LIS: A Knowledge Graph-Based Line Information System

Irlan Grangel-González[1(✉)], Marc Rickart[2], Oliver Rudolph[2], and Fasal Shah[3]

[1] Robert Bosch GmbH, Corporate Research, Renningen, Germany
Irlan.GrangelGonzalez@de.bosch.com
[2] Robert Bosch GmbH, Automotive Electronics, Reutlingen, Germany
[3] Robert Bosch Engineering and Business Solutions Private Limited, Karnataka, India

Abstract. In a manufacturing enterprise, like Bosch, answering business questions regarding production lines, involves different stakeholders. Production planning, product and production process development, quality management, and purchase have different views on the same entity "production line". These different views are reflected in data residing in silos as Manufacturing Execution Systems (MES), Enterprise Resource Planning (ERP) systems as well as Master Data (MD) systems. To answer these questions, all data have to be integrated and semantically harmonized conciliating the different views in a uniform understanding of the domain. To fulfill these requirements in this specific domain, we present the Line Information System (LIS). LIS is a Knowledge Graph (KG)-based ecosystem capable of semantically integrating data from MES, ERP, and MD. LIS enables a 360° view of manufacturing data for all stakeholders involved while resolving Semantic Interoperability Conflicts (SICs) in a scalable manner. Furthermore, as a part of the LIS ecosystem, we developed the LIS ontology, mappings, and a procedure to ensure the quality of the data in the KG. The LIS application comprises many functionalities to answer business questions that were not possible without LIS. LIS is currently in use in 12 Bosch plants semantically integrating data of more than 1.100 production lines, 16.000 physical machines, as well as more than 400 manufacturing processes. After the rollout of LIS, we performed a study with 21 colleagues. In general, the study showed that LIS in particular, and KG-based solutions in general, paves the way of exploiting the knowledge in manufacturing settings in a reusable and scalable way.

Keywords: Knowledge Graph · Industry 4.0 · Smart Manufacturing · Semantic Data Integration · Ontology

1 Introduction

In numerous corporate-level manufacturing organizations, the *digital transformation*, which has seen its onset a few years back and is still ongoing, is considered a key-enabler in order to remain competitive as a company. In essence, digital transformation involves not only establishing connectivity, but also driving

C. Pesquita et al. (Eds.): ESWC 2023, LNCS 13870, pp. 591–608, 2023.
https://doi.org/10.1007/978-3-031-33455-9_35

accessibility and especially an understanding of the increasingly present digital assets. An essential part of the digital transformation thus is faster provisioning of easily digestible data to answer business-crucial questions used in decision-making processes. Data required for production planning purposes, traceability of products, production process optimization, and production operation is currently distributed in different IT-systems as there are ERP, MES, and MD. Also non-IT-available data of production process experts is of high interest, but is not directly accessible. These data typically rely in the head of experts as well as in unstructured formats, e.g., intranets, word, and pdf documents. The collection of all data required for decision-making causes high efforts in terms of time and cost to draw the right conclusions. Typically, data reside in above described silos which are not interconnected but contain semantically related terms with redundant but inconsistent information. Unlocking the potential of these data in a combined set-up by creating knowledge out of it is one of the biggest challenges of competing enterprises. Of particular importance in this setting is to achieve semantic interoperability among all data assets. Achieving semantic interoperability is challenging due to the fact that SICs demand to be resolved accordingly. This, in turn, requires domain knowledge and organizational consensus [12]. To that end, and particularly in the context of the Industry 4.0 movement, KGs have demonstrated to be a solution for that problem [1,6,9,14,16]. At this point, it is worth to give a word on SICs. In general, interoperability can be defined as a measure of the degree to which diverse systems, organizations, and/or individuals are able to work together to achieve a common goal. A more detailed overview on SICs in this context is available in [5,11].

To tackle this problem, we propose a KG-based solution for efficient integration and quick access of manufacturing data. The above described solution has been implemented at Bosch in a KG-based approach called *L*ine *I*nformation *S*ystem (`LIS`). We show the implementation of this approach on integration of heterogeneous data sources by a specially developed user interface according to user experience requirements. The proposed approach enables semantic harmonization of the data on production lines as this was found already in precedent works [3,8]. Key to the success of the presented approach is the encoded domain knowledge in the ontology combined with the semantically enriched KG mappings. The mappings bridge the mismatch between the data layer and the ontology layer as semantically heterogeneous data from different sources using distinct representations are integrated by these mappings to resolve SICs at the level of the KG. This allows querying available data in an integrated way by exploiting the semantics of the provided information. In the following, we present the contributions of the `LIS` approach. *(i)* We developed the `LIS` ontology describing the concepts and properties of a production line together with domain knowledge from various experts. This ontology is the basis of the KG-based approach for semantic data integration in manufacturing. *(ii)* We built a mapping layer that semantically connects the `LIS` ontology to the ERP, MES, and MD data sources. *(iii)* We integrated data from 12 Bosch plants around the world while providing a structured way of dealing with SICs. *(iv)* We devel-

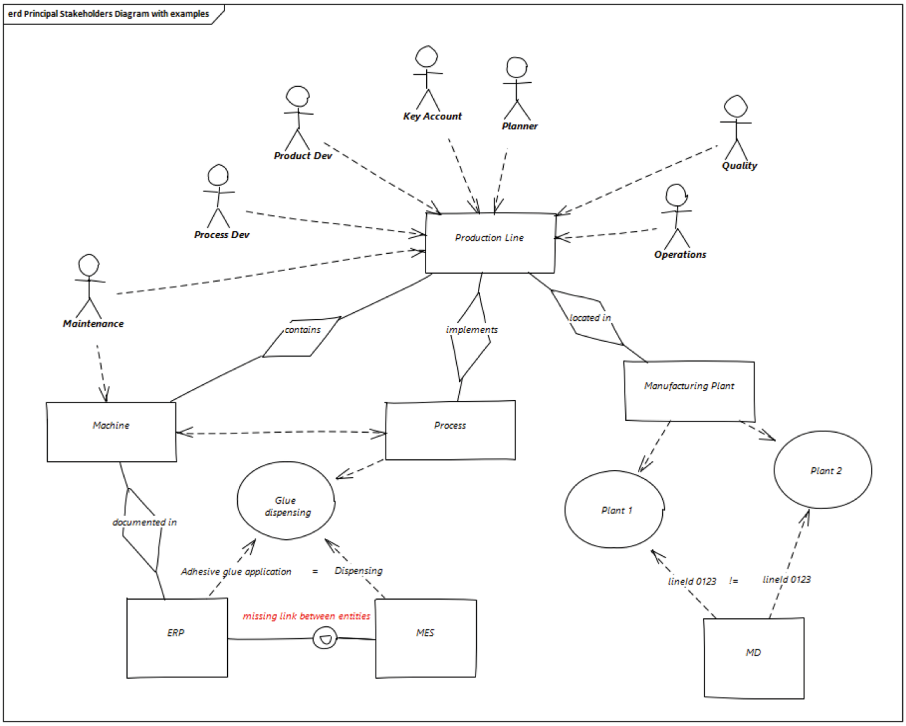

Fig. 1. Motivating example. The *Production Line* exposes multiple entities, e.g., *Machines* and *Processes*, to different stakeholders such as process and product developers, key accountants, planners, maintenance crew, quality engineers, and shopfloor operations. These stakeholders hold differing responsibilities and interest in these entities, hence leading to fundamentally different views. Different views on top of the data silos generate SICs. The example depicted is on the *Process* entity, which lives in both data assets *ERP* and *MES* under differing names yet referring to the exact same process in reality. One further example of such conflicts is the use of an identical production line identifier *lineId* used separately for completely different lines on two geo-locations.

oped the LIS ecosystem comprising a dedicated Web application that access the semantically harmonized data; a validation process to ensure that the data have the demanded quality so that it can be used by the domain experts, as well as a data sharing service on top of the LIS KG. In addition, LIS acts as a master data management system and a procedure to share as well as reporting on top of the semantically harmonized data. *(v)* We provided an automatic procedure to check that the data in the KG is correct with a set of dashboards to make the data quality transparent.

The paper is organized as follows: Sect. 2 provides background and motivation for this work. Section 3 describes the available components for the KG-based approach and the challenges for integrating data from these systems with the correct semantic description. In Sect. 4 we performed and study to get the feedback from the main stakeholders. In Sect. 5 we discuss the lessons learned. Section 6 presents the related endeavours and Sect. 7 concludes the paper with an outlook and future work.

2 Motivation

One of the larger pain points observed is that information is spread heterogeneously within the organization, be it in terms of the technology applied or the terminology in use within the data sets. This drives a culture of "expertism" and siloed applications. Instead of company-wide shared access to knowledge, a funneling approach that requires mostly human resources to retrieve and provide information by case can lead to bottlenecks and delays in the provisioning of those business-relevant details. Data collection to answer consumer-specific questions hence becomes time consuming and tedious, already starting with the search for the right expert or domain owner. In case the question asked renders certain levels of complexity, multiple entities and domain experts may become necessary, adding further to the resource constraints. If the right expert cannot be identified, typically data exploration and search with the risk of misinterpretations based on incorrect assumptions or outdated records can be the result, possibly posing risk to the business. The entities present in a manufacturing organization with varying functions and responsibilities, may have very differing and domain-specific questions regarding production lines.

A production line consists of a defined number and sequence of production processes with specified capabilities to manufacture or assemble a product until ready for shipment to the customer. These processes are realized by physical assets or machines. Along the processes, value is added to the product by materials and resource consumption, e.g., operations personnel, machine wear, maintenance and so forth. In this context, a plethora of questions can arise. While some plants make use of the MES in a similar way, lack of standardization across all plants causes differences in the usage and consequently population of certain fields. Next, we present some concrete examples (cf. Fig. 1).

1. Manufacturing planning generates a forecast about production requirements from one up to eight years. This requires a complete overview about available production lines in the plants in an international production network (IPN). Furthermore, these data need to be correlated with the customers demand of products to take investments decisions. The identification of the production lines is usually a number which is unique only in one plant but not across the IPN. As concrete example we found the case that in two plants, *different lines were having the same identifier but actually were referring to distinct lines in the reality*. Thus, a SIC of type *homonym* is present here (cf. [11]).
2. Data about machines and production processes with their names are distributed in different ERP and MES systems. A process is implemented by a machine as a physical asset. The description of a machine refers in most

cases to the process and can be found like this in the ERP system. In MES the description of a process is defined by its name. There is no possibility to link *physical assets with logical processes* in the current setting. Thus, a SIC of type *missing items* is detected here.

3. Due to missing standardization rules, processes have different names in one or even many MESs but referring to the identical process in the reality. To optimize the adhesive dispensing process[1] across multiple manufacturing plants in the IPN, process development requires the count, location, and manufacturer of assets implementing this process along with a means to identify the sources for process data. This data is usually connected to the name of the production process. Unfortunately the name of the process is not standardized and many names for our example adhesive dispensing are found in the systems as e.g., *dispense sealant* and *dispensing sealing compound*. Here a SIC of type *synonym* is required to be resolved.

To achieve a semantically correct answer, SICs have to be resolved. In addition, a scalable approach has to provide a common semantic understanding of the domain that can be exploited and further enhanced. The challenges which we address here are the time consumption and iterations found in human-based communication to retrieve information from knowledge domains in a manufacturing enterprise, i.e., Bosch. Additionally, the risk of interpretative errors causing misinformation, and the SICs present in data queries employing multiple tables or sources. To tackle these challenges, we propose `LIS`: a KG-based approach to enable a semantically harmonized access to knowledge for stakeholders within an organizational unit of Bosch.

3 `LIS`: A Knowledge Graph-Based Line Information System

In this section, we present a KG-based approach for semantically integrating data and its specific application at Bosch. Based on the architectural view in Fig. 2, we describe the key elements required to establish the `LIS` ecosystem from bottom to top.

3.1 Data Sources

In this section, we describe the key data sources to provide the information from various domains, i.e., from MES, tools, i.e., ERP, and reference, i.e., MD.

MES. Manufacturing execution systems act as control instance to assert proper sequence, completion, and performance of the work in progress during manufacturing. As such, MES is a primary source of information in each value-adding

[1] A production process used to mix, meter, and dispense adhesives, i.e., glues and thermal interface materials on one or more components of a product to bond those components with each other to withstand defined stress-levels.

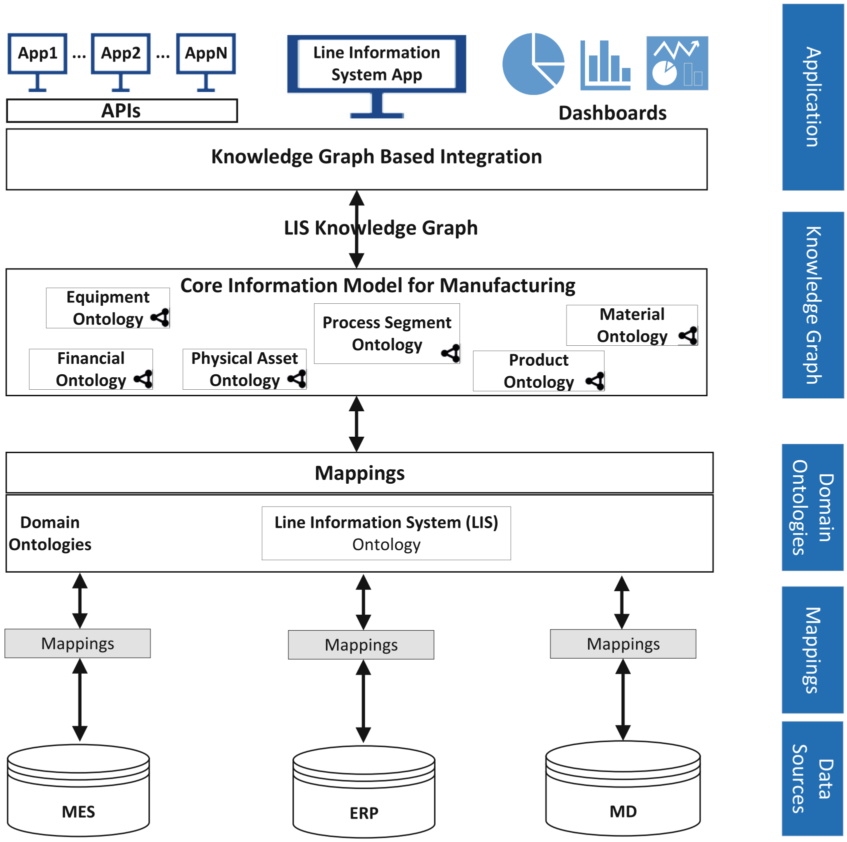

Fig. 2. The architecture in use for the proposed KG-based *Line Information System (LIS)* The LIS architecture comprises five layers, i.e., Data Sources, Mappings, Domain Ontologies, Knowledge Graph, and Application. *Data Sources* provide the information combined with the *Domain Ontologies* to shape the LIS KG. *Mappings* ensure the correct connection between the tables and fields within the underlying structured sources. The LIS Ontology is maintained in a central repository, and serves as the merging layer of interoperability between the sources. The KG provides semantically connected data, which can be retrieved from the *Application* via *SPARQL* queries, visualized using dashboards, as well as served to other applications via an API service.

step, as it typically involves part, machine, and process information, and possibly even results. In case of LIS, MES data serves as the logical view on the production line, as it abstracts the actual physical implementation by an asset or machine from the viewpoint of the mere process. In other words, MES describes what needs to happen in which sequence to a product, but does not interfere with the actual doing which is taken over by the equipment. As MES is close to real-time critical for manufacturing, in an effort to decouple loads the data sources are split into OLTP and OLAP stores. This aims at satisfying both, the immediate execution request from the connected equipment and the retroactive analysis of records taken. OLAP sources thus provide the point of connection.

ERP. Tools for enterprise resource-planning are mainly centered around the financial and controlling aspects of the company. In case of a manufacturing line, information such as asset tags, procurement or maintenance cost, depreciation cost, total cost of ownership, spare part stocks, and similar is present. In contrast to MES, ERP is more involved in the aspects of planning activities and customer-delivery fulfillment rather than day to day operations on the shop floor.

MD. Master data are crucial for clustering, categorizing, and querying correctly. Along other aspects, MD is considered to be a single source of truth with the intent to serve the organization in the endeavor of like treatment of data assets. In our example, the naming of plants, lines, and processes is to be considered unique and present. As with any organization undergoing the digital transformation, however, we have found this not to be the case in every situation. Thus, missing or redundant records have to be addressed. Since improvement data quality is a process and cannot be achieved within a short duration, intermediate solutions based on mere comma-separated value files have been installed as mitigating action. With increasing maturity of the organization, these sources can be replaced by their improved versions at a later time, requiring updates to the mappings but concealing these changes to the layers above. Hence, the application and users will not notice those changes. As a result of organically grown organizational diversity, while integrating our various sources, we have observed SICs frequently.

3.2 Mappings

A *mapping* is a set of assertions specifying how the classes and properties of the ontology are populated with data from our sources described in Sect. 3.1.

```
INSERT DATA {
GRAPH <http://bosch.com/kg/lis#> {
  ?line_instance a lis:LISLine ; lis:lineId ?line_id .
} }
WHERE {
    ?uri    a   tmpschema:MESClass ;
                  tmpschema:plant_id    ?plant_id ;
                  tmpschema:system_id   ?system_id ;
                  tmpschema:line_number  ?line_number .
   # generate unique key
BIND (CONCAT("Plant", ?plant_id, "_System", ?system_id, "_line", ?
     line_number) AS ?line_id)
   # instantiate classes based on their unique keys
BIND (IRI(CONCAT("http://bosch.com/ontologies/lis#LISLine_", ?line_id))) AS
      ?line_instance) }
```

Listing 1. Example of a mapping for URI creation. For every production line across all plants a unique key is generated based on several attributes to be consequently used in the instantiations of the `lis:LISLine` class.

Accordingly, they link the records from the sources to the concepts described in the ontology. Typically, mappings are used to resolve SICs. One of the central features for LIS is the creation of correct URI for every instance of the classes. As described in the motivation, lines are not uniquely identify across different plants causing a SIC of type homonym. We resolve this SIC by generating an unique identifier of the lines combining the plant and the system to where they belong in addition to its number (cf. Listing 1).

```
INSERT DATA {
GRAPH <http://bosch.com/kg/lis#> {
  ?process_instance a proc:Process ;
                      proc:processNumber  ?process_number;
                      proc:processName    ?process_name .
} }
WHERE{
      ?uri  a       tmpschema:MESClass ;
                tmpschema:process_number    ?process_number .
                tmpschema:process_name      ?process_name .

BIND (IRI (CONCAT("https://open-manufacturing.org/ontologies/I4.0/
     ProcessSegment#Process_", ?process_number))) AS ?process_instance)
}
```

Listing 2. Example of a mapping for creating an instance of process. Processes instances are created based on process number since the process name can vary or not exist. This resolves the synonym SIC.

Next, to resolve the synonym SIC we first employed the Listing 2. In this particular case, despite the process names may be distinct in different plants, the process number remains the same. Thus, we created the process instance based on the process number. We do not intent here to provide an exhaustive list of all the SICs that occur in the manufacturing context. On the contrary, the goal is to set the basis to resolve some of the most important ones and show that the `LIS` approach can scale in this context.

3.3 LIS Ontology

The `LIS` ontology serves as an abstraction over the manufacturing-related sources and is used by the domain experts to formulate queries over the data. The `LIS` ontology reuses most of the concepts and relations described in the set of ontologies that are part of the Core Information Model for Manufacturing (**CIMM**) [3]. This set of ontologies are published in the context of the Industrial Digital Twin association[2]. With the aim to create the `LIS` ontology, several iterations were performed with different groups of experts. It is important to note that the aim was to cover, as much as possible, the manufacturing domain in `LIS` by directly reusing the concepts included in CIMM. For instance, a Bosch plant which is described in CIMM as a class belonging to the equipment ontology, i.e., `eq:Plant` (cf. Fig. 3). Then, specific concepts and relationships were introduced to capture the particularities of the `LIS` approach. These concepts are not included in the CIMM but are demanded to be incorporated in the `LIS` ontology to capture the requirements of the manufacturing domain and the proposed approach.

2 https://github.com/eclipse-esmf/esmf-manufacturing-information-model.

3.4 LIS Knowledge Graph

After executing the mappings, the data that rely on the three data sources under use is transformed into an RDF KG. It is important to realize that every data transformation process may lead to data leakage. We took this fact into account and developed a procedure to ensure that no data leakage is occurring. For instance, part of the procedure is focused on determining whether the number of instances of the classes that are built, e.g., `lis:LISLine`, `eq:Plant`, `proc:Process`, match with what exist in the data sources. `LIS` is intended to be used by manufacturing experts and for them, the quality of the presented data is of key relevance. The `LIS` KG comprises data from 12 Bosch plants around the globe, distributed in different regions and time zones. Additionally, it contains semantically integrated data of more than 1.100 production lines, more than 16.000 physical machines, as well as more than 400 manufacturing processes.

Writing Links into the `LIS` KG. In what follows, we outline two of the main requirements where we need to not only use the typical ETL process but also to enrich the LIS KG by including the input from the domain experts.

Linking Physical Machines and Logical Work Units. Despite the huge effort employed in data harmonization, an automatic link between physical machines presented in ERP and logical stations in MES was not possible to realize. Thus, we provide a user interface to enable domain experts for the manual creation of these links. Of core relevance here is to note that this approach is not only transforming the data that relies in existing sources but also creating links that do not exist. In this case, the experts in the plants are linking a physical machine with a logical work unit. The link is described in the ontology with the property `c:isImplementedBy` which has the `eq:WorkUnit` as domain and `phys:Machine` as range. This link is then written and maintained in the `LIS` KG.

Master Data Management. Combining data from different plants across countries and time zones is a challenging task, particularly, when it comes to semantic data integration. For instance, the line names were maintained manually in CSV files in the plants. We used the CSV files as starting point. The files were ingested into the `LIS` KG. Then, `LIS` also provides the possibility to maintain these and other data relevant to many different stakeholders in a centralized and harmonized manner. In this case, the line names are not maintained in the local CSV files anymore rather using `LIS` as a centralized solution.

3.5 Application

`LIS` is focused to create an ecosystem that integrates and harmonizes different heterogeneous data sources in the manufacturing domain. The integration has the goal to enable users to answer business-relevant questions (cf. Sect. 2). To that end, we designed the Application layer at the top of the `LIS` architecture. This layer comprises: 1) the web-based `LIS` application, with different

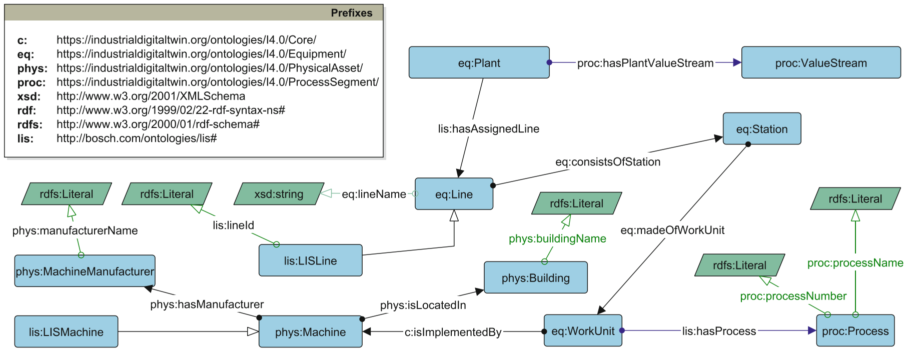

Fig. 3. Main classes and relations of the LIS Ontology. The LIS Ontology reuses classes from the **CIMM**, e.g., `phys:Machine` and `eq:Line` and further specifies subclasses, e.g., `lis:LISMachine` and `lis:LISLine` to reflect a more fine grained representation of the concepts in the manufacturing scenario in the Bosch Plants.

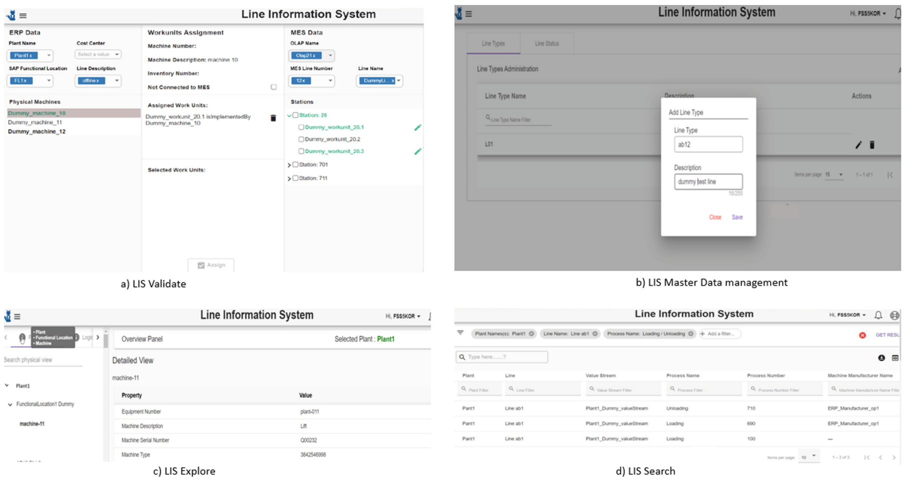

Fig. 4. Core functionalities of the LIS application. The Figure depicts four of the core functionalities of the LIS application. a) The Validate, which enables users to manually create links between ERP and MES; b) The Master Data management, which allows users to add new Lines or Statuses; c) The Explore, that aims to browse the integrated data of the plants; and d) The Search, which permit to access the data in any order on various entities, e.g., lines, processes, machines.

functionalities to browse and edit the LIS KG; 2) APIs service to enable data sharing by accessing the LIS KG from other applications; and 3) a set of dynamic dashboards to directly interact with LIS KG covering a huge range of possible combinations of data requirements to answer business questions.

```
SELECT DISTINCT ?plant_id ?work_unit_label ?process_name ?process_number ?
    building_name ?manufacturer_name
WHERE {
  ?plant eq:plantId ?plant_id ;
         lis:hasAssignedLine ?line .
  ?line eq:consistsOfStation  ?station ;
        eq:madeOfWorkUnit ?work_unit .
    ?work_unit  c:isImplementedBy  ?machine ;
                rdfs:label         ?work_unit_label ;
                lis:hasProcess     ?process .
    ?machine    phys:isLocatedIn  ?building ;
                phys:hasManufacturer ?manufacturer .
    ?building   phys:buildingName ?building_name .
    ?manufacturer phys:manufacturerName ?manufacturer_name .
    ?process proc:processNumber  ?process_number ;
  OPTIONAL{
    ?work_unit c:isImplementedBy  ?machine .
    ?process proc:processName  ?process_name .
  }
```

Listing 3. SPARQL query linking data from the physical assets and the logical processes

`LIS` Web Application. The `LIS` Web application is developed with KG at its crux. The application comprises different functionalities like explore, search, master data management, and linking information across data silos. Figure 4 depicts four of the core functionalities of the `LIS` web application. As discussed in Sect. 3.4 `LIS` provides a user interface for domain experts to perform the link creation between ERP and MES systems. This function is named "Validate" since for the domain experts they are performing a validation on top of the data. Currently, the application supports two types of validation. Firstly, it allows domain experts to select their respective plant, once selected it proceeds to query the KG and fetch the list of physical machines from the ERP system. These machines are manually mapped to the respective work units under MES data by the experts. Links are stored in the form of triples and are colored green once assigned (cf. Fig. 4a). The second form of the validation is done when a physical machine cannot be mapped to the MES system and is marked as "Not Connected to MES", displayed in red color. This leads to the creation of new relations within the KG which were not present earlier and hence semantically connecting the two data silos.

Another important feature of `LIS` is the data management system. It replaces the traditional way of maintaining CSV files with a centralized and semantically harmonized system. For instance, lines can be maintained locally but linked to the central line entity. The newly created data is appended into `LIS` KG as RDF triples (cf. Fig. 4b). Even more, the `LIS` application provides functionality to browse/explore and search across the KG (cf. Figs. 4c and 4d). Search allows domain experts to access the data in a non-sequential manner based on values selected, e.g., plants, lines, etc. Listing 3 showcases the query used in the search feature to fetch data like work unit, process name, and number from the MES system along with building and manufacturer name from the ERP system. This is possible since `LIS` enables the possibility to create links between the two systems. Thus resolving the missing items SIC. Particularly, the resolution of this SIC has made LIS very attractive for other applications since these links are typically non existing in the manufacturing domain.

LIS API Service. In the manufacturing context, there exist other applications that require the data managed by LIS. Typically, these applications create their own data silos on top of what already existed. LIS enables the reuse of data between these applications by providing an easy way of sharing the data. With that we ensure that the applications access to the same harmonized data which is also maintained and semantically curated in LIS. To meet this requirement, we implemented an API service based on the needs of every application. The requirements of the consumer applications are transformed into SPARQL queries. They are then executed against the LIS KG and then served as a JSON-LD payload.

Dashboards on Top of the LIS KG. The LIS application is designed to cover main business questions regarding manufacturing. However, once the data from the plants is integrated many more questions can be answered. The complete set of possible questions is not covered by the LIS application. Thus, as a part of the LIS ecosystem, we provided the possibility to create dashboards on top of the LIS KG. The goal of the dashboards is twofold. First, to be able to provide to a different set of stakeholders the freedom of interacting with the semantically harmonized data to get answers to many different business questions. Second, the dashboards are also used by the developers in order to continuously validate the data quality. For instance, to ensure that the same amount of instances of processes, lines, machines is on the LIS KG as in the original sources.

4 Feedback of Main Stakeholders

In this section we report on the results of a study performed to collect feedback from the main stakeholders. A questionnaire is designed with eight questions to specific topics of interest. Seven questions are measured following the five-point Likert scale with values ranging between Strongly agree and Strongly disagree. We employ the mode to measure them since the data have ordinal nature. The first five questions, i.e., from **Q1** to **Q5**, are specifically evaluating LIS as a solution. **Q6** collects the current vision regarding the use of KGs for manufacturing and engineering domain. Further, the importance and availability of training courses in this regard in the manufacturing industry is included in **Q7**. Finally, **Q8** is a free-text question to collect feedback w.r.t. the possible obstacles that the stakeholders perceive to apply solutions like LIS. Table 1 outlines the questions and the mode value of the answers, depicting what is the general consensus of the stakeholders. A total of 21 stakeholders participated in the study. They are divided in three different groups of seven participants. The characteristics of the groups are as follows: **Managers**: They are business responsible and as such have a high interest in trust-worthy data. In digital transformation, they have the role of project sponsors, dedicate resources to projects, and need to trust in the technology. As the provided data offers completely new insights they are challenged to think out of the box for further business improvements. **Users**:

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

Question (with Mode values M we provide general consensus of our survey)	**M**
Q1. Did the developed LIS semantic model (ontology) meet your expectations?	Agree
Q2. How do you evaluate the perceived benefit of LIS?	Agree
Q3. How do you evaluate the benefit of data curation and integration in the LIS and its impact on data quality?	Strongly Agree
Q4. Do you think investing in knowledge graph-based technologies as LIS is based on can result in a good Return of Invest (ROI) in future?	Strongly Agree
Q5. Do you consider a high value of reuse data from LIS as a semantically curated central Master Data System in your organization?	Strongly Agree
Q6. Do you consider knowledge graph-based technologies fit for usage in the manufacturing and engineering domain?	Strongly Agree
Q7. Do you think a broader community should achieve the knowledge about and get trained in knowledge engineering?	Strongly Agree
Free-text questions:	
Q8. What would be the biggest obstacles for the successful use of knowledge graph-based technologies at Bosch?	

Mostly domain experts use the application and interact with the data. Partially they are also use case providers who specified their requirements for consideration for implementation. Their feedback is strongly required to improve the solution. They are mainly responsible for data curation and validation to raise the quality of the data in the systems and avoid SICs. **Developers**: This group has the deepest insight in the technology, i.e., understanding of the functionality, vision of the potential, but also the risks of this new technology. The developers have a strong identification with the product as they know the value of a solution like this for the organization. Figure 5 depicts the results of the questionnaire for the first seven questions. Regarding **Q8** we received some useful feedback w.r.t. possible obstacles for the adoption of LIS but also regarding KG-related technologies at Bosch. For instance, “degree of novelty of the technologies employed very high”, and “tools for developing and maintaining ontologies and KGs need to get easier to use for non-experts” thus the need to have KG experts to ensure the success is crucial; “huge effort to bring humans to a shared understanding”, and “management is lacking the deeper understanding of the importance of semantics for data sharing, reusability, and its economic impact” requires better communication for management level to achieve the right awareness for the topic.

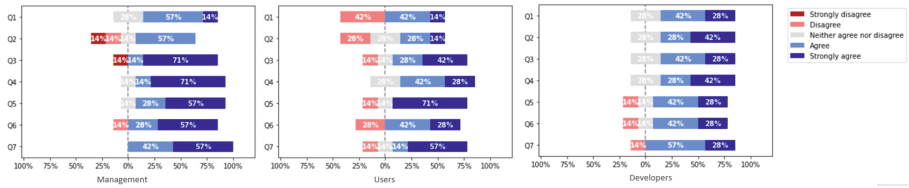

Fig. 5. Results of the questionnaire. The Figure shows the results of the questionnaire divided in three groups, i.e., Management, Users, and Developers. In general, the feedback from stakeholders is highly positive regarding `LIS` as well as for KG-based solutions.

5 Discussion and Lessons Learned

In this section, we analyze the outcome of the performed study. Initially, we discuss the results of the study. Further, we describe main lessons learned. The overall result is a very positive reception of the approach as the majority of the answers of all groups agreed or even strongly agreed with the `LIS` approach and the technology behind it. Though the group of users shows more a critical view on the application as this might be related to their role of use case providers and like this a high expectation is present. Moreover this group has the task of data cleansing which causes extra effort not seen before as this limits the advantage of usage of the application in their view. The question related to the perceived benefit of `LIS` was in the group of managers connected with some disagreement due to wrong expectation for a fast ROI than a strategic development. Furthermore the application revealed a data quality in the investigated systems which raises directly efforts to be taken in order to clean the data properly for the responsible system owners. From management side this is noted as a risk as possibly high personnel effort is required to ensure the data quality. Nevertheless the overall introduction of KG-based technology is seen as positive in all three groups and a strong belief in a success of KG-based technologies is present in the community. Further use cases were discussed with managers and users group to see a benefit of reusing the data of the `LIS` as this application as once the data set is curated, it can be used as a Master Data System for other applications. The developers group has a different view on this topic as technological requirements of a Master Data System is in principal possible but might require some adaptions. The usage of KG-based technologies in the domains engineering and manufacturing is perceived as appropriate to introduce as it offers more opportunities specifically for data re-use. All groups give a clear feedback that the introduction of a new technology requires setup of training possibilities to face future tasks with sufficient skills and competency. The result of the good acceptance of the application might be surprising, because compared to the developer group the managers and users have not the detailed insight of the technology used. Both groups see only a graphical user interface which hides any underlying technology.

In general, the `LIS` approach was well acknowledged in the organization. In the following we describe the main lessons learned for a production enterprise.

Main difficulty in the beginning was to understand how to use available data in different systems and to connect them to one and the same semantic model. The data base received from the different plants was required to be cleaned, sorted, and newly adjusted. These data bases are to be consumed in a way that users from other domains as manufacturing, e.g., product development or process engineering were able to understand the provided data. Corresponding table structures would have to be defined with their data models. The effort would have been to fit into these models. However, the reality is much more heterogeneous due to historical circumstances. In this respect, the use of a KG is of core importance here, as the ontology and thus also a semantically harmonization of the data can be handled and integrated much more flexibly. The possibility of easily adding further information by mapping of new data sources was found to be most efficient in terms of time and cost. The lesson learned of this approach was the power and flexibility of the KG-based technology. Even without data cleansing the information at once glance saves already many engineering hours to search for the right information in the organization. `LIS` is far more than a proof on concept and it is currently in use in one Bosch division. `LIS` will be stabilized as a product for re-use in other Bosch divisions. Though it has some limitations due to actuality of data[3], it serves already as a single source of truth for many other applications depending on data from manufacturing domain. Next, we list the essential points of our lessons learned.

Integrated View of Data Provided by LIS Ontology: For the first time, data from ERP, MES, and MD was integrated. This enabled the possibility that domain experts formulate their information needs as queries, as they were working on the level of incompatible raw data sources before. The integration of MD was required to harmonize the data available in the plant specific master data systems.

Impact on Data Quality: For the first time it was possible to make the data quality of the integrated systems transparent. The findings were used to fix undetected failures which were not present to the data stewards responsible for these systems. For example, the location identifier in the MES was in some cases too short and will probably lead into a failure. A correction of the location identifier was initiated.

Involvement of Domain Experts: Involving domain experts in the process of building, and maintaining the `LIS` approach is of paramount importance for its success. Training in KG-based technologies are provided to help experts to better understand the problems which `LIS` is tackling. Despite initial resistance to the KG-related technologies, the impact of the semantically integrated data has convinced most of the domain experts.

[3] The `LIS` KG is generated periodically once a day. The plant ERP and MES systems are mirrored accurately with a one-day uncertainty which is due to the definition of the use-cases acceptable.

LIS as Data Provider for Different Applications: Associated to the previous lesson, we observed an increasing need for other applications requesting data to LIS. This re-use of data enables the whole organization to refer with its questions to one single source of truth. The availability of a reliable master data system with all information about production lines is of high interest for implementation in other applications. The need for information typically arises in non-manufacturing domains such as product engineering or controlling and planning. By connecting to the LIS via an API these data can be retrieved now easily with the additional benefit that it is actual.

6 Related Work

In this section, we survey the current state of the art presenting similar approaches as LIS. Yan *et al.* [17] outlines an approach for building KGs in the context of manufacturing equipment. In this work, several data sources are integrated with particular focus on lathes, conveyors, and robots. Cheng *et al.* [2] designed an ontological approach for dealing with production lines in the context of Industry 4.0. Five ontologies for representing knowledge regarding production lines are developed, i.e., Base, Product, Device, Parameter, Process. Petersen *et al.* [13] utilize an RDF-based information model to semantically integrate different data sources in a manufacturing company. Kasrin *et al.* [7] present an RDF-based framework for system architecture and domain modelling to create semantic data management solutions for manufacturing. The framework comprises heterogeneous data utilising data lake and enhances knowledge with semantic metadata. Kalayci *et al.* [6] outlined a virtual KG approach for semantically integrating data for Surface Mount Technologies. Mehdi *et al.* [10] tackle a similar problem for performing semantic data integration on top of manufacturing related data. In the previously mentioned works, no extended and concrete implementation and rollout of such a KG-based solution was performed. Furthermore, no scalable way of describing and semantically harmonizing and resolving SICs in manufacturing context is provided.

7 Conclusions and Outlook

In this paper, we presented the LIS, a Knowledge Graph-based ecosystem capable of semantically harmonizing and integrating manufacturing data at Bosch. LIS enables a 360° view of manufacturing data while resolving Semantic Interoperability Conflicts in a scalable manner. Furthermore, we developed the LIS ontology, mappings, and a procedure to ensure the quality of the data in the LIS KG. The LIS application comprises many functionalities to answer business questions that were not possible without LIS. LIS is currently in use in 12 Bosch plants integrating data of more than 1.100 production lines, 16.000 physical machines, as well as more than 400 manufacturing processes. The feedback collected showed that LIS in particular, and KG-based solutions in general, paves the way of exploiting the knowledge in manufacturing settings in a reusable and scalable way. Despite the efforts employed in the data integration to create the LIS KG, still some room for improvement exists regarding its completeness and the semantic disambiguation of some of the core entities, e.g., processes. In future work, we envision to develop entity disambiguation algorithms on top of the LIS KG. Moreover, we plan to applying Machine Learning techniques to further improve the completeness of a manufacturing KG like the one presented here, e.g., in terms of Link Prediction [4,15].

References

1. Buchgeher, G., Gabauer, D., Gil, J.M., Ehrlinger, L.: Knowledge graphs in manufacturing and production: a systematic literature review. IEEE Access **9**, 55537–55554 (2021)
2. Cheng, H., Zeng, P., Xue, L., Shi, Z., Wang, P., Yu, H.: Manufacturing ontology development based on Industry 4.0 demonstration production line. In: 2016 Third International Conference on Trustworthy Systems and their Applications (TSA), pp. 42–47 IEEE (2016)
3. Grangel-González, I., Lösch, F., ul Mehdi, A.: Knowledge graphs for efficient integration and access of manufacturing data. In: 25th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA, Vienna, Austria, 8–11 September, pp. 93–100. IEEE (2020)
4. Grangel-González, I., Shah, F.: Link prediction with supervised learning on an industry 4.0 related knowledge graph. In: 26th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA, Vasteras, Sweden, 7–10 September, pp. 1–8. IEEE (2021)
5. Jirkovský, V., Obitko, M., Marík, V.: Understanding data heterogeneity in the context of cyber-physical systems integration. IEEE Trans. Ind. Inform. **13**(2), 660–667 (2017)
6. Kalaycı, E.G., et al.: Semantic integration of Bosch manufacturing data using virtual knowledge graphs. In: Pan, J.Z., et al. (eds.) ISWC 2020. LNCS, vol. 12507, pp. 464–481. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-62466-8_29
7. Kasrin, N., Qureshi, M., Steuer, S., Nicklas, D.: Semantic data management for experimental manufacturing technologies. Datenbank-Spektrum **18**(1), 27–37 (2018)
8. Kharlamov, E., et al.: Semantic access to streaming and static data at siemens. J. Web Semant. **44**, 54–74 (2017)
9. Li, X., Lyu, M., Wang, Z., Chen, C.H., Zheng, P.: Exploiting knowledge graphs in industrial products and services: a survey of key aspects, challenges, and future perspectives. Comput. Ind. **129**, 103449 (2021)
10. Mehdi, A., Kharlamov, E., Stepanova, D., Loesch, F., Grangel-Gonzalez, I.: Towards semantic integration of Bosch manufacturing data. In: Proceedings of ISWC, pp. 303–304 (2019)
11. Melluso, N., Grangel-González, I., Fantoni, G.: Enhancing industry 4.0 standards interoperability via knowledge graphs with natural language processing. Comput. Ind. **140**, 103676 (2022)
12. Nilsson, J., Sandin, F.: Semantic interoperability in industry 4.0: survey of recent developments and outlook. In: 2018 IEEE 16th International Conference on Industrial Informatics (INDIN), pp. 127–132. IEEE (2018)

13. Petersen, N., Halilaj, L., Grangel-González, I., Lohmann, S., Lange, C., Auer, S.: Realizing an RDF-based information model for a manufacturing company - A case study. In: ISWC-17, pp. 350–366 (2017)
14. Sjarov, M., Franke, J.: Towards knowledge graphs for industrial end-to-end data integration: technologies, architectures and potentials. In: Behrens, B.-A., Brosius, A., Drossel, W.-G., Hintze, W., Ihlenfeldt, S., Nyhuis, P. (eds.) WGP 2021. LNPE, pp. 545–553. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-78424-9_60
15. Xia, L., et al.: A knowledge graph-based link prediction for interpretable maintenance planning in complex equipment. In: 2022 13th International Conference on Reliability, Maintainability, and Safety (ICRMS), pp. 301–305. IEEE (2022)
16. Yahya, M., Breslin, J.G., Ali, Intizar, M.: Semantic web and knowledge graphs for industry 4.0. Appl. Sci. **11**(11), 5110 (2021)
17. Yan, H., Yang, J., Wan, J.: KnowIME: a system to construct a knowledge graph for intelligent manufacturing equipment. IEEE Access **8**, 41805–41813 (2020)