats based on a harmonized model. This improves the discoverability and comprehensibility of said data.

**Future work:**

- Implement use cases for richer discoverability, harmonized querying and support for different content types.
- Explore how ML can benefit regarding quality in practice when providing consolidated and cleaned data via dataspaces.
- Discuss how we can extend the IDS RAM with services to improve the support for semantic interoperability provided by our solution.