

Realizing an RDF-based Information Model for a Manufacturing Company – A Case Study^{*}

Niklas Petersen^{1,2}, Lavdim Halilaj^{1,2}, Irlán Grangel-González^{1,2},
Steffen Lohmann², Christoph Lange^{1,2}, Sören Auer^{3,4}

¹ Enterprise Information Systems (EIS), University of Bonn

{petersen|halilaj|grangel|langec}@cs.uni-bonn.de

² Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS)

steffen.lohmann@iais.fraunhofer.de

³ Computer Science, Leibniz University of Hannover, Germany

⁴ TIB Leibniz Information Center for Science and Technology, Hannover, Germany
soeren.auer@tib.eu

Abstract. The digitization of the industry requires information models describing assets and information sources of companies to enable the semantic integration and interoperable exchange of data. We report on a case study in which we realized such an information model for a global manufacturing company using semantic technologies. The information model is centered around machine data and describes all relevant assets, key terms and relations in a structured way, making use of existing as well as newly developed RDF vocabularies. In addition, it comprises numerous RML mappings that link different data sources required for integrated data access and querying via SPARQL. The technical infrastructure and methodology used to develop and maintain the information model is based on a Git repository and utilizes the development environment VoCol as well as the Ontop framework for Ontology Based Data Access. Two use cases demonstrate the benefits and opportunities provided by the information model. We evaluated the approach with stakeholders and report on lessons learned from the case study.

1 Introduction

Although the vision of digitizing production and manufacturing has gained much traction lately (viz. Industry 4.0), it is still not clear how it can actually be *implemented* in an interoperable way using concrete standards and technologies [4]. A key challenge is to enable industrial devices to communicate and to *understand* each other as a prerequisite for cooperation scenarios [13]. Different standards, such as those from the ISO and IEC series, are used to describe information about manufacturing, security, identification and communication, among other areas.

Integrating all relevant information and automating as many production steps as possible is the central goal of the Industry 4.0 vision [11]. Instead of envisioning one monolithic system or database, we pursue a decentral semantic integration, i.e., the formal description and linking of all relevant assets and data sources based on an aligned

^{*} This work has been supported by BMBF (German Federal Ministry of Education and Research) grants 01IS15035C (SDI-X) and 01IS15054 (Industrial Data Space).

set of RDF vocabularies – the *information model*. This avoids unnecessary data redundancy and allows for structured querying and analyses across individual assets and data sources based on SPARQL. The information model serves as a crystallization point and reference for data structures and semantics emerging from the data sources and value chains. Furthermore, it is aligned to important industry standards, such as *RAMI* [2] and *IEC 62264* [1], to additionally foster data exchange and semantic interoperability.

The information model is more than a set of aligned RDF vocabularies. It comprises i) a methodology for the development, curation and integration of vocabularies for the domain in focus; ii) a technical infrastructure for following the methodology; iii) governance procedures, which are aligned with the corporate organizational structure.

In this paper, we report on a case study in which we realized such an information model for a global manufacturing company, and discuss findings and lessons learned derived from the case study. In Section 2, we describe the context, requirements and motivating scenario of this work. The core contributions—the information model and its implementation—are presented in Section 3 and Section 4. In Section 5, we apply the information model to two use cases demonstrating its benefits and opportunities. Section 6 reports on stakeholder feedback and summarizes the lessons learned. Section 7 reviews related work, and Section 8 concludes with an outlook to future work.

2 Motivating Scenario

The information modeling project involved employees from different departments and hierarchical levels of the manufacturing company, external consultants and a third party IT provider. The company itself realized that their IT infrastructure has reached a level of complexity making difficult to manage and effectively use their existing systems and data. While adding new sensors to production lines is straightforward, using the sensor data effectively to improve the production process and decision-making can be cumbersome. The need to share production data with clients led them to evaluate the fitness of semantic technologies. For example, the production of bearing tools is fairly quality-driven depending on the customer specifications. Sharing the production details in a more processable format (compared to non-machine-comprehensible formats) aroused interest. Further goals were to gain a *bigger picture of the company's assets* (physical and non-physical) and to capture as much expert knowledge as possible.

2.1 Use Cases

The concrete use cases are based upon a machine newly introduced into the production lines of the company, a so-called *machine tool*. This is a machine that requires the mounting of tools to assemble specific metal or rigid products. Compared to older generations, the new machine features more than 100 embedded sensors that monitor the production.

Tool Management Possible tools to be mounted into the machine are cutters, drillers or polishers. A tool usually consists of multiple parts. The number of parts depends on the manufacturer of the tool, which is not necessarily the same as the manufacturer of the machine. Mounting tools into a machine is a time-consuming task for the machine

operator. Uncertain variables of the tools, such as location, availability and utilization rate, play a major role in the efficiency of a work shift and of a machine in particular. The production of certain goods may wear a tool out quickly, thus decreasing its overall lifetime and forcing the machine operator to stop the machine and replace it with a new tool. Reducing the idle time for remounting the machine by clearly describing its configuration, location and weariness was therefore one concrete goal to be addressed by the information model.

Energy Consumption Producing goods with the machine tool is an energy-intensive process. Before we started the information modeling project, only the energy costs per factory were known. Sensors were added to track the energy consumption per machine and processed work order. For the cost calculation, data from the added sensors and the work orders, which resides in different data sources, needs to be linked and jointly queried. Therefore, integrating this data to be able to retrieve the information at run-time was another concrete goal addressed by the information modeling project.

2.2 Data Sources

Three types of data were of particular interest in the project: i) Sensor Data (SD), ii) the Bill of Materials (BOM), and iii) data from the Manufacturing Execution System (MES). The SD comprises sensor measurements of the machine tool. These measurements record parameters needed for the continuous monitoring of the machine, such as energy, power, temperature, force, vibration, etc. The MES contains information about work orders, shifts, material numbers, etc. The machine produces assets based on the work order details, which provide the necessary information for the production of a given asset. The BOM contains information about the general structure of the company, such as work centers, work units, associated production processes, as well as information related to the work orders and the materials needed for a specific production.

3 Realizing the RDF-based Information Model

The information model aims at a holistic description of the company, its assets and information sources. The core of the model is based on a factory ontology we developed in a previous project [18], which describes real world objects from the factory domain, including factories, employees, machines, their locations and relations to each other, etc. In addition, the information model comprises the mappings between ontologies that represent the data sources (i.e., SD, BOM, MES) and their corresponding schemes.

3.1 Development Methodology

Our development methodology was based on the approach proposed by Uschold et al. [22]. We first defined the purpose and scope of the information model; then, we captured the domain knowledge, conceptualized and formalized the ontologies and aligned them with existing ontologies, vocabularies and standards. Finally, we created the mappings between the data sources and ontological entities. In line with best practices, we

followed an iterative and incremental development process, i.e., with an increased understanding of the domain, the information model was continuously improved.

All artifacts were hosted and maintained by VoCol [10], a collaborative vocabulary development environment which we adapted for the purpose of this project. VoCol supports the requirements of the stakeholders: i) version-control of the ontology; ii) online and offline editing; and iii) support for different ontology editors (by generating a unique serialization before changes are merged to avoid false-positive conflicts [9]). In addition, it offers different web-based views on the ontology, including a human-readable documentation, a visualization and charts generated from queries applied to the ontology and instance data. These views are designated to ease the collaboration of domain experts in the development process, i.e., enabling them to participate without having to set up and maintain a proper infrastructure themselves.

Purpose and Scope The information model comprises i) a formal description of the physical assets of the company, ii) mappings to database schemas of existing production systems, and iii) a formalization of domain-related knowledge of experienced employees about certain tasks and processes within the company.

The heart of the information model represents the aforementioned *machine tool*, including its sensor data, usage processes and human interaction. Therefore, the majority of concepts are defined by their relation to this machine.

The scope is set by the motivating examples *energy consumption* and *tool management* introduced in Section 2.1. Nevertheless, the management considered it also *nice to have* to gain a clearer picture of all assets of the company. For example: What local knowledge exists in the factory? What kind of data exists for which machine? Where is that data? Who has access to it? Discussions on fully automated order-driven production sites are ongoing. The management hopes for this to be supported by the information model, and we aim to provide the basis for that goal.

Capturing Domain Knowledge We captured the domain knowledge in different ways:

1. The company provided descriptive material of the domain, including maps of factories, descriptions of machines and work orders, information about processes, sensor data and tool knowledge. The types of input material ranged from formatted and unformatted text documents to spreadsheets and SQL dumps.
2. An on-site demonstration of the machine within the factory was given during the project kick-off, including a discussion of further contextual information missing in the material. In subsequent meetings, open questions were clarified and concrete use cases for the information model were discussed.
3. We reviewed relevant existing ontologies and industry standards, intending to build on available domain conceptualizations and formalizations.
4. We created customized document templates to enable easy participation of domain experts by collecting input on the ontology classes and properties in a structured way. We collected names and descriptions of all properties having a given class as their domain in one table with one row per property; additional details about the domains and ranges of properties were collected in a separate table. These documents were handed over to the domain experts to be reviewed and completed.

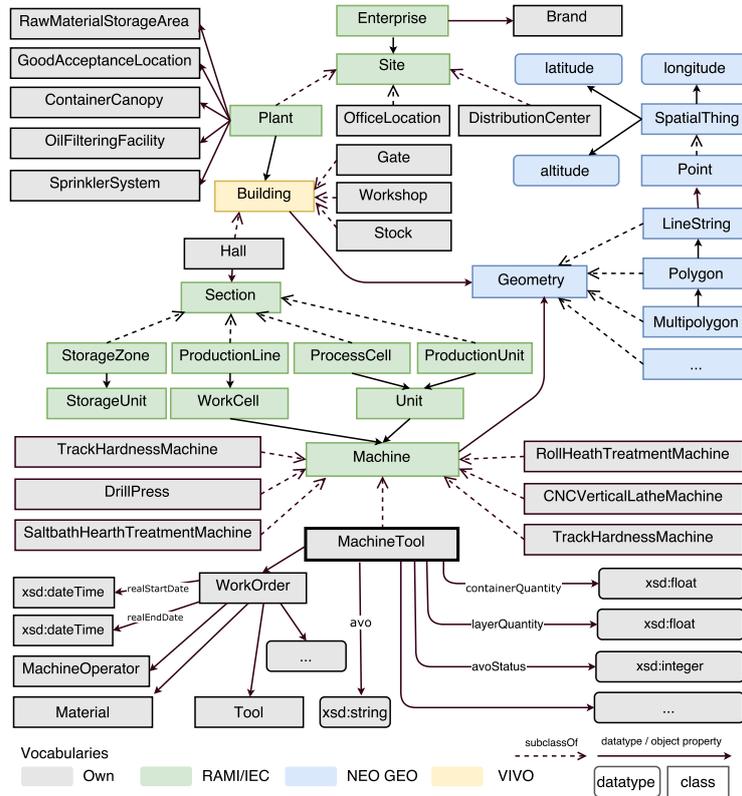
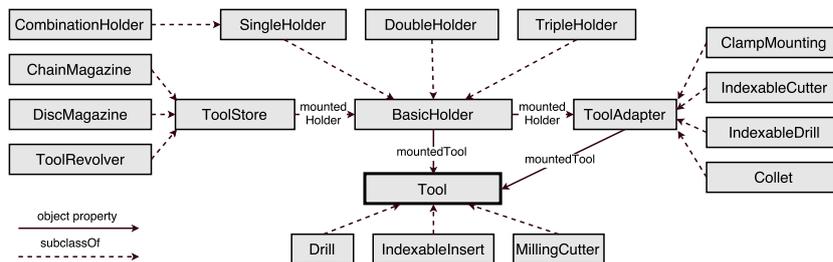


Fig. 1: Core concepts of the information model

5. We trained IT-affine employees of the company on modeling ontologies using editors such as Protégé, TopBraid Composer and the Turtle editor integrated into the VoCol environment.⁵

Conceptualizing and Formalizing Figure 1 shows the core concepts of the developed information model. The colors group the different subdomains and reused vocabularies. Since Machine(s) are the main assets of the manufacturing company, they have been used as a starting point for creating the ontology. Each machine contains a geo-location (property with range Geometry) and is part of a certain Section, which is in a certain Hall. Each Hall can contain multiple sections and also has a geo-location (inherited from Building). Plant, OfficeLocation and DistributionCenter are different types of Site(s), each serving a specific purpose. The MachineTool comprises domain-related properties to describe its AVO (operation status). Next, it is connected with WorkOrder(s) to be processed. Each WorkOrder defines the re-

⁵ <http://protege.stanford.edu/>, <http://www.topquadrant.com/tools/modeling-topbraid-composer-standard-edition/>, <https://github.com/vocol/vocol>

Fig. 2: Ontological view of the *Tool* concept

quired *Material* and *Tool(s)*, as well as which machine should be used by which operator to execute a particular task.

Figure 2 provides a more detailed view on the *Tool*-related concepts. Machines have different interfaces, called *ToolStore(s)*. Tool stores can be equipped with different *BasicHolder(s)*. There exist three kinds of basic holders: *SingleHolder(s)*, *DoubleHolder(s)* and *TripleHolder(s)*. The name indicates the number of tools a holder can be assembled with. A *CombinationHolder* is a special kind of *SingleHolder* that can only be combined with a specific tool. The majority of holders are to be combined with a *ToolAdapter*, on which the actual *Tool* is mounted. Certain tools can be mounted directly onto the holder. Tools are the parts that wear out over time and need to be replaced. As with machines, their geolocation is defined, such that their position can be shown on a map. The tool ontology part reflects the configuration options for tools of multiple tool manufacturers.

In total, the developed ontologies comprise of 148 classes, 4662 instances, 89 object and 207 datatype properties. We focused on the description of the core concepts here that are needed to understand this work; a description of all ontology concepts would be out of the scope of this paper.

Aligning with Existing Ontologies & Standards The developed information model consists of concepts from existing vocabularies and industrial standards that we formalized during the project. In particular, concepts from the VIVO (*vivo:Building*), NeoGeo (*ngeo:Geometry*), FOAF (*foaf:Person*) and Semantic Sensor Network (*ssn:Sensor*) vocabularies and ontologies are reused.⁶

Furthermore, we aligned the ontologies of the information model to RDF vocabularies of the industry standards RAMI [2, 7] and IEC 62264 [1] that we developed.⁷ RAMI is a reference model aiming at structuring the interrelations between IT, manufacturing and product life cycles, whereas IEC 62264 defines hierarchical levels within a company. RAMI includes the IEC concepts and adds the “connected world” as an additional level on top, which aligns well with the basic idea and motivation of this work.

⁶ <http://vivoweb.org/>, <http://geovocab.org/doc/neogeo.html>, <http://xmlns.com/foaf/spec/>, <https://www.w3.org/TR/vocab-ssn/>

⁷ <https://w3id.org/i40/rami/>, <https://w3id.org/i40/iec/62264>

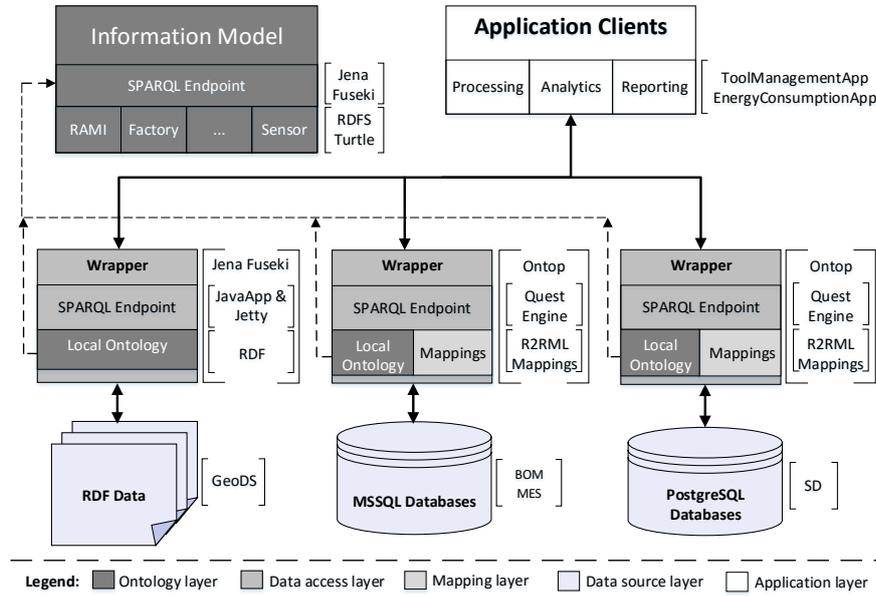


Fig. 3: The implemented architecture comprised of different layers

4 Architecture and Implementation

With the objective to provide a uniform interface for accessing heterogeneous distributed data sources, we designed and implemented the architecture illustrated in Figure 3. It is extensible and able to accommodate additional components for accessing other types of data sources as well as supporting federated query engines. The architecture distinguishes the following four main layers, some of which are orthogonally located across different components:

The **ontology layer** consists of several ontologies that have been created to conceptualize a unified view of the data (cf. section 3). Wache et al. distinguish three main approaches of using ontologies to explicitly describe data sources [23]: i) *Global Ontology Approach*—all data sources are described in an integrated, global ontology; ii) *Multiple Ontology Approach*—separate local ontologies represent the respective data sources, and mappings between them are established; the iii) *Hybrid Ontology Approach*—a combination of the two previous approaches with the aim to overcome the drawbacks of maintaining a global shared ontology and mappings between local ontologies.

We followed the third approach, which enables new data sources to be added easily, avoiding the need for modifying the mappings or the shared ontology. Accordingly, our ontologies are organized in two groups: i) a shared ontology to represent the highest level of abstraction of concepts and mappings with external ontologies; and ii) local ontologies representing the schemas of the respective data sources. This makes our architecture quite flexible with respect to the addition of diverse types of data sources [3].

The **data access layer** consists of various wrappers acting as bridges between client applications and heterogeneous data sources. It receives user requests in the form of SPARQL queries, which are translated into the query languages of the respective data sources, and returns the results after query execution. Accessing relational databases is realized using the *Ontology-Based Data Access* (OBDA) paradigm, where ontologies are used as a conceptualization of the unified view of the data, and mappings to connect the defined ontology with the data sources [19]. In particular, the Ontop [5] framework is used to access the data sources, i.e., the BOM, MES, and SD data, which exposes the relational databases as virtual RDF graphs, thus eliminating the requirement to materialize data into RDF triples. Additionally, Jena Fuseki is used as in-memory triple store for loading GeoDS, since the information about the geo-locations of the machines are less than 20,000 RDF triples.

The **mapping layer** deals with the mappings between the data stored in the data sources and the local ontologies. For the definition of the mappings, we used *R2RML*⁸, the W3C standard RDB-to-RDF mapping language. As a result, it is possible to view and access the existing relational databases in the RDF data model.

The **data source layer** comprises the external data sources, i.e. databases and RDF datasets as described in Section 2.2. Due to the high dynamicity and the great amount of incoming data, the data sources are replicated and synchronized periodically. As a result, any performance and safety degradation of the production systems is avoided. Additional types of data sources can be easily integrated in the overall architecture by defining local ontologies, mappings with the global ontology and data sources as well as choosing an appropriate wrapper.

The **application layer** contains client applications that benefit from the unified access interface to the heterogeneous data sources. These applications can be machine agents or human interface applications able to query, explore and produce human-friendly presentations.

5 Application to the Use Cases

We applied the developed information model to the use cases introduced in Section 2.1 to demonstrate the possibilities resulting from semantically integrated data access.

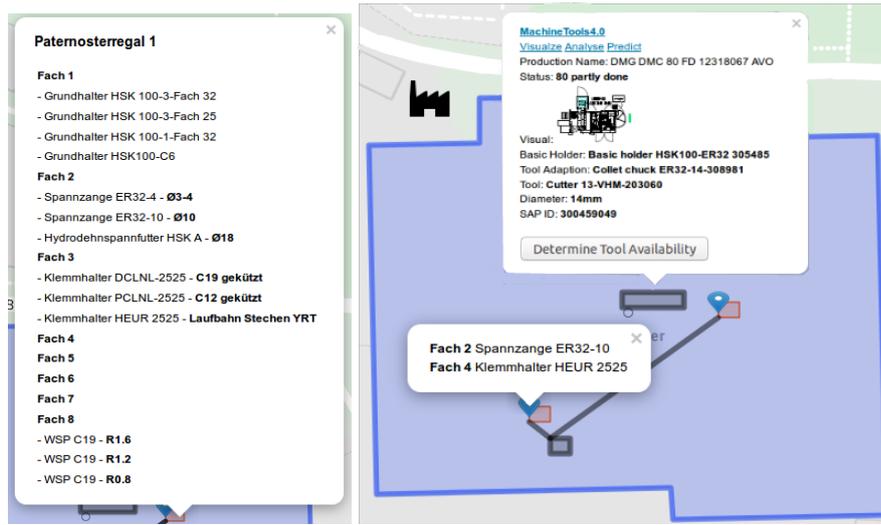
5.1 Tool Management

Figure 4 displays different views on the assets of the company. On a world map (see Figure 4a), the sites of the company are highlighted based on their geo-location given in the information model. By zooming in, the different locations can be investigated w.r.t. their functionality, address, on-site buildings up to the level of machines, etc. By clicking on the objects on the map, static and live production data is displayed. As an example, Figure 4b shows all tools stored in a certain paternoster system, grouped by drawer. Figure 4c provides an example of a machine with its properties: production name, current status, self-visualization, mounted basic holder, tool with its diameter,

⁸ <http://www.w3.org/TR/r2rml/>



(a) Global view of the company sites



(b) Paternoster view

(c) Factory view

Fig. 4: Various views of the tool management application

etc. Further, it contains links to existing external analytical web pages. A “Determine Tool Availability” function is offered for locating the tools to be assembled in the closest paternoster storage system based on the location of the machines.

Each time a certain view is opened, a SPARQL query is executed to retrieve the required data in the information model. Geo-locations are drawn to the world map view using the Leaflet JavaScript library⁹.

```

1 @prefix rr: <http://www.w3.org/ns/r2rml#> .
2 @prefix im: <http://iaais.fraunhofer.de/vocabs/infomodel#> .
3 <WorkOrderMap> a rr:TriplesMap ;
4   rr:logicalTable [ rr:tableName "WorkOrders" ];
5   rr:subjectMap [ rr:template
6     "http://.../infomodel/WorkOrder/{WorkOrderId}"; rr:class im:WorkOrder];
7   rr:predicateObjectMap
8     [rr:predicate im:workOrderId;   rr:objectMap [rr:column "WorkdOrderId"]],
9     [rr:predicate im:beginWorkOrder; rr:objectMap [rr:column "BeginWorkOrder"]],

```

⁹ <http://leafletjs.com/>

```

10 [rr:predicate im:targetAmount; rr:objectMap [rr:column "TargetAmount"]],
11 [rr:predicate im:totalExecTime; rr:objectMap [rr:column "TotalExecTime"]],
12 [rr:predicate im:matDesc; rr:objectMap [rr:column "MatDesc"]].
13 ...

```

Listing 1.1: R2RML mapping for work orders

5.2 Energy Consumption

Information about the energy, power or temperature are critical for the company to forecast the production process, expenses and maintenance. In the second use case, we asked the following question: what is the *energy consumption* of a given machine for a given day for a particular *work order*? To answer this question, data from sources introduced in Section 2.2 (SD, MES, BOM) needs to be taken into account. Since the SD lacks work order definitions, we used time intervals to access the required energy stream data. Next, we linked the work order IDs in the BOM and MES databases. Listing 1.1 displays an excerpt of the R2RML mappings for the relational database table *WorkOrders*. Among others, it includes its material number, total execution time and target production amount.

Based on these mappings, we defined two queries: The first one retrieves information about work orders (cf. Listing 1.2); the second one retrieves the energy consumption values for a work order in a specific time interval (cf. Listing 1.3).

```

1 PREFIX im: <http://iaais.fraunhofer.de/vocabs/infomodel#>
2 SELECT DISTINCT ?workOrderId ?materialDesc ?materialNumber
3 ?beginTime ?dateFrom ?dateTo ?totalExecTime ?targetAmount
4 WHERE {
5   ?o a im:workOrder; im:workOrderId ?workOrderId ;
6   im:beginWorkOrder ?beginTime; im:dateFrom ?dateFrom;
7   im:dateTo ?dateTo; im:targetAmount ?targetAmount;
8   im:totalExecTime ?totalExecTime; im:matNr ?materialNumber;
9   im:matDesc ?materialDesc.
10 FILTER (?dateFrom >= dateFrom && ?dateTo <= dateTo)}

```

Listing 1.2: Query to retrieve information about work orders for a time interval

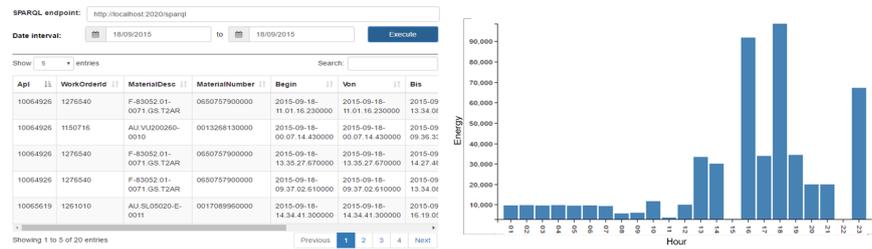
```

1 PREFIX im: <http://iaais.fraunhofer.de/vocabs/infomodel#>
2 SELECT ?hour ((?latest - ?earliest) AS ?measurementByHour) {
3   { SELECT ?hour (MIN(?measurement) AS ?earliest)
4     WHERE { ?machineSensor im:energyTime ?time; im:energyValue ?measurement.
5     } GROUP BY (HOURS(?time) AS ?hour) }
6   { SELECT ?hour (MAX(?measurement) AS ?latest)
7     WHERE { ?machineSensor im:energyTime ?time; im:energyValue ?measurement.
8     } GROUP BY (HOURS(?time) AS ?hour) }
9   FILTER (?hour >= timeFrom && ?hour <= timeTo) } ORDER BY ?hour

```

Listing 1.3: Query to retrieve the energy consumption per machine per hour

This allowed us to integrate the information from the three data sources using the information model and SPARQL queries. Figure 5a depicts the integrated information of work orders for a given machine, and Figure 5b shows the energy consumption per hour for that machine for a given day. Overall the performance of the implemented solution was satisfactory, i.e., time to retrieve the information for energy consumption of particular work order was less than five seconds.



(a) Work order data for a given machine in a time interval of one day (b) Energy consumption of a given work order within a day

5.3 Information Model Governance

Introducing new technologies is often a challenge for companies. The introduction has to be well-aligned with the organizational structure of the company to balance the added value produced for the information model to the business and the maintenance costs of the technology. Thus, in parallel with the introduction of the information model, we defined a procedure to support the *governance* of information to ensure the maintenance of the model and uniform decision-making processes.

Since the core of the information model is a network of ontologies and vocabularies with a clear hierarchical and modular structure, there are boards of experts assigned to each part, which are responsible for its maintenance. Decisions cover, for instance, new terms to be included or existing ones to be removed, external vocabularies to be reused and aligned, and the continuous alignment with industry standards implementing the Industry 4.0 vision, e.g., RAMI, IEC, ISO. Additionally, we provided concrete guidelines for maintaining the information model along with the use of VoCol, for example¹⁰

- detailed documentation of all terms defined in the vocabulary by `skos:prefLabel`, `skos:altLabel`, and `skos:definition`;
- multilingual definition of labels, i.e., in English and German;
- definition of `rdfs:domain` and `rdfs:range` for all properties;
- inclusion of provenance metadata, licenses, attributions, etc.

6 Evaluation and Lessons Learned

To gain feedback from the stakeholders involved in the information modeling project, we designed a questionnaire and sent it to the stakeholders, asking for anonymous feedback. Table 1 lists the questions and results of the questionnaire. We were interested in how the stakeholders evaluate the developed information model and semantic technologies in general, based on the experience they gained in the project.

¹⁰ These are based on the W3C “Data on the Web Best Practices” Recommendation, <https://www.w3.org/TR/dwbp/>

6.1 Stakeholder Feedback

Five employees of the manufacturing company (three IT experts, one analyst, and one consultant) who were actively involved in the project answered the questionnaire. The results varied across the stakeholders: While some regarded the information model and future potential of semantic technologies as promising, others remained skeptical about its impact within the company. Question 6 asking for the expectations towards semantic technologies (cf. Table 1) was answered by nearly all as an “enabler for autonomous systems” and by one as a “potential technology to reduce the number of interfaces”. One stakeholder praised the “integration and adaption” capabilities of semantic technologies. Question 7 asking for the biggest bottleneck yielded the following subjective answers: “lack of standardized upper ontologies”, “lack of field-proven commercial products”, “lack of support for M2M communication standards”, “skepticism of the existing IT personnel”. While the stakeholders find the advantages of semantic technologies appealing, the lack of ready-to-use business solutions, industrial ontologies and available IT personnel is halting their efforts to move forward. As a result of the project, the company is actively seeking IT personnel with a background in semantic technologies.

Table 1: Questions of the questionnaire and answers of the stakeholders

Question (with Likert scale of 1 to 5, 1 = not at all, 5 = very much; M = mean value, SD = standard deviation)	M	SD
1. Did the developed RDF-based information model meet your expectations?	2.4	0.9
2. Do you think investing in semantic technologies can result in a fast ROI?	3.0	1.4
3. Do you consider semantic technologies fit for usage in the manufacturing domain?	3.6	0.9
4. Are you satisfied with the software for semantic technologies available on the market?	2.8	1.3
5. Is it easy to hire personnel with knowledge in semantic technologies?	1.8	0.4
Free-text questions:		
6. What do you expect from semantic technologies in manufacturing contexts?		
7. What is the biggest bottleneck in using semantic technologies in manufacturing contexts?		

6.2 Lessons Learned

Technology awareness within the company After all, the majority of the stakeholders were enthusiastic and committed to developing an integrated information model and applications on top of it. Nevertheless, reservations on the fitness of the technology and methodology existed from the start. A few stakeholders preferred a bottom-up approach of first gathering and generating internally an overview of the existing schemas and models before involving external parties (such as our research institute). However, the management preferred an *outside view* and put a focus on quick results. Instead of spending time on finding an agreement on how to proceed, speed was the major driving force. Thus, they preferred to try out a (for them) “new” technology and methodology, which does not yet have the reputation of strong industrial maturity.

Perceived maturity of semantic technologies While semantic technologies are already widely used in some domains (e.g., life sciences, e-commerce or cultural heritage), there is a lack of success stories, technology readiness and show-case applications in most industrial areas. With regard to smaller and innovative products, the penetration of semantic technologies is still relatively small. A typical question when pitching semantic technologies within companies is “Who else in our domain is using them already?”. Therefore, it is important to point to successful business projects, even if details on them are usually rare.

Lack of semantic web professionals on the job market Enabling the employees of the manufacturer to extend the information model by themselves is crucial for the success of the project. Consequently, it is necessary to teach selected stakeholders the relevant concepts and semantic technologies. Hiring new staff experienced with semantic technologies is not necessarily an easy alternative. Compared to relational data management and XML technologies, there is still a gap between the supply of skilled semantic technology staff and the demand of the market.¹¹

Importance of information model governance Of major importance for the company is a clear governance concept around the information model, answering questions such as who or which department is allowed to access, modify and delete parts of the information model. An RDF-based information model has advantages in this regard: i) it enables people across all sites of the company to obtain a holistic view of company data; ii) current data source schemes are enriched with further semantic information, enabling the creation of mappings between similar concepts; and ii) developers can follow a defined and documented process for further evolving and maintaining the information model.

Building on top of existing systems Accessing data from the existing infrastructure as a virtual RDF graph was an important requirement of the manufacturing company. It avoids the costs of materializing the data into RDF triples and maintaining them redundantly in a triple store, and at the same time, benefits from mature mechanisms for querying, transaction processing, security, etc. of the relational database systems. Three different data access strategies were considered:

DB in Dumps Relational data to be analyzed is dumped in an isolated place away from the production systems, as not to affect their safety and performance. This strategy is used in cases where the amount of data is small and most likely to be static or updated very rarely.

DB in Replication All data is replicated, allowing direct access from both production systems and new analytic platforms. This solution was considered in cases where data changes frequently and the amount of data is relatively high. It requires allocation of additional resources to achieve a “real-time” synchronization and to avoid

¹¹ For the related field of data science, the European Data Science Academy has conducted extensive studies highlighting such a skill/demand gap all over Europe; cf. Deliverables D1.2 and D1.4 (“Study Evaluation Report 1/2”) downloadable from <http://edsa-project.eu>.

performance degradation of the systems in production. We used this strategy to implement our solution, since it allows accessing the data sources as a virtual RDF graph and benefit from the maturity of relational database systems.

DB in Production The strategy of accessing data in real-time systems does not require allocating additional resources, such as investment in new hardware or software. Since this strategy exposes a high risk for performance degradation of the real-time systems, whereas sensitive information requires high availability and not providing it on time can have hazardous consequences, we did not apply it in our scenario.

7 Related Work

In this section, we give an overview on the development and usage of semantic models in related industrial scenarios: Siemens developed an *ontology-based access* to their wind turbine stream data [12, 15]. The ontology serves as a global view over databases with different schemata. It thus enables SPARQL queries to be executed on different databases without having to take the different schemas into account. Statoil ASA also established a “single point of semantic data access” through an ontology-based data integration for oil and gas well exploration and production data [14]. They thus reduced the time-consuming data gathering task for their analysts by hiding the schema-level complexity of their databases. Ford Motor Company captures knowledge about manufacturing processes in an ontology such that their own developed AI system is able to “manage process planning for vehicle assembly” [21]. Furthermore, Ford examined the potential of federated ontologies to support reasoning in industry [17] as well as detecting supply chain risks [16]. Volkswagen developed a *Volkswagen Sales Ontology*¹² to provide the basis for a contextual search engine [8]. Renault developed an ontology to capture the performance of automotive design projects [6]. With regard to Ontology-Based Data Access (OBDA), Statoil chose the *Ontop* [20] framework because of its efficient query processing. While Siemens initially favored *Ontop* as well, they developed their own system in the end to further optimize stream data processing. Based on these experiences and our own tests of OBDA tools (mainly *Ontop* and *D2RQ*¹³), we chose *Ontop* as well. Regarding semantic models for companies, none of the existing works has specifically addressed machine tools and factory infrastructures. While it is understandable that companies prefer not to share internal details of their methodologies and infrastructure, there is nevertheless very limited evidence of semantic technologies being deployed in the manufacturing industry.

8 Conclusion and Future Work

We have presented a case study on realizing an RDF-based information model for a global manufacturing company using semantic technologies. The information model is centered around machine data and describes all relevant assets, concepts and relations

¹² <http://www.volkswagen.co.uk/vocabularies/vvo/ns>

¹³ <http://d2rq.org>

in a structured way, making use of existing as well as specifically developed ontologies. Furthermore, it contains a set of RML mappings that link different data sources required for integrated data access and SPARQL querying. Finally, it is aligned to relevant industry standards, such as *RAMI* [2] and *IEC 62264* [1], to additionally foster data exchange and semantic interoperability. We described the used methodology to develop the information model, its technical implementation and reported on the gained results and feedback. Additionally, we reflected on the lessons learned from the case study.

As for the enterprise, a high-level ontology is under development to extend the existing information model with our guidance. Its goal is to describe entire business units, their processes and assets for the entire organization.

The use of data-centric approaches in engineering, manufacturing and production is currently a widely discussed topic (cf. the related initiatives concerning Industry 4.0, the Industrial Internet or Smart Manufacturing). The challenges and complexity of data integration are perceived as a major bottleneck for the comprehensive digitization and automation in these domains. A key issue is to efficiently and effectively integrate data from different sources to ease the management of individual factories and production processes up to complete companies. The presented information model is envisioned to serve as a crystallization point and semantic reference in this context.

Future work concerns the continuous translation of relevant industry concepts and standards into RDF as well as their integration and alignment with existing ontologies and vocabularies. In addition, further ontologies for different industry domains need to be developed to enable data integration and semantic interoperability within and between companies. There is also a lack of related business processes and governance models, as it was shown by the case study.

From a technical point of view, further support for ontology-based data access would be needed to achieve the envisioned scalability. While we had good experiences with Ontop as an OBDA framework, complete coverage of SPARQL 1.1 is not yet given and challenging to achieve [5]. Further, R2RML supports only RDB-to-RDF mappings, while other types of data sources are not covered. Although there exist first proposals to extend R2RML beyond relational databases (e.g., RML¹⁴), tool support for these extensions is limited and not yet on the same maturity level as for R2RML.

References

1. IEC 62264-1: Enterprise-control system integration part 1: Models and terminology. Standard, IEC (2013)
2. Adolphs, P., et al.: Reference Architecture Model Industrie 4.0 (RAMI4.0). Status report, ZVEI and VDI (2015)
3. Bellatreche, L., Pierra, G.: OntoAPI: An ontology-based data integration approach by an a priori articulation of ontologies. In: DEXA (2007)
4. Brettel, M., Friederichsen, N., Keller, M., Rosenberg, M.: How virtualization, decentralization and network building change the manufacturing landscape: An Industry 4.0 perspective. *Int. J. Mechanical, Aerospace, Industrial and Mechatronics Engineering* 8(1), 37–44 (2014)

¹⁴ <http://rml.io/spec.html>

5. 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* 8(3), 471–487 (2017)
6. Golebiowska, J., Dieng-Kuntz, R., Corby, O., Mousseau, D.: Building and exploiting ontologies for an automobile project memory. In: K-CAP (2001)
7. Grangel-González, I., Halilaj, L., Auer, S., Lohmann, S., Lange, C., Collarana, D.: An RDF-based approach for implementing industry 4.0 components with administration shells. In: ETFA (2016)
8. Greenly, W., Sandeman-Craik, C., Otero, Y., Streit, J.: Case study: Contextual search for volkswagen and the automotive industry (2011), <https://www.w3.org/2001/sw/sweo/public/UseCases/Volkswagen/>
9. Halilaj, L., Grangel-González, I., Vidal, M., Lohmann, S., Auer, S.: Proactive prevention of false-positive conflicts in distributed ontology development. In: KEOD (2016)
10. Halilaj, L., Petersen, N., Grangel-González, I., Lange, C., Auer, S., Coskun, G., Lohmann, S.: VoCol: An integrated environment to support version-controlled vocabulary development. In: EKAW (2016)
11. Hermann, M., Pentek, T., Otto, B.: Design principles for industrie 4.0 scenarios. In: 49th Hawaii Int. Conf. on System Sciences (HICSS). pp. 3928–3937. IEEE (2016)
12. Kharlamov, E., Brandt, S., Jimenez-Ruiz, E., Kotidis, Y., Lamparter, S., Mailis, T., Neuenstadt, C., Özçep, Ö., Pinkel, C., Svingos, C., et al.: Ontology-based integration of streaming and static relational data with OPTIQUE. In: SIGMOD (2016)
13. Kharlamov, E., Grau, B.C., Jiménez-Ruiz, E., Lamparter, S., Mehdi, G., Ringsquandl, M., Nenov, Y., Grimm, S., Roshchin, M., Horrocks, I.: Capturing industrial information models with ontologies and constraints. In: ISWC (2016)
14. Kharlamov, E., Hovland, D., Jiménez-Ruiz, E., Lanti, D., Lie, H., Pinkel, C., Rezk, M., Skjæveland, M.G., Thorstensen, E., Xiao, G., et al.: Ontology based access to exploration data at Statoil. In: ISWC (2015)
15. Kharlamov, E., Mailis, T., Mehdi, G., Neuenstadt, C., Özçep, Ö., Roshchin, M., Solomakhina, N., Soylu, A., Svingos, C., Brandt, S., et al.: Semantic access to streaming and static data at Siemens. *Web Semantics* (2017, in press)
16. Kim, M., Wang, S.T., Ostrowski, D., Rychtycky, N., Macneille, P.: Technology outlook: Federated ontologies and industrial applications. *Semantic Computing* 10(01) (2016)
17. Ostrowski, D., Rychtycky, N., MacNeille, P., Kim, M.: Integration of big data using semantic web technologies. In: ICSC (2016)
18. Petersen, N., Galkin, M., Lange, C., Lohmann, S., Auer, S.: Monitoring and automating factories using semantic models. In: JIST (2016)
19. Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R.: Linking data to ontologies. *J. Data Semantics* 10, 133–173 (2008)
20. Rodriguez-Muro, M., Kontchakov, R., Zakharyashev, M.: Ontology-based data access: Ontop of databases. In: ISWC (2013)
21. Rychtycky, N., Raman, V., Sankaranarayanan, B., Kumar, P.S., Khemani, D.: Ontology re-engineering: a case study from the automotive industry. In: AAAI (2016)
22. Uschold, M., Gruninger, M.: Ontologies: Principles, methods and applications. *Knowledge engineering review* 11(2), 93–155 (1996)
23. Wache, H., Voegele, T., Visser, T., Stuckenschmidt, H., Schuster, H., Neumann, G., Huebner, S.: Ontology-based integration of information - a survey of existing approaches. In: IJCAI-01 Workshop: Ontologies and Information (2001)