

Serving Bosch Production Data as Virtual KGs [★]

Elem Güzel Kalaycı^{1,2}, Irlan Grangel González³, Felix Lösch³, Guohui Xiao¹,
Anees ul-Mehdi³, Evgeny Kharlamov^{4,5}, and Diego Calvanese¹

¹ Free University of Bozen-Bolzano, ² Virtual Vehicle Research GmbH,
³ Bosch Corporate Research, ⁴ Bosch Center for AI, ⁵ University of Oslo

Abstract. Analyses of manufacturing processes is vital for effective and efficient manufacturing. In complex industrial settings, such analyses should account for data that comes from many different and highly heterogeneous machines, and thus are affected by the data integration challenge. In this work, we show how this challenge can be addressed with semantics using Virtual Knowledge Graphs. For this purpose, we propose the **SIB** Framework, in which we semantically integrate Bosch manufacturing data. In this demo we present **SIB** in action on 2 scenarios for the analysis of the Surface Mounting Process (SMT) pipeline.

1 Introduction

The digitization trend in manufacturing industry, known as Industry 4.0, leads to a huge growth of volume and complexity of data generated by machines involved in manufacturing processes. These data become an asset of key relevance for enhancing the efficiency and efficacy of manufacturing. However, unlocking the potential of these data is a major challenge for many organizations. Indeed, often the data naturally reside in silos and its effective analyses requires costly data integration that includes cleaning, de-duplication, and semantic homogenization. Such integration can consume up to 70-80% of the overall data analyses time [4].

A prominent approach to address this data integration challenge is semantic data integration that is based on Virtual Knowledge Graphs (VKG) [5]. In such approach, one relies on ontologies that mediate between the data and analytical applications and expose the domain of data rather than the data itself in terms classes and properties, while the data is connected to the ontology by making use of semantic mappings [1]. Then, the data can be analysed by posing queries over the ontologies and the VKG system takes care of transforming them into queries over the data and pushing them to the data for further processing.

The above solution has been implemented at Bosch in the VKG-based data integration framework called **SIB** (for Semantic Integration at Bosch) by relying on the state-of-the-art VKG engine Ontop [2] and deployed at Bosch. The purpose of this deployment is to evaluate the feasibility of using semantic technologies and data integration based on them for supporting product quality analysis.

In this demo we will present **SIB** and its Bosch deployment. The demo is based on the data from a Bosch plant located in Salzgitter, Germany, that produces

[★] Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

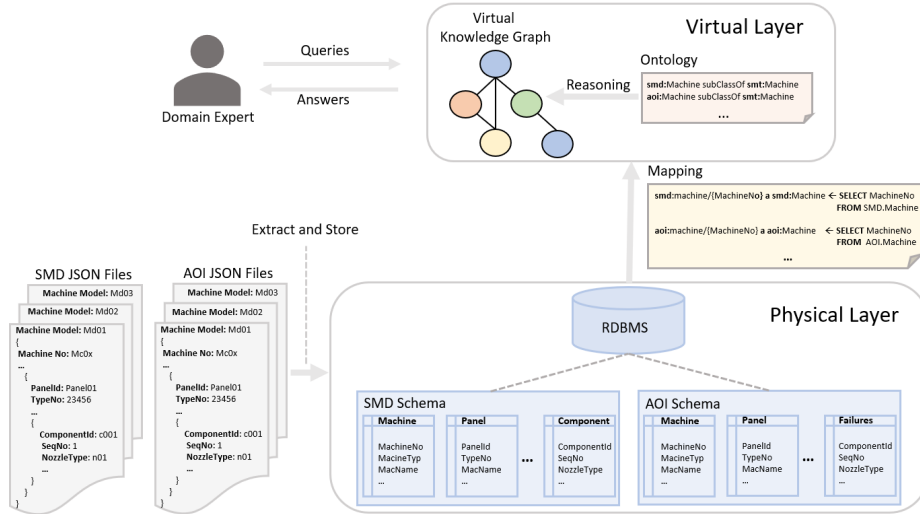


Fig. 1. SIB exemplified over Bosch SMT scenario.

electronic control units. The scenario of the demo is the product quality analysis that is performed at the plants and that requires integration of vast amounts of heterogeneous data. More precisely, the demo is focused on failure detection for *Surface Mounting Process* (SMT) that fundamentally relies on the integration and analysis of data generated by the machines deployed in different phases of the process. Such machines, e.g., for *placing electronic components* (SMD) and for *automated optical inspection* (AOI) of solder joints, usually come from different suppliers and they rely on distinct formats and schemata for managing the same data across the process. Hence, the raw, non-integrated data does not give a coherent view of the whole SMT process and hampers analysis of the manufactured products. During the demo the attendees will be able to explore the SMT Ontology we developed, observe sample SMT data, and mappings between the data and ontology. Moreover, we encoded relevant product analysis tasks into a catalog of SPARQL queries formulated over the SMT Ontology. The demo attendees will be able to explore the product analyses tasks, how they were encoded in SPARQL, and how easy such complex tasks can be achieved with the help of SIB. In particular, the latter will be shown by comparing SPARQL queries to the native database queries over the underlying SMT data.

This demo accompanies our accepted in-use track paper at ISWC’20 [3].

2 Our Solution

Our SIB solution for semantic integration of manufacturing data of the SMT process is depicted in Figure 1. Note that the raw manufacturing log data comes in JSON files generated by various machines and then it is extracted and loaded into a PostgreSQL database. For processing of queries posed over the VKG, SIB

relies on the state-of-the-art VKG framework Ontop that computes answers end-user SPARQL queries by translating them into SQL queries, and delegating the execution of the translated SQL queries to the original data sources.

Note that the VKG approach does not require to materialize into a KG all facts entailed by the ontology. Moreover, the workflow of Ontop can be divided into an *off-line* and an *online* stage. As the first step at the *off-line stage*, Ontop loads the OWL 2 QL ontology and classifies it via the built-in reasoner, resulting in a directed acyclic graph stored in memory that represents the complete hierarchy of concepts and that of properties. In the second step, Ontop constructs a so-called *saturated mapping*, by compiling the concept and property hierarchies into the original VKG mapping. This aspect is important also in SIB, since the domain knowledge encoded in the ontology allows for simplifying the design of the mapping layer. During the offline stage, Ontop also optimizes the saturated mapping by applying structural and semantic query optimization.

During the *online stage*, Ontop takes a SPARQL query and translates it into SQL by using the saturated mapping. To do so, it applies a series of transformations that we briefly summarize here [2,6]: *(i)* it rewrites the SPARQL query w.r.t. the ontology; *(ii)* it translates the rewritten SPARQL query into an algebraic tree represented in an internal format; *(iii)* it unfolds the algebraic tree w.r.t. the saturated mapping, by replacing the triple patterns with their optimized SQL definitions; and *(iv)* it applies structural and semantic techniques to optimize the unfolded query. One of the key points in the last step is the elimination of self-joins, which negatively affect performance in a significant way. To perform this elimination, Ontop utilizes in an essential way the key constraints defined in the data sources. In those cases where it is not possible to define these key constraints explicitly in the data sources, or to expose them as metadata of the data sources so that Ontop can use them, Ontop allows one to define them implicitly, as part of the mapping specification. The data we have been working with in the Bosch use case was mostly log data and stored as separate tables containing often highly denormalized and redundant data. Consequently, there were a significant amount of constraints in the tables that are not declared as primary or foreign keys, which brought significant challenges to the performance of query answering. To address these issues, we had to declare these constraints manually, and supply them as separate inputs to Ontop.

3 Demonstration Scenarios

We prepared two scenarios for the demo:

[S1:] SIB Deployment over Bosch data. In this scenarios the attendees will get a better understanding of the data integration challenge with the Bosch SMT use case and how it can be addressed with the help of semantic technologies offline, prior performing the actual data analyses. In particular, the attendees will be able to look closer at Bosch manufacturing data, to understand particularities of SMD and AOI data formats. Then, the attendees will study the Bosch SMT Ontology by zooming into its classes and properties. Finally, they will be able to study mappings relating the ontology and the data.

[S2:] Product analysis with SIB. In this scenario the attendees will be able to benefit from the deployed Bosch VKG solution. In particular, the attendees will study several product analysis tasks for the Bosch SMT use case. Then, they will study how these tasks can be expressed by means of suitable SPARQL queries over the SMT Ontology. Notably, such queries make use of ontology terms to refer to the relevant information assets, and thus are very close to the natural language formulation of the analysis tasks, which in turn makes it easy for Bosch engineers to formulate them. Then, the attendees will experience how to obtain the respective analysis data coming from the process logs, by simply executing such queries over the underlying database via the SIB VKG engine. Finally, the attendees will compare the SPARQL queries and their SQL counterparts to witness how much easier the former are comparing to the latter in terms of the size, number of joins, readability of schema elements.

We now illustrate the data and the queries for the two scenarios. The data is mainly based on two sets of relational tables: SMD Tables where the most notable ones are `smd_event`, `smd_location`, `smd_panel`, `smd_components`, and AOI Tables with `aoi_event`, `aoi_location`, `aoi_panel`, and `aoi_failures`. Consider a sample example record in one of these tables:

smd_panel				
panelId	boardNo	machineName	processedTS	location
p01	b01	SMD Machine 1	24-04-2020	mes01

We prepared 13 analytical tasks for the demo and they were the result of a collaborative work and a careful selection during two visits to Bosch plants and meetings with Bosch line engineers and line managers. The queries offer a good balance among three dimensions: they are representative for product analyses, offer a good coverage of product analyses tasks, and they are complex enough to account for a reasonable number of domain terms. Consider one such query in natural language and in SPARQL:

Query q3: “Return all panels processed from a given time T up to the detection of a failure.”

Despite the temporal nature of the query it can be realized in SPARQL:

```

1 SELECT DISTINCT ?panel ?ts ?eventTime
2 WHERE {?panel psmt:pTStamp ?ts . {
3     SELECT ?eventTime
4     WHERE {?eventfailure fsmt:eTStamp ?eventTime .
5     FILTER (?eventTime > '2018-06-01T00:06:00.000+02:00'^^xsd:dateTimeStamp)}
6     ORDER BY (?eventTime) LIMIT 1 }
7 FILTER (?ts > '2018-06-01T00:06:00.000+02:00'^^xsd:dateTimeStamp && ?ts < ?eventTime) }
```

References

1. Bienvenu, M., Rosati, R.: Query-based comparison of mappings in ontology-based data access. In: Proc. KR, AAAI Press (2016) 197–206
2. Calvanese, D., Cogrel, B., Komla-Ebri, S., Kontchakov, R., Lanti, D., Rezk, M., Rodriguez-Muro, M., Xiao, G.: Ontop: Answering SPARQL queries over relational databases. *Semantic Web J.* **8**(3) (2017) 471–487

3. Kalaycı, E.G., González, I.G., Lösch, F., Xiao, G., ul Mehdi, A., Kharlamov, E., Calvanese, D.: Semantic integration of bosch manufacturing data using virtual knowledge graphs. In: Proc. ISWC. (2020)
4. Kharlamov, E., Hovland, D., Skjæveland, M.G., Bilidas, D., Jiménez-Ruiz, E., Xiao, G., Soylu, A., Lanti, D., Rezk, M., Zheleznyakov, D., Giese, M., Lie, H., Ioannidis, Y., Kotidis, Y., Koubarakis, M., Waaler, A.: Ontology based data access in Statoil. *J. Web Semantics* **44** (2017) 3–36
5. Xiao, G., Ding, L., Cogrel, B., Calvanese, D.: Virtual knowledge graphs: An overview of systems and use cases. *Data Intelligence* **1**(3) (2019) 201–223
6. Xiao, G., Kontchakov, R., Cogrel, B., Calvanese, D., Botoeva, E.: Efficient handling of SPARQL optional for OBDA. In: Proc. ISWC. LNCS, Springer (2018) 354–373