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

1 Acknowledgements

2 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.

3 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**

4 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

5 Use Cases

2 running projects: DataBri-X

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

UNDERPIN

- refinery predictive maintenance
- wind farms predictive maintenance

6 Approach

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

- Ontology-Based Data Access (OBDA)
- RDB to RDF Mapping Language (R2RML)
- CSV on the Web (CSVW)

Unstructured

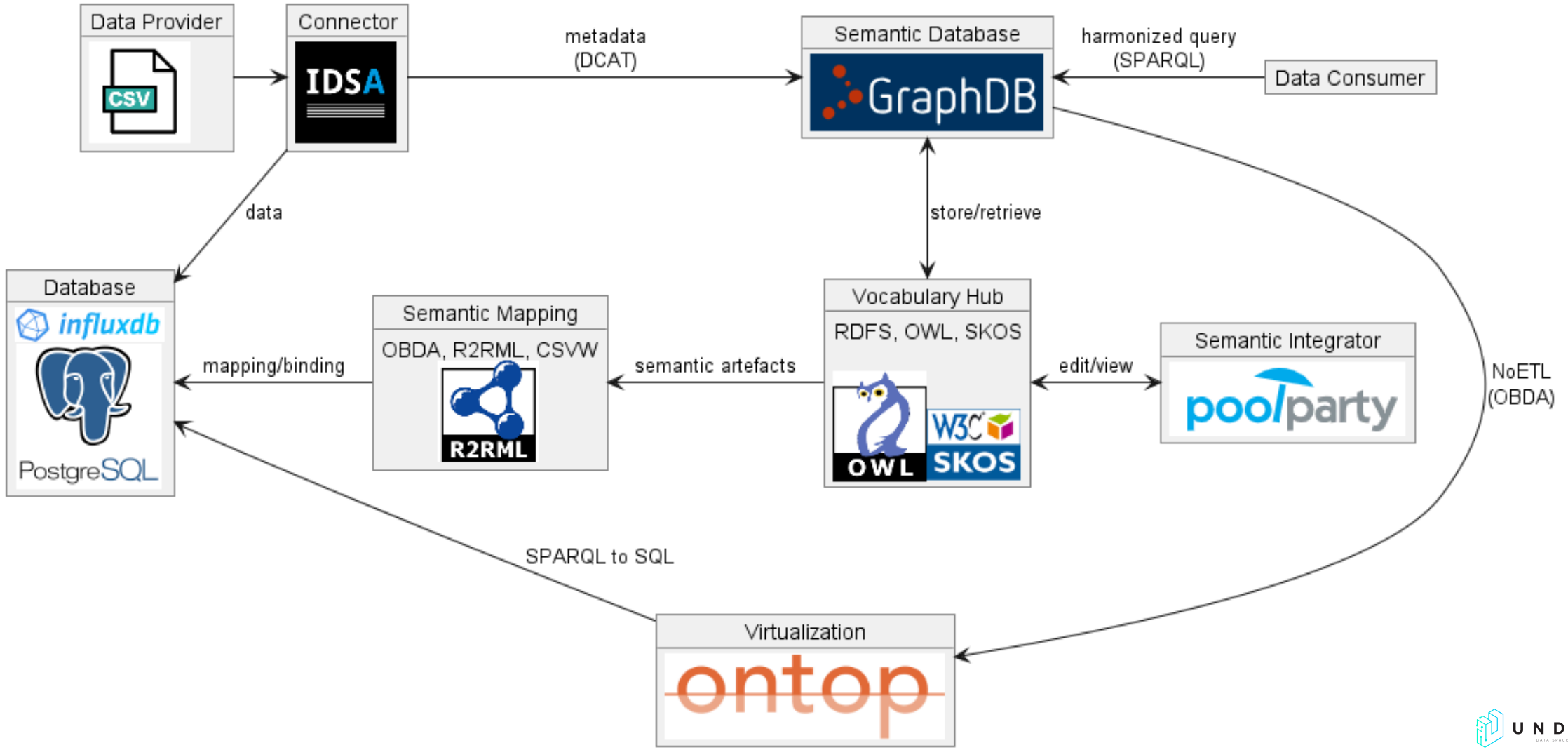
- semantic document annotations via KG entities

7 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

8 Integration Architecture



9 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

10 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.