

A benchmark dataset with Knowledge Graph generation for Industry 4.0 production lines

Muhammad Yahya^{a,*}, Aabid Ali^b, Qaiser Mehmood^c, Lan Yang^c, John G. Breslin^a and Muhammad Intizar Ali^d

^a *Confirm Centre for Smart Manufacturing, National University of Ireland Galway, Ireland*

E-mails: m.yahya1@nuigalway.ie, john.breslin@nuigalway.ie

^b *Manufacturing, Forward Group Lahore, Pakistan*

E-mail: abid.zalan@yahoo.com

^c *Insight centre for Data Analytics, National University of Ireland Galway, Ireland*

E-mails: qaiser.mehmood@nuigalway.ie, lan.yang@nuigalway.ie

^d *Dublin City University, Ireland*

E-mail: ali.intizar@dcu.ie

Editors: Bahar Aameri, University of Toronto, Canada; María Poveda-Villalón, Universidad Politécnica de Madrid, Spain; Emilio M. Sanfilippo, ISTC-CNR Laboratory for Applied Ontology, Italy; Walter Terkaj, STIIMA-CNR, Italy

Solicited review: three anonymous reviewers

Abstract. Industry 4.0 (I4.0) is a new era in the industrial revolution that emphasizes machine connectivity, automation, and data analytics. The I4.0 pillars such as autonomous robots, cloud computing, horizontal and vertical system integration, and the industrial internet of things have increased the performance and efficiency of production lines in the manufacturing industry. Over the past years, efforts have been made to propose semantic models to represent the manufacturing domain knowledge, one such model is Reference Generalized Ontological Model (RGOM).¹ However, its adaptability like other models is not ensured due to the lack of manufacturing data. In this paper, we aim to develop a benchmark dataset for knowledge graph generation in Industry 4.0 production lines and to show the benefits of using ontologies and semantic annotations of data to showcase how the I4.0 industry can benefit from KGs and semantic datasets. This work is the result of collaboration with the production line managers, supervisors, and engineers in the football industry to acquire realistic production line data^{2,3}. Knowledge Graphs (KGs) or Knowledge Graph (KG) have emerged as a significant technology to store the semantics of the domain entities. KGs have been used in a variety of industries, including banking, the automobile industry, oil and gas, pharmaceutical and health care, publishing, media, etc. The data is mapped and populated to the RGOM classes and relationships using an automated solution based on JenaAPI, producing an I4.0 KG. It contains more than 2.5 million axioms and about 1 million instances. This KG enables us to demonstrate the adaptability and usefulness of the RGOM. Our research helps the production line staff to take timely decisions by exploiting the information embedded in the KG. In relation to this, the RGOM adaptability is demonstrated with the help of a use case scenario to discover required information such as current temperature at a particular time, the status of the motor, tools deployed on the machine, etc.

Keywords: Industry 4.0, production line, Knowledge Graphs, Industry 4.0 Knowledge Graph

* Corresponding author. E-mail: m.yahya1@nuigalway.ie.

¹ <https://w3id.org/rgom>

² <https://github.com/MuhammadYahta/ManufacturingProductionLineDataSetGeneration-Football>

³ <https://zenodo.org/record/7779522>

1. Introduction

The Industry 4.0 (I4.0) production line is a trending research topic that necessitates the automation of the production line to meet smart manufacturing [34]. The goal of I4.0 is to use various technologies and artificial intelligence to make better use of resources and optimize production [29,36]. Meanwhile, in I4.0, the sensing technologies hosted on machines generate an immense amount of data in different types and formats. During the manufacturing process, the machines and tools required to be automatically handled according to the context of the retrieved information to enhance the efficiency of the overall production line [33].

In the I4.0 manufacturing production line., various operators require access to different types of information according to their interests and roles. For instance, an engineer from the maintenance department would be involved in knowing the *Critical To Process* (CTP) or Key Process Index Variables (KPIV) information of machines on a daily basis. The CTP or KPIV of machines includes information such as their nominal power requirement, current temperature and/or pressure, and operating frequencies of the tools. On another side, an engineer from the production department would like to analyze the product's quality and would be interested in retrieving the CTP parameters set down by the maintenance department for each product. These operators can be facilitated with such information retrieved from information systems.

In order to enable the production line operators to access such information, the semantic web is one of the possible solutions. The semantic web has transformed the current document-based web into a more intelligent system. The semantic web combines the data and content into a structured web environment, allowing software agents to execute tasks autonomously for users. The semantic web uses ontologies, RDF model and syntax, SPARQL Queries, etc. to represent the information in a machine-processable structure, in which ontologies sit in the centre among other building blocks [1]. According to Feilmayr et al. 2016 and Gruber et al. 1993, an ontology can be defined as *a formal, explicit specification of a shared conceptualization that is characterized by high semantic expressiveness* [8,14]. The assertion of the data into ontology becomes a KG. According to Hogan et. al. 2021, a KG can be thought of as a graph of data intended to gather and convey knowledge of the real world, with nodes denoting objects of interest and edges denoting relationships between these objects [15]. The semantic web recommends best practices for exposing, sharing, and integrating data, information, and knowledge, known as linked data. The data distributed under an open license is referred to as linked open data [22]. In the last decade, KGs have been widely used in many fields, e.g., geography, health care, education, news, social networks, cyber-security, and I4.0 [39].

Industry 4.0 Knowledge Graphs (I4.0 KGs) have been receiving significant attention recently, and many researchers are working to build them such as manufacturing production lines KGs. However, most of the time, they are limited to a specific use case [20]. These use cases are based on two possibilities (1) researchers are using synthetic data, or (2) a use case is coming from an industry based on their private company data. There are a few datasets available for Industry 4.0 production lines, such as Bosch dataset.⁴ The Bosch dataset reflects the production line processes including its machines and their operations, and targets applications such as predictive maintenance. This dataset does not provide any description of its attributes and therefore, it is very ambiguous and hard to understand. For example, *L3_S36_F3939* stands for a feature measured on line 3, station 36, that has a feature number 3939. However, there is no explanation about what sort of station and feature they are. Moreover, private company data is only accessible to enterprise researchers. For the earlier use case, there is an issue with reproducibility and making the data findable, accessible, interoperable, and reusable, also known as the FAIR principles. However, to the best of our knowledge, there is no open dataset (and associated KGs) for an Industry 4.0 production line.

Based on our investigations, the goals of this research work are (1) to prepare a realistic production line dataset and (2) to generate a semantic representation of the data with the help of the Reference Generalized Ontological Model (RGOM). To achieve these goals, collaboration is made with the football industry (explained in Section 3) and the results contribute to

- one of the first datasets following the FAIR principles representing the realistic data collected from the football production line,
- an automated approach for mapping the data into RGOM to build a KG,

⁴<https://www.kaggle.com/c/bosch-production-line-performance/data>

Table 1
Key facts about the dataset

Name:	Football manufacturing production line dataset
URL:	https://doi.org/10.5281/zenodo.7779522
Version number:	v1.0
Version date:	30-03-2023
Use of established vocabularies:	RGOM, Dublin Core, SSN, Time, OM, OWL RDF, RDFS.
Licensing:	MIT Licence

- a demonstration of the RGOM adaptability in the form of a catalogue of competency questions (SPARQL queries) that is provided by the collaborators.

The dataset is evaluated using the FAIR evaluation tool,⁵ obtaining a total FAIR score of 79% (Findable: 77%, Accessible: 100%, Interoperable: 75%, Reusable: 62%).

The results demonstrate the efficiency of the RGOM against the dataset that could be utilized by the researcher's community. Table 1 illustrates some of the key facts about the Football production line dataset. Additionally, this could motivate other industries to set up a linked data-based production line.

The rest of the paper is organized as: Section 2 reviews the Related work by exploring state-of-the-art techniques in this area. The dataset from the football production line is analyzed in Section 3. An approach to integrate the production line data is illustrated in Section 4. The use case to verify the RGOM adaptability with the generated dataset is presented in Section 5. Section 6 explains the extension of the benchmark to other use cases. Finally, the paper is concluded in Section 7.

2. Related work

In recent years, a number of initiatives have been undertaken to semantically represent the I4.0-based production line domain knowledge where an application is built on top of it. An overview of the existing literature on using datasets following ontologies to construct a knowledge graph is presented in this section. A summary of the ontologies and the datasets being used to develop an I4.0 production line KG is presented here.

2.1. Extant ontologies with real use cases

In recent years, domain and application ontologies have been developed in different I4.0 settings. Along with the development, approaches and use cases have been made available for reuse and validation. In 2016, Cheng et al. proposed a modular approach to semantically model a demonstration production line in which the product, process, parameter, and device ontologies were merged using a base ontology [3]. However, their paper lacks any examples or evaluation of how the ontology can be used in practice. In another study by Jarvenpaa et al. 2019, an ontology was proposed to represent both simple and combined capabilities of manufacturing resources [17]. Unlike Cheng et al. [3], the authors evaluated resource capabilities based on their ontology using laboratory-based test data. Ramírez Durán et al. 2020 [27], proposed an ontological model (ExtruOnt) to semantically describe a manufacturing machine known as an extruder machine. Specifically, their work focuses on extruder machines, including extruder components, three-dimensional representations of components, spatial connections, features, and sensors that provide information on the performance of machines. As a result, it can be used as a reference model to construct ontologies that represent other manufacturing machines in I4.0. The authors were able to evaluate their ontology with real data taken from the urola solutions. In a recent study, we developed RGOM [37] which is an ontological model including a domain core terminology spanning from raw material to manufacturing a product. The construction of RGOM has reused several classes and properties from previous studies [3,17,30] and also modeled a number of missing concepts and relations. The common limitation of the extant ontologies is that although the ontologies

⁵<https://satifyd.dans.knaw.nl/>

are developed based on domain knowledge, use cases, and datasets, the knowledge resources especially the datasets are not publicly available, making it difficult to evaluate the ontologies.

2.2. I4.0 production line with KG generation

In line with smart manufacturing, many other studies have empowered ontologies with applicable use cases and explored the advantages of KGs on top of ontologies to configure, manage, and optimize the production line. In 2018, in order to represent knowledge of manufacturing resource reconfiguration, a configuration-based ontology was developed using the Web Ontology Language (OWL) [34]. This research focuses on the integration of manufacturing Cyber-Physical Systems (CPS) equipment based on an ontology-based resource integration architecture. Ontology-based architecture is evaluated using data generated by the smart manipulators as clients and data stored on the Raspberry Pi as servers. A KG is constructed by mapping the data instances to ontology terminology. Kovalenko et al. 2018, presented AutomationML ontology to enhance the engineering processes in CPS design [23]. The ontology addresses the CAEX section of the AutomationML standard. However, the authors did not explain how their ontology can be used with data instances to generate a KG. In another study, Liebig et al. 2019 proposed a KG pipeline for industrial automation and control [24]. A KG is constructed by combining data collection, data cleaning, and the use of various technologies. There has been a revolution in KG-based approaches in academia as well as in manufacturing companies, such as Bosch, and Siemens [13,16]. Irlan et al. 2020, proposed Bosch I4.0 KGs on the top of a set of domain ontologies [13]. The purpose of their work is to integrate data from different silos in order to address the problem of interoperability. According to Kalaycı et al. 2020, a Semantic Integration at Bosch (SIB) framework was introduced to integrate Bosch manufacturing data to analyze the surface mounting process pipeline [20]. To experiment with their framework, they developed surface mounting to map production line data. In all, KGs have become a potential enabler to facilitate I4.0 production lines, as can be seen from the various initiatives and interests. However, there lacks a clear explanation of the data instances to generate KGs and the mapping mechanism among the datasets, ontologies and KGs.

To summarize, the current literature demonstrates an emerging trend towards the implementation of ontologies and KGs to I4.0 production lines. Nevertheless, many studies did not demonstrate how to populate the ontologies with data instances to construct a KG. There also lacks public access to the existing knowledge resources, particularly the datasets that constitute the ontologies. As a solution to the lack of availability of production line datasets, a collaborative effort is established with the football manufacturing industry with the goal of obtaining realistic data. In order to construct a KG, dataset instances are populated into RGOM. The development of a KG based on this methodology can serve as a motivation for other smart manufacturing industries.

3. Football production line

This section explains the data acquisition and dataset construction. The production floor consists of several production lines, each consisting of 9 machines with five operators (humans) performing manual operations. A typical football construction requires Thermoplastic polyurethane (TPU) roll, football cores, printing colours, glue, Laser Cutting Machine, Oval Printing machine, High-Frequency machine, Glue Spraying machine, Heat activating Conveyor machine, Forming Moulding machine, Ball Shaping machine, Ball Seam Gluing machine, and Heat drying machine. During a single production process, these machines perform different processes on different materials and produce four footballs as a finished product. Figure 1. depicts the single process flow of football production and the flow of the sensors data in a manufacturing production line.

The sensors installed on the machines in the I4.0-based production line generate data that are sent to the monitoring unit. In order to collect the first real instance of the data, several meetings were held with production line managers and engineers of Forward group limited regarding the operations of the machines, resources, processes, and production. It generally involved recording the power consumption, temperature, pressure, location, and type of process performed at a given timestamp by the machines. Also, the working status and rotational speed of the motor and other attributes were also recorded. Table 2 depicts the tools and machines parameters which include

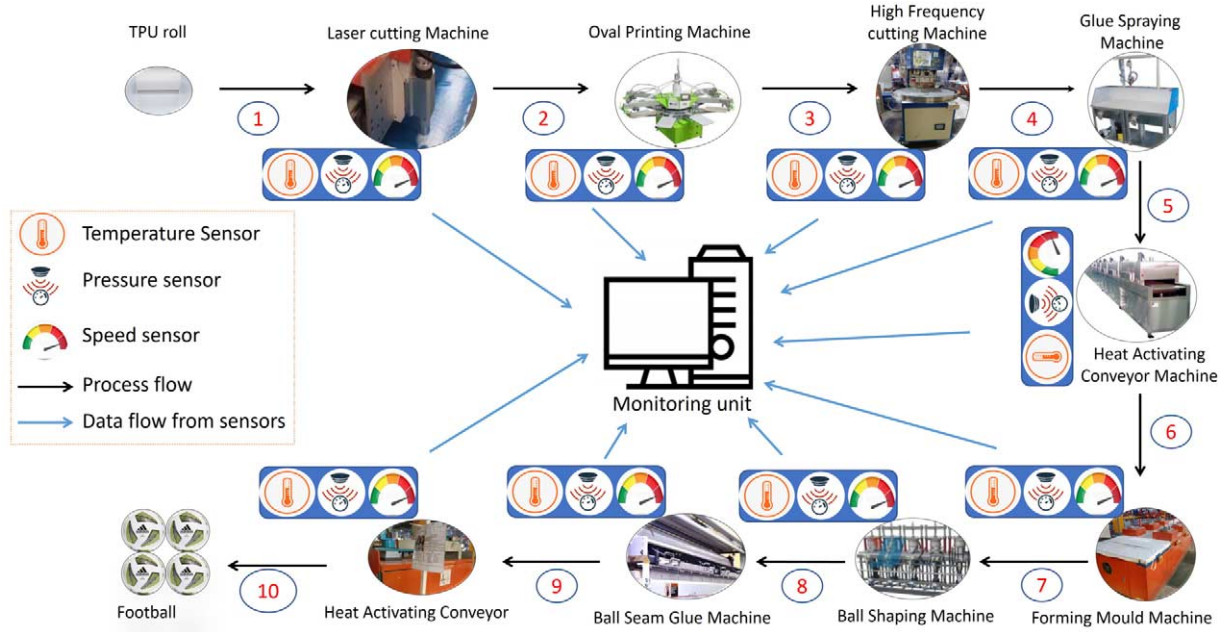


Fig. 1. Flow of a single production process. The black arrows show the process flow in the production line and the blue arrows show the data flow from the sensors to the monitoring unit. ① In a single process, the TPU roll is fed into Laser Cutting Machine. ② Laser Cutting Machine converted the TPU roll into patches. ③ Patches are printed via squeegee by the oval printing machine. ④ Printed patches are cut into panels. ⑤ Back sides of panels and cores are sprayed with glue. ⑥ Glued panels and cores are passed by the heated conveyor to form a moulding machine. ⑦ Cores and panels are moulded. ⑧ Balling shaping machine gives football shape to the moulded cores and panels. ⑨ The gaps between the panels are sealed with glue via a Ball seam glue machine. ⑩ The glue is dry by the heat-activating conveyor and 4 footballs are produced.

machine name, timestamp, temperature, pressure, power, laser die, bed, squeegee, heater, high-frequency die and many others.

Initially, the collected data were stored in a file comprising two types of attributes such as static attributes and variable attributes under the supervision of production line engineers and managers. Static attributes contain those attributes of the machine whose values remain the same in each process of manufacturing e.g. Machine model, process location, motor ID etc. On the other hand, variable attribute value changes in each process based on the performance condition of the machine, e.g. temperature, pressure, diameter, etc. Using the minimum and maximum values as well as the real value measured by the sensor, we are able to obtain the realistic data with the help of uniform probability distribution in each sub-processes. Uniform probability distribution takes input in a range bounded between the possible minimum and maximum value describing the possible likelihood and values of a variable [21]. The values of generated instances are computed using base values plus the random value generated using uniform probability distribution in the range $[a, b]$. For instance, to generate a temperature value of *machine1*, we can formally compute a new temperature value $Temp_2$ from the previous value $Temp_{previous}$ as follows:

$$Temp_2 = Temp_{previous} + f(x), \quad \text{where } f(x) = \begin{cases} 0, & \text{if } x < a \\ \frac{1}{b-a}, & \text{if } a < x < b \\ 0, & \text{if } x > b \end{cases} \quad (1)$$

Where a and b are the maximum and minimum reference values from the production line supervisors and engineers. For example, to compute an instance of the temperature for *machine1* during the TPU roll-cutting process, a base value of temperature is taken that is 41° celsius. Using the uniform probability distribution $f(x) = \text{random.uniform}(a, b)$ in equation (1), where a is set to 1 and b is set to 5, this provides an $f(x)$ of 0.25. Upon summing the $f(x)$ value with the previous temperature value, it gives the temperature value for *machine1*,

41.25° celsius. The same procedure is used for other variable attributes. The values generated are consistent with actual observations and have been verified and validated by the engineers on the production line.

After clarifying the two types of attributes regarding the entire production line, the data type for each machine is analyzed and generated. As mentioned before, the production includes 9 machines that are explained in the following subsections.

3.1. Laser Cutting Machine

Laser Cutting Machine (LCM) is a manufacturing machine that performs the first process in football production, known as the cutting TPU process. A laser-based cutting tool is hosted by the LCM to convert TPU rolls into patches. The laser is rotated with the help of a motor. LCM produces six patches in a single process as a refined material that are given as input to the Oval Printing Machine (OPM).

3.2. Oval Printing Machine

The OPM hosts six tools i.e., three beds, three squeegees, and three heaters. Each tool has a different function, for instance, the bed acts as a container to hold the patch, each squeegee prints different colours (colour 1, colour 2 and colour 3), and the heater dries the printed patches with a temperature ranging from 55 to 65-degree Celsius. Squeegee has various attributes such as power consumption, pressure, hardness, etc. The six patches produced by LCM are passed as input material to the OPM. This machine performs a total of eight step-wise processes to print the colours on the patches in a single production process. In the first process, patch one is placed on the first bed, which is then forwarded to the squeegee for printing colour one. The printed patch placed on bed one is dried in the heat of heater one. In the second process performed by OPM, the same operation is repeated by squeegee one and heater one on patch two placed on bed two. Squeegee two prints colour two on patch one and heater two dry the printed patch. Now, the patch contains two colours. In the third process by OPM, bed three contain patch three, while bed two and bed one contain patch two and patch three, respectively. At the end of the third process, beds one, two, and three contain patches one, two, and three with printed colours one, one and two, one, two and three, respectively. The rest of the eight processes are performed in the same flow to print the three colours on the 4 to 6 patches. The output of the machine is passed to the High-Frequency Cutting Machine (HFCM)

3.3. High-Frequency Cutting Machine

The dry-printed patches are transferred to the bed of an HFCM. A die-cutting tool hosted by HFCM cut off the printed patch into four panels. HFCM performs a single process for each patch, a total of 6 processes are performed, and 24 panels are produced in a single production process. An operator plucks the panels from the HFCM and matches the panels for a single football which is passed to the next machine known as Glue Spraying Machine (GSM).

3.4. Glue Spraying Machine

The GSM in the production line receives two input materials, i.e., 6 panels and a rubber material inside the football known as the core. The glue is sprayed on the backside of the 6 panels and core with the help of a needle (diameter of the needle is 0.5 millimetres) hosted on the GSM. The glue panels and core are sent to Heat Activating Conveyor (HAC) machine.

3.5. Heat Activating Conveyor machine

Heat Activating Conveyor (HAC) is a conveyor machine. It has a heating tool that generates heat with a temperature ranging from 45 to 55 degrees Celsius. The function of the HAC is to dry the glue on the backside of the core and panels. The dried core and panels are sent to a ball shaping machine.

3.6. Ball shaping machine

The panels attached to the core are provided as input to the ball shaping machine. The core and panels are placed inside the ball shaping machine, where pressure with a 60-degree Celsius temperature is applied to convert the panels on the core into a round shape. This manufacturing process results in the production of a semi-finished football. The semi-finished football is passed to the form moulding machine.

3.7. Form moulding machine

A form moulding machine is an assembly machine. It is used to assemble the panels on the core. It performs a total of 4 processes in a single production process. This machine output is provided to the ball seam glue machine as an input.

3.8. Ball seam glue machine

The ball seam glue machine performs a manufacturing process. The ball seam glue machine hosts a needle with a diameter of 0.5 millimeters, aiming to fill the gap between panels with the glue. The filled gap of the product is then sent to the heat drying conveyor machine.

3.9. Heat drying conveyor machine

This is the final machine in the football production process. The function of the heat drying conveyor machine is similar to that of machine 5. The glue (wet) football is then passed through a Heat drying conveyor to become dry. After the process of machine 9, operators clean the ball, pack it in polybags, and then in the cartoon.

4. An approach to Integrate the I4.0 production line data

The acquired data becomes increasingly available in the data storage. However, most of the time, the raw data are stored in various formats, disregarding their semantics and relations. It may restrict the usability of the data, e.g., querying information, data analysis, etc. Therefore, there is a pressing need to represent this data in a semantic representation, i.e., Linked Open Data. Semantically enriched representation of data or KGs adds meaning and context to data through ontologies and vocabularies that make it more easily understood and interpreted by humans and machines [25]. This leads to several benefits including improved data integration, data understanding, data interoperability, and faster discovery of knowledge via more powerful data querying and analysis [7,10,19]. We present the new release of the I4.0 KG dataset, comprising comprehensive semantic descriptions and values of the machine's processes involved in a single football production line.

This section describes the approach to building I4.0 KGs from the football production line data that can be queried by any staff in the production line. Figure 2 illustrates the workflow for constructing the KG which is comprised of four layers, Layer 1: Unstructured Data Sources, Layer 2: Knowledge Graph Construction, Layer 3: Football Production Line Knowledge Graph, and Layer 4: Users and Applications. The following subsection describes the layers and the interaction of the different components.

4.1. Layer 1: Unstructured data sources

The tools and sensors hosted by machines generate a huge amount of data at different timestamps in the manufacturing production line. The generated data is usually unstructured and is usually stored in different formats (e.g., csv, xml, json, text, etc.) by the data storage. Accessing the unstructured data in terms of information requires a lot of pre-processing and manual efforts. It is difficult for the production line staff to access the information from the data.

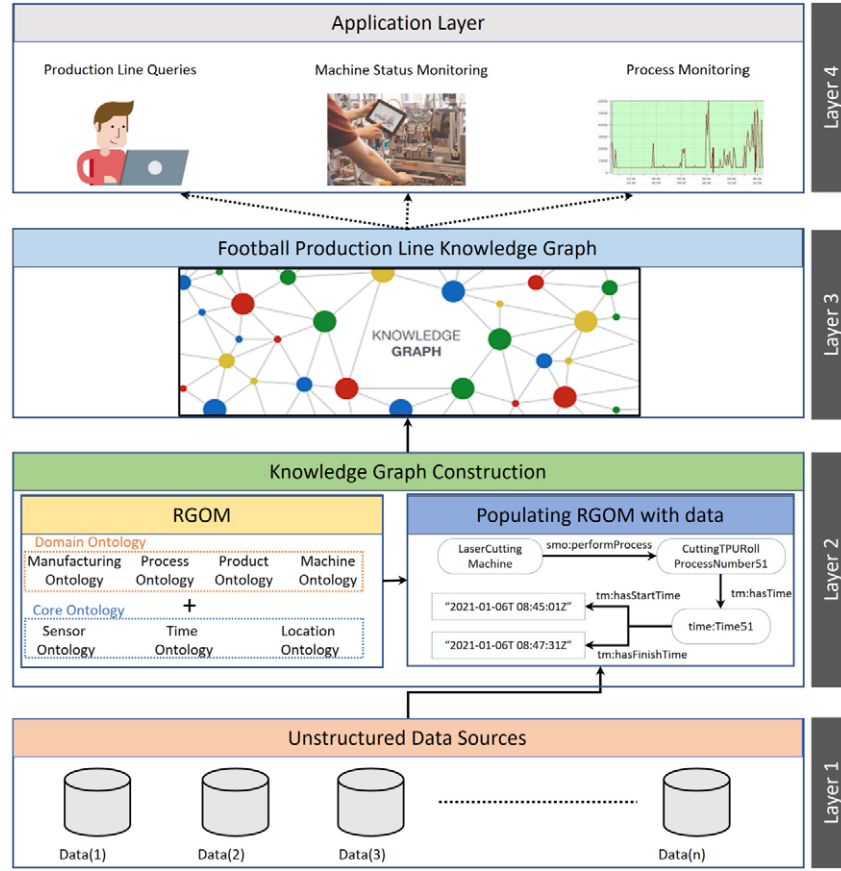


Fig. 2. An illustration of the approach to integrate data.

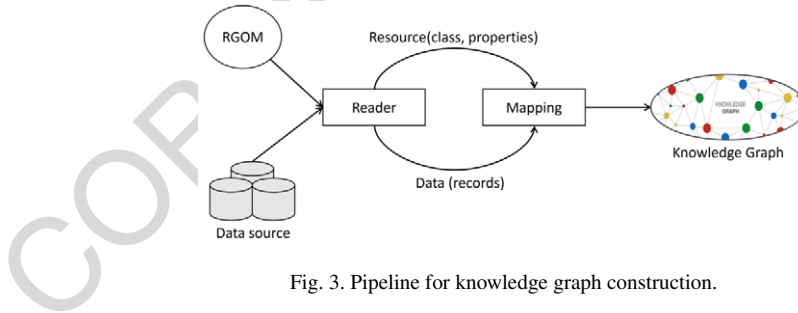


Fig. 3. Pipeline for knowledge graph construction.

4.2. Layer 2: Knowledge Graph Construction

The goal of building an I4.0 KG can be accomplished with core and domain ontologies of the Reference Generalized Ontological model (RGOM), to which the data from Layer 1 is mapped to construct a KG. Figure 3 depicts the pipeline to construct I4.0 KGs. The RGOM and data sources are given as input to the reader component. The reader component read the ontology resources i.e., class, object and data properties from the RGOM and parsed data records from the data sources. The data instances are mapped with RGOM classes and properties by the mapping components that are described in Section 4.2.2. The subsections below describe the components involved in Layer 2.

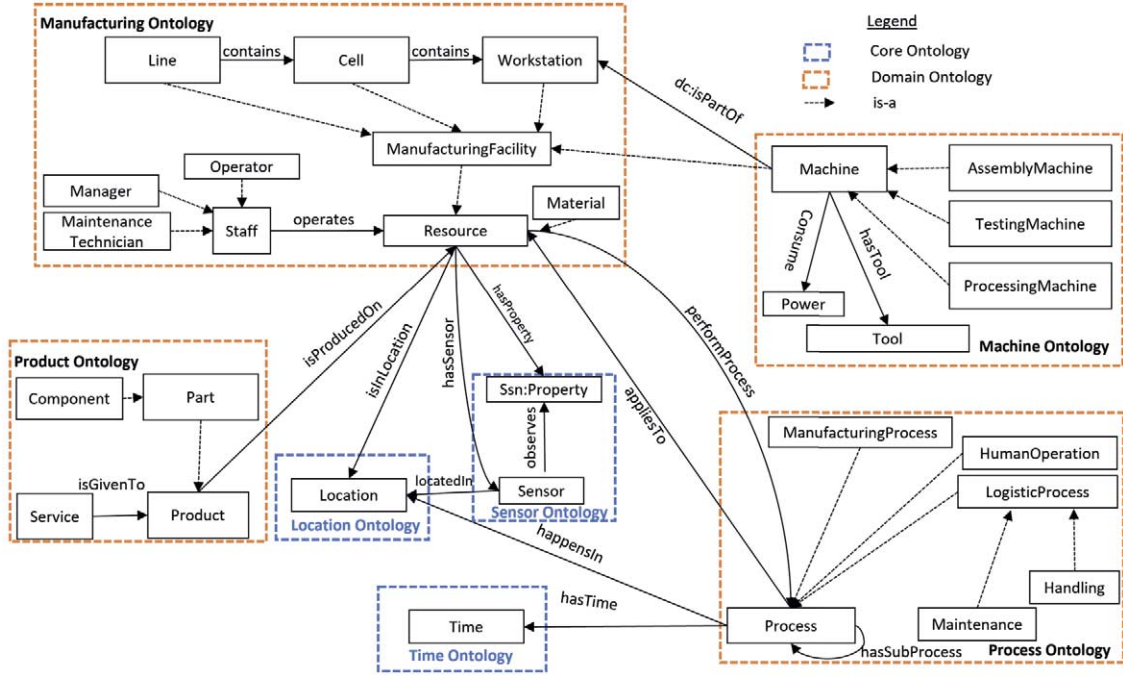


Fig. 4. Figure illustrates some of the main classes, sub classes and relations of domain and core ontologies in RGOM ontology.

Table 3
An overview of RGOM highlighting key concepts and relations, along with references to other reused ontologies

RGOM components	Reused concepts (References)	RGOM new concepts/relations
Manufacturing ontology	[3,12,17,30]	Material, Container (Pellet), isProcessedby, isPlaceOn
Machine ontology	[3,12,27,30]	MachinePart, Tool, Capabilities, Current, Power, consumesPower, hasInputMaterial, hasTool, useTool
Process ontology	[3,12]	ManualProcess, MeasurementProcess, ConveyorOperation, FeederOperation
Product ontology	[12]	Service, isGivenTo
CoreOntology	[4,12,26,28]	X

4.2.1. RGOM

The development of the football production line KG is based on our initial work that is about RGOM [37]. RGOM is a generalized ontology representing the essential concepts of I4.0, illustrated in Fig. 4. One of the main characteristics of RGOM is its modular design and is the combination of domain ontologies (device ontology, process ontology, manufacturing ontology, etc.) and core ontologies (sensor ontology, time ontology, location ontology) which is the extended version of the work by Giustozzi, F. et al. [12]. The use of core ontologies helps to benefit from the context of the data and provides added value information. Additionally, RGOM is inspired by the standards adopted by Reference Architectural Model Industry 4.0 (RAMI4.0) which is of key importance for developing the I4.0 applications.

In order to develop the KG based on RGOM, only the most relevant ontologies to the I4.0 are selected which are the core ontologies and domain ontologies. Figure 4 depicts the concepts and relations reused from relevant existing ontologies (ssn:Sensor, ssn:Property, sosa:madeObservation, sosa:observes, time:Time, time:hadTime, dc:isPartOf), while there are several newly defined concepts such as Power, Tool, Part, Service, etc. Due to the page limit of the paper, we could not show the full picture of the RGOM in detail, however, Table 3 provides the overview of RGOM with emphasis on key RGOM concepts/relations as well as other reused ontologies and their references. Moreover, the modules in the RGOM ontology are discussed as follows.

Manufacturing ontology The manufacturing ontology objective is to semantically describe the resources in the manufacturing production line. The *Staff* concept represents all the people participating in the production activities, i.e., technicians, operators, and managers. The *ManufacturingFacility* concepts characterize different physical entities and hardware modules in the factory. The concepts of the production line are decomposed into; *Workstations* contains the physically integrated machines; *Cell* is the combination of the workstation to perform a complex task; a *line* includes cells. This decomposition presents a potential reconfigurable processing line. Additionally, this taxonomy makes it possible to describe the manufacturing facility context at various next levels such as the characterization of a line or to illustrate the context of the cell that belongs to that line. The physical entities in the manufacturing facility are the resources linked to other ontologies via object properties.

Machine ontology The machine is the main resource to process raw or refined material into semi or finished production on the production line. The machine performs the process with the help of tools by itself or with intervention from the human. It can be either a processing or assembly machine processing a raw or refined material or assembling the refine parts. The machine is a manufacturing facility and is part of the workstation.

Process ontology A set of tasks or operations completed by a resource is known as Process(es). The process(es) performed by a resource(s) can be controlled operations as well as machining or assembly ones. The process ontology represents the fundamental taxonomy of all the processes executed in the manufacturing and is specified with contextual such as process *happensIn* location, process *appliesTo* a product, machine performProcess process, etc.

Product ontology Product ontology covers the basic taxonomy related to products based on the IEC 62890. The components are the parts assembled by an assembling machine into a finished product. The customers can place an order of one of the two, i.e., for service to the bought product or buying a new product.

Core ontologies Establishing the context of certain manufacturing tasks is difficult and is a major problem in the industrial domain since it includes many different entities linked to time and locations. *Time*, *Location*, *Process*, *Machine* and *Resource* are the primary concepts to semantically represent the manufacturing knowledge in line with domain ontologies. In addition, the terminology reused from the sensor ontology i.e. SSN enhanced the semantic representation of the collected sensor data. The concepts for measuring the sensor data are reused from the Ontology of units of Measure (OM) [28]. Thus, the use of core ontologies e.g, *Time* [26], *Location* (adopted from [12]) and *Sensor* [4] with the domain ontologies present useful information regarding various situations, i.e., inquiring about the status of the motor at a particular time (Listing 3).

4.2.2. Populating RGOM with data

Now we discuss populating RGOM with Data which adds the structured data from the production line to the ontological terms of RGOM, i.e., classes and properties. Algorithm 1 outlines the data population process, starting with initializing the ontology classes and properties required for the data population. It then goes through each record in the data, the values in the columns and creates individuals and their associated object/data properties. Using the semantics defined in the ontology model, an object property is chosen to form a triple with the subject and object individuals. For example, as can be seen from Fig. 2 in Layer 2 of the data population into ontology terms, the algorithm gets the class type *ManufacturingMachine* and *ManufacturingProcess* from the ontology file, and iterates over the data records in the data file. In a single iteration, the columns *machine1* and *machine1_process* from the spreadsheets are created as a subject and object with the aforementioned class types, respectively. Based on the condition, the parsed literal values are linked via data property to the resources. For instance, object property, such as *performsProcess*, from the ontology file which links the resources to form a triple i.e., *data:LaserCuttingMachine smo:performsProcess data:CuttingTPURollProcessNumber51*. Moreover, the values in the database column can also be mapped to the attributes of the KG via column mapping using R2RML [5].

As a result, data will be successfully populated to RGOM. In Fig. 5, we show a snippet of the *machine_9* individuals along with their features i.e *processMaterial*, *hasTools*, *consumesPower*, *performsProcess*, *isInLocation*, *hosTools*.

Algorithm 1: Populating data to rgom

Input: ontologymodel, data
Output: KG

```

1 namespace ← IRI ;
2 resource ← ontologymodel.getResource(namespaces + "classtype"); // e.g., Machine, Process etc.
3 r1 ← Null ;
4 r2 ← Null ;
5 ⋮
6 rnth ← Null ;
7 objectProperty = ontologymodel.getProperty(namespace + "objectproperty"); // e.g., performProcess etc.
8 dataProperty = ontologymodel.getProperty(namespace + "dataproperty"); // e.g., hasName, hasTime etc.
9 for each record in data do
10   if (record.column{1} != Null && record.column{1} != (""))
11   then
12     r1 ← ontologymodel.createIndividual(namespace+record.column{1}, resource);
13     r1.addProperty(dataProperty, record.column{l}); //literal column
14   if (record.column{2} != Null && record.column{2} != (""))
15   then
16     r2 ← ontologymodel.createIndividual(namespace+record.column{2}, resource);
17     r2.addProperty(dataProperty, record.column{l}); //literal column
18     r2.addProperty(objectProperty, r1);
19   ⋮
20   if (record.column{ith} != Null && record.column{ith} != (""))
21   then
22     rnth ← ontologymodel.createIndividual(namespace+record.column{ith}, resource);
23     rnth.addProperty(dataProperty, record.column{l}); //literal column
24     rnth.addProperty(objectProperty, r2);
25 KG = ontologymodel.write(path); //store the KG on the provided path of the local machine

```

4.3. Layer 3: Knowledge Graphs

Layer 3 represents the KG generated from layer 2. It is often used to store interlinked descriptions of entities – objects, events, situations, or abstract concepts. Whilst the data is mapped into the ontology concepts and properties, it becomes a KG⁶. In our case, the I4.0 KG contains the football production line data that the engineers use to analyze the machines and process CTP parameters on a daily basis.

4.4. Layer 4: Users and application layer

At this layer, different users and applications can access the connected data via the KG. The KG offers a SPARQL endpoint to take queries from the users and applications and return the results to them, i.e., what tools are hosted by machines, monitoring the status of a machine motor, etc. SPARQL⁷ is a standard query language used to access data from KG based on RDF and OWL. The users and applications benefit from the semantics provided by the KG in a standardized format.

⁶<https://web.stanford.edu/class/cs520/>

⁷<https://www.w3.org/TR/sparql11-query/>

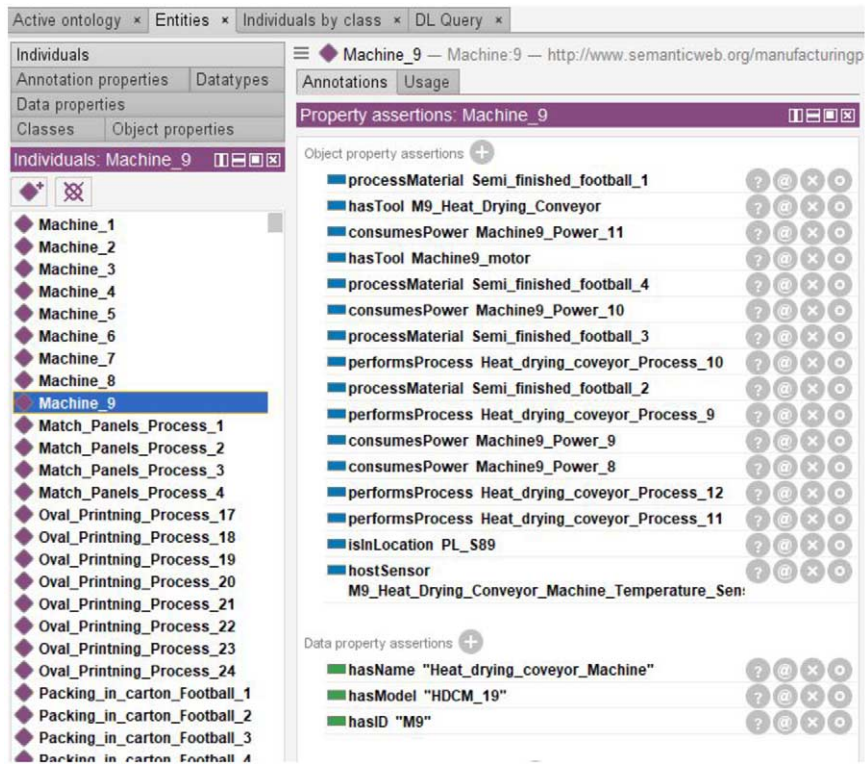


Fig. 5. Illustration of machine 9 instance with their attributes instances.

5. Use case: RGOM adaptability

In this section, a use case on the KG-based dataset is presented. The use case deals with the adaptability of RGOM on realistic football production line data. The RGOM semantically integrates the heterogeneous data of the machines in a football production line into I4.0 KG⁸. The data from the I4.0 KG is queried without requiring extra time and manual effort.

The production process of a football contains nine machines and each performs different sub-processes (Fig. 1). The data produced during this process contains the domain knowledge of the machines, tools hosted on the machines, processes performed on them, tools deployed on the machines, tool's critical parameters, and contextual data generated by the sensors hosted on the machines at some timestamp. The approach presented in Section 4 is followed to semantically integrate the data. At first, we gathered the data sources containing the data about all the machines which are then analyzed in line with the RGOM classes and relations. The main classes and relations of RGOM are depicted in Fig. 3. Next, the data is populated to the ontology terms, i.e., classes and relations with Jena API⁹. Upon the population of data into ontology terms, an RDF triple store is obtained, known as an I4.0 KG.

To produce a single football, an average of 1730 triples, 1355 logical triples, and 233 declaration triples are produced. In one hour of the production line, a total of 9 main processes are executed, producing 36 footballs and 22150 triples on average from 2903 individuals.

Besides, three I4.0 KG-based datasets are generated to provide the researcher's community to evaluate their tools and techniques. These datasets are comprised of ten days, twenty days, and thirty days of data from a football production line. The types of machines and their parameters are explained in Section 3. The total number of axioms,

⁸<https://drive.google.com/drive/folders/15G4tgVheu-gOHg8Ia4VKwF2UgXcZFWzn?usp=sharing>

⁹<https://jena.apache.org>

Table 4
Summary of the axioms in each Knowledge graph

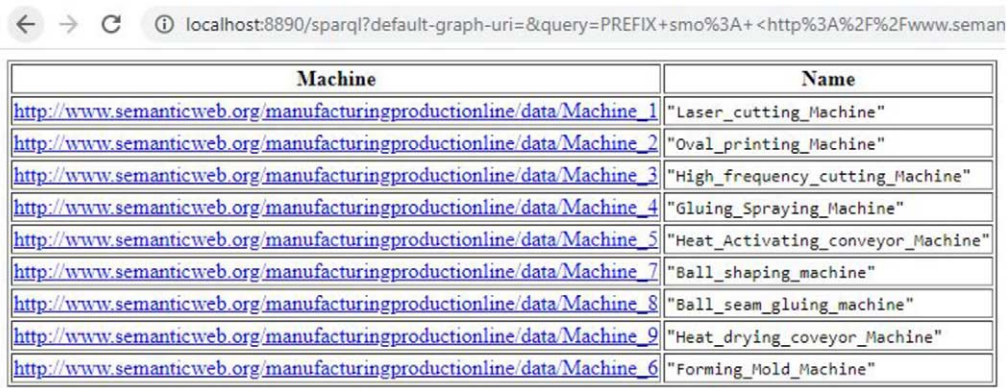
KGs	Total number of Axiom	Logical Axiom Count	Declaration axioms Count	Individual Count
10 Days	525865	525503	225	166273
20 Days	1050535	1050173	225	332363
30 Days	1470280	1469918	225	465238

```

PREFIX smo: <http://www.semanticweb.org/manufacturingproductionline/>
SELECT *
WHERE { { ?Machine a smo:ProcessingMachine; } UNION
        { ?Machine a smo:AssemblingMachine; }
        ?Machine smo:hasName ?Name. }

```

Listing 1. Query to retrieve the machines involved in the production line



Machine	Name
http://www.semanticweb.org/manufacturingproductionline/data/Machine_1	"Laser_cutting_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_2	"Oval_printing_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_3	"High_frequency_cutting_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_4	"Gluing_Spraying_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_5	"Heat_Activating_conveyor_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_7	"Ball_shaping_machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_8	"Ball_seam_gluing_machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_9	"Heat_drying_coveyor_Machine"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	"Forming_Mold_Machine"

Fig. 6. Listing 1 query provides the number of machines involved in the production line with their names.

```

PREFIX smo: <http://www.semanticweb.org/manufacturingproductionline/>
PREFIX d: <http://www.semanticweb.org/manufacturingproductionline/data/>
SELECT ?machine ?tool
WHERE { d:Machine_2 smo:hasName ?machine;
        smo:hasTool ?tool. }

```

Listing 2. Query to retrieve the tools hosted on machine 2

logical axioms count, declaration axioms count, and individual count in each KG are illustrated in Table 4. The number of classes, object properties, and datatype properties are the same for each KG.

After the construction of the KG, several queries are provided by the domain experts to find the usefulness of the KG. The SPARQL endpoint at the application layer of I4.0 KGs paves the way for different applications to access the required information embedded in the KG. Given a production line where the job at hand is to utilize the query drawn from Listing 1 in order to access the type of machines and their names involved.

Figure 6 shows the results returned from the Listing 1 query. It can be seen from the figure that the production line consists of a single assembly line and eight processing machines, each with their name.

Furthermore, the production line manager can utilize the Listing 2 query to find the tools present on a machine. The result of the Listing 2 query is shown in Fig. 7. It can be seen from the figure that machine 2 has a name and uses different types of tools such as one motor, three beds, three heaters, and three squeegees. The query fetches machine 2 name and different types of tools that resides in the KG.

← → ↻ ⓘ localhost:8890/sparql?default-graph-uri=&query=PREFIX smo:<http%3A%2F%2Fwww.semanticweb.org%2F

machine	tool
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Bed_1
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Bed_2
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Bed_3
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Heater_1
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Heater_2
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_Heater_3
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/Machine2_motor2
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/machine2_Squeegee_1
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/machine2_Squeegee_2
"Oval_printing_Machine"	http://www.semanticweb.org/manufacturingproductionline/data/machine2_Squeegee_3

Fig. 7. Listing 2 query provides tools hosted by machine 2.

```

PREFIX smo:<http://www.semanticweb.org/manufacturingproductionline/>
PREFIX d:<http://www.semanticweb.org/manufacturingproductionline/data/>
PREFIX tm:<http://www.w3.org/2006/time#>
SELECT DISTINCT ?Motor_Name ?Status ?Start_time
WHERE {
  d:Machine_1 smo:hasTool ?motor.
  ?motor smo:hasName ?Motor_Name.
  ?motor smo:hasMotorState ?state.
  ?process tm:hasTime ?time.
  ?state smo:hasState ?Status.
  ?time tm:hasStartTime ?Start_time.
  FILTER (?Start_time > "2021-06-01T 10:11:00Z"^^xsd:dateTime &&
    ?Start_time < "2021-06-01T 10:12:55Z"^^xsd:dateTime).
}

```

Listing 3. Query to retrieve the status of machine 2 motor at certain time period

← → ↻ ⓘ localhost:8890/sparql?default-grap

Motor_Name	Status	Start_time
"LCM_Motor"	"working"	"2021-06-01T 10:12:28Z"
"LCM_Motor"	"working"	"2021-06-01T 10:11:32Z"
"LCM_Motor"	"working"	"2021-06-01T 10:11:57Z"
"LCM_Motor"	"working"	"2021-06-01T 10:12:13Z"

Fig. 8. Result returned by the query in Listing 3.

Similarly, an engineer from the maintenance department wants to query the KG for CTP parameters to check the current observation of the sensor or the status of the motor. For instance, a maintenance engineer can retrieve the status of a motor at a particular period of time by using the query in Listing 3. The query fetches the status of the motor at different timestamps as illustrated in Fig. 8.

In order to retrieve the temperature (a CTP parameter) query in Listing 4 is utilized. The reuse principle of Linked Open Data has been followed by reusing the *SOSA* vocabulary as depicted in list 4. Figure 9 shows the fetched results of the query in Listing 4.

Furthermore, a query in Listing 5 is used to find the total number of processes performed by a machine and the total number of tools that each machine used during a given time. The query returns information about all the machines with the total number of processes they performed during a given time and the total number of tools used

```

PREFIX smo: <http://www.semanticweb.org/manufacturingproductionline/>
PREFIX d: <http://www.semanticweb.org/manufacturingproductionline/data/>
PREFIX tm: <http://www.w3.org/2006/time#>
PREFIX sosa: <http://www.w3.org/ns/sosa#>
SELECT DISTINCT ?machine ?Start_time ?result
WHERE {
  ?machine smo:hasTool ?tool.
  ?tool sosa:madeObservation ?observation.
  ?observation sosa:hasSimpleResult ?result.
  ?time tm:hasStartTime ?Start_time.
  FILTER (?tool != d:M6_Folding_Mold_machine_Pressure_Sensor.)
}

```

Listing 4. Query to retrieve the CTP parameter (Temperature) with time

← → ↺ ① localhost:8890/sparql?default-graph-uri=&query=PREFIX+smo%3A+<http%3A%2F%2Fwww.sema

machine	Start_time	result
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T12:30:31Z	"34 celcius"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T12:31:01Z	"34 celcius"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T12:36:59Z	"34 celcius"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T12:37:29Z	"34 celcius"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T12:37:59Z	"34 celcius"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T12:38:29Z	"35 celcius"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T12:44:27Z	"34 celcius"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T12:44:57Z	"35 celcius"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T12:45:27Z	"34 celcius"
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	2021-06-12T12:45:57Z	"35 celcius"

Fig. 9. Listing 4 query provides the retrieval of the temperature (a CTP parameter).

```

PREFIX smo: <http://www.semanticweb.org/manufacturingproductionline/>
PREFIX d: <http://www.semanticweb.org/manufacturingproductionline/data/>
SELECT DISTINCT ?machine (count(distinct ?process) as ?process_count)
(count(distinct ?tool) as ?tool_count)
WHERE {
  ?machine smo:performsProcess ?process.
  ?machine smo:hasTool ?tool.
  ?process sosa:hasTime ?time.
  ?time tm:hasStartTime ?start_time.
  ?time tm:hasFinishTime ?finish_time.
  FILTER (?start_time > "2021-06-08T 12:55:13Z"^^xsd:dateTime &&
  ?finish_time < "2021-06-12T 14:36:04Z"^^xsd:dateTime
} GROUP BY ?machine order by ?machine

```

Listing 5. Query to retrieve the count of processes performed by machines and count of tools used by them during a time period

by them, shown in Fig. 10. For instance, in the list, machine_1 has performed a total of 100 processes and used a total of three tools during the time from 12:55:13 to 14:36:04.

6. Extension of the benchmark to other use cases

The current work showcases the potential of applying the RGOM to a larger picture. The concepts introduced in RGOM can be utilised for the majority of production lines which will facilitate wider use of our approach for generating KGs for different use cases. For example, this approach can be extended not only to other similar manu-



machine	process_count	tool_count
http://www.semanticweb.org/manufacturingproductionline/data/Machine_1	110	3
http://www.semanticweb.org/manufacturingproductionline/data/Machine_2	1054	10
http://www.semanticweb.org/manufacturingproductionline/data/Machine_3	792	2
http://www.semanticweb.org/manufacturingproductionline/data/Machine_5	134	3
http://www.semanticweb.org/manufacturingproductionline/data/Machine_6	536	3
http://www.semanticweb.org/manufacturingproductionline/data/Machine_7	538	3
http://www.semanticweb.org/manufacturingproductionline/data/Machine_8	539	3
http://www.semanticweb.org/manufacturingproductionline/data/Machine_9	539	3

Fig. 10. Listing 5 query provides the count of processes performed by machines and count of tools used by them during a time period.

facturing processes such as volleyball and rugby ball production, but also to other more generic production lines that incorporate welding processes. This can help other industries map their customised data into RGOM and construct an industry-specific KG. Additionally, to build KGs for a different manufacturing industry, one should adopt the RGOM framework to add definitions of required classes and relations. For instance, in ongoing work, we have been collaborating with Bosch to incorporate RGOM in their welding production line. We have created new concepts such as *WeldingMachine* as *Machine*, *WeldingRobot*, *WeldingGun*, and *Electrode* as *Tool*, *WeldingOperation* and *MaintenanceOperation* as *Operation*, and *Workpiece* as *Material*. This developed welding ontology demonstrates the adaptability of RGOM [38]. Adopting a similar mechanism will help digital transformation for those who have not set up a Linked Data-based production line.

The I4.0 KG dataset can be utilised in predictive maintenance [2]. For example, one useful use case of the I4.0 KG dataset is predicting the temperature of a machine. In manufacturing factories, the temperature of the machine is of high significance and critical. During the process execution, the tools are operating under a set point. The increase in temperature can adversely affect the machine which impacts the product quality. The assessment of the temperature information enables the setting up of condition-based machine tool temperature monitoring and prevents any impact on the quality of the end product. Furthermore, the I4.0 KG can be used by deep learning models to carry out entity matching, nodes classifications, link prediction, and knowledge graph completion [18,31,35].

7. Conclusion

The purpose of this research is to introduce a benchmark dataset with KG generation for I4.0 production lines, and specifically leverage the advantages of ontologies and semantic annotation of the data to showcase how the I4.0 industry can benefit from KGs and semantic datasets. We identified that the current research work has been limited to two use case possibilities, (1) using synthetic data and (2) an Industry-based use case that comes with private company data. private company data is made accessible only to people associated with the enterprise. Moreover, the earlier use case faces an issue of reproducibility and FAIR principles. This work tries to bridge the gap between academic researchers and industry by providing the former (first case) access to data that is not easily and publicly available. To address this limitation, a dataset is designed by acquiring realistic data from a football production line. An automated solution is developed that maps the dataset attributes with RGOM terms, i.e., classes and relations in the form of I4.0 KGs. The results of the competency questions on the I4.0 KG illustrate the adaptability of the RGOM in a production line. The adaptability of RGOM demonstrates the usefulness of the dataset, Hence, it can be extended to similar production lines such as Volleyball and Rugby ball.

Further studies need to be carried out in order to validate the dataset on other ontological models. Further research should be undertaken to apply KG embedding and diffusion techniques to have a complete I4.0 KG.

Funding

This publication has emanated from research supported in part by a grant from Science Foundation Ireland (SFI) under Grant Number SFI/16/RC/3918 (Confirm), by a grant from SFI under Grant Number SFI 12/RC/2289_P2 (Insight), and by a grant from the European Union under Grant Number 958371 (OntoCommons). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

- [1] Z. Ahmed and D. Gerhard, Role of ontology in semantic web development, 2010, arXiv preprint [arXiv:1008.1723](https://arxiv.org/abs/1008.1723).
- [2] Q. Cao, A. Samet, C. Zanni-Merk, F.d.B. de Beuvron and C. Reich, An ontology-based approach for failure classification in predictive maintenance using fuzzy C-means and SWRL rules, *Procedia Computer Science* **159** (2019), 630–639. doi:[10.1016/j.procs.2019.09.218](https://doi.org/10.1016/j.procs.2019.09.218).
- [3] H. Cheng, P. Zeng, L. Xue, Z. Shi, P. Wang and H. Yu, Manufacturing ontology development based on Industry 4.0 demonstration production line, in: *2016 Third International Conference on Trustworthy Systems and Their Applications (TSA)*, IEEE, 2016, pp. 42–47. doi:[10.1109/TSA.2016.17](https://doi.org/10.1109/TSA.2016.17).
- [4] M. Compton, P. Barnaghi, L. Bermudez, R. Garcia-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog et al., The SSN ontology of the W3C semantic sensor network incubator group, *Journal of Web Semantics* **17** (2012), 25–32. doi:[10.1016/j.websem.2012.05.003](https://doi.org/10.1016/j.websem.2012.05.003).
- [5] S. Das, R2RML: RDB to RDF mapping language, 2011, <http://www.w3.org/TR/r2rml/>.
- [6] L. Ehrlinger and W. Wöb, Towards a definition of knowledge graphs, *Semantics (Posters, Demos, SuCCESS)* **48**(1–4) (2016), 2.
- [7] S. El-Sappagh, J.M. Alonso, F. Ali, A. Ali, J.-H. Jang and K.-S. Kwak, An ontology-based interpretable fuzzy decision support system for diabetes diagnosis, *IEEE Access* **6** (2018), 37371–37394. doi:[10.1109/ACCESS.2018.2852004](https://doi.org/10.1109/ACCESS.2018.2852004).
- [8] C. Feilmayr and W. Wöb, An analysis of ontologies and their success factors for application to business, *Data & Knowledge Engineering* **101** (2016), 1–23. doi:[10.1016/j.datak.2015.11.003](https://doi.org/10.1016/j.datak.2015.11.003).
- [9] D. Fensel, U. Simsek, K. Angele, E. Huaman, E. Kärle, O. Panasiuk, I. Toma, J. Umbrich and A. Wahler, *Knowledge Graphs*, Springer, 2020.
- [10] N. Freire and S. de Valk, Automated interpretability of linked data ontologies: An evaluation within the cultural heritage domain, in: *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, 2019, pp. 3072–3079. doi:[10.1109/BigData47090.2019.9005491](https://doi.org/10.1109/BigData47090.2019.9005491).
- [11] J.H. Gennari, M.A. Musen, R.W. Fergerson, W.E. Grosso, M. Crubézy, H. Eriksson, N.F. Noy and S.W. Tu, The evolution of protégé: An environment for knowledge-based systems development, *International Journal of Human-computer studies* **58**(1) (2003), 89–123. doi:[10.1016/S1071-5819\(02\)00127-1](https://doi.org/10.1016/S1071-5819(02)00127-1).
- [12] F. Giustozzi, J. Saunier and C. Zanni-Merk, Context modeling for industry 4.0: An ontology-based proposal, *Procedia Computer Science* **126** (2018), 675–684. doi:[10.1016/j.procs.2018.08.001](https://doi.org/10.1016/j.procs.2018.08.001).
- [13] I. Grangel-González, F. Lösch and A. ul Mehdi, Knowledge graphs for efficient integration and access of manufacturing data, in: *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Vol. 1, IEEE, 2020, pp. 93–100. doi:[10.1109/ETFA46521.2020.9212156](https://doi.org/10.1109/ETFA46521.2020.9212156).
- [14] T.R. Gruber, A translation approach to portable ontology specifications, *Knowledge acquisition* **5**(2) (1993), 199–220. doi:[10.1006/knac.1993.1008](https://doi.org/10.1006/knac.1993.1008).
- [15] A. Hogan, E. Blomqvist, M. Cochez, C. d’Amato, G. de Melo, C. Gutierrez, S. Kirrane, J.E.L. Gayo, R. Navigli, S. Neumaier et al., Knowledge graphs, *Synthesis Lectures on Data, Semantics, and Knowledge* **12**(2) (2021), 1–257. doi:[10.1007/978-3-031-01918-0](https://doi.org/10.1007/978-3-031-01918-0).
- [16] T. Hubauer, S. Lamparter, P. Haase and D.M. Herzig, Use cases of the industrial knowledge graph at siemens, in: *International Semantic Web Conference (P&D/Industry/BlueSky)*, 2018.
- [17] E. Järvenpää, N. Siltala, O. Hylli and M. Lanz, The development of an ontology for describing the capabilities of manufacturing resources, *Journal of Intelligent Manufacturing* **30**(2) (2019), 959–978. doi:[10.1007/s10845-018-1427-6](https://doi.org/10.1007/s10845-018-1427-6).
- [18] S. Ji, S. Pan, E. Cambria, P. Martinen and S.Y. Philip, A survey on knowledge graphs: Representation, acquisition, and applications, *IEEE Transactions on Neural Networks and Learning Systems* **33**(2) (2021), 494–514. doi:[10.1109/TNNLS.2021.3070843](https://doi.org/10.1109/TNNLS.2021.3070843).
- [19] H. Jung, H. Yoo and K. Chung, Associative context mining for ontology-driven hidden knowledge discovery, *Cluster Computing* **19** (2016), 2261–2271. doi:[10.1007/s10586-016-0672-8](https://doi.org/10.1007/s10586-016-0672-8).
- [20] E.G. Kalaycı, I.G. González, F. Lösch, G. Xiao, E. Kharlamov, D. Calvanese et al., Semantic integration of bosch manufacturing data using virtual knowledge graphs, in: *International Semantic Web Conference*, Springer, 2020, pp. 464–481.
- [21] S. Kamthe, S. Assefa and M. Deisenroth, Copula flows for synthetic data generation, 2021, arXiv preprint [arXiv:2101.00598](https://arxiv.org/abs/2101.00598).
- [22] S. Khusro, F. Jabeen, S.R. Mashwani and I. Alam, Linked open data: Towards the realization of semantic web-a review, *Indian Journal of Science and Technology* **7**(6) (2014), 745. doi:[10.17485/ijst/2014/v7i6.7](https://doi.org/10.17485/ijst/2014/v7i6.7).
- [23] O. Kovalenko, I. Grangel-González, M. Sabou, A. Lüder, S. Biffl, S. Auer and M.-E. Vidal, *AutomationML Ontology: Modeling Cyber-Physical Systems for Industry 4.0*, IOS Press, 2018.
- [24] T. Liebig, A. Maisenbacher, M. Opitz, J.R. Seyler, G. Sudra and J. Wissmann, Building a knowledge graph for products and solutions in the automation industry, 2019.

- [25] F. Özcan, C. Lei, A. Quamar and V. Efthymiou, Semantic enrichment of data for AI applications, in: *Proceedings of the Fifth Workshop on Data Management for End-To-End Machine Learning*, 2021, pp. 1–7.
- [26] F. Pan and J.R. Hobbs, Time Ontology in OWL, 2006, W3C working draft, W3C.
- [27] V.J. Ramírez-Durán, I. Berges and A. Illarramendi, ExtruOnt: An ontology for describing a type of manufacturing machine for Industry 4.0 systems, *Semantic Web* **11**(6) (2020), 887–909. doi:[10.3233/SW-200376](https://doi.org/10.3233/SW-200376).
- [28] H. Rijgersberg, M. Van Assem and J. Top, Ontology of units of measure and related concepts, *Semantic Web* **4**(1) (2013), 3–13. doi:[10.3233/SW-2012-0069](https://doi.org/10.3233/SW-2012-0069).
- [29] C. Santos, A. Mehraei, A. Barros, M. Araújo and E. Ares, Towards Industry 4.0: An overview of European strategic roadmaps, *Procedia Manufacturing* **13** (2017), 972–979. doi:[10.1016/j.promfg.2017.09.093](https://doi.org/10.1016/j.promfg.2017.09.093).
- [30] A. Seyedamir, B.R. Ferrer and J.L.M. Lastra, An ISA-95 based ontology for manufacturing systems knowledge description extended with semantic rules, in: *2018 IEEE 16th International Conference on Industrial Informatics (INDIN)*, IEEE, 2018, pp. 374–380. doi:[10.1109/INDIN.2018.8471929](https://doi.org/10.1109/INDIN.2018.8471929).
- [31] B. Steenwinckel, G. Vandewiele, M. Weyns, T. Agozzino, F.D. Turck and F. Ongenaes, INK: Knowledge graph embeddings for node classification, *Data Mining and Knowledge Discovery* **36**(2) (2022), 620–667. doi:[10.1007/s10618-021-00806-z](https://doi.org/10.1007/s10618-021-00806-z).
- [32] Y. Tang, T. Liu, G. Liu, J. Li, R. Dai and C. Yuan, Enhancement of power equipment management using knowledge graph, in: *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, IEEE, 2019, pp. 905–910. doi:[10.1109/ISGT-Asia.2019.8881348](https://doi.org/10.1109/ISGT-Asia.2019.8881348).
- [33] D. Thakker, P. Patel, M.I. Ali and T. Shah, in: *Semantic Web of Things for Industry 4.0, Semantic Web*, 2020, pp. 1–2.
- [34] J. Wan, B. Yin, D. Li, A. Celesti, F. Tao and Q. Hua, An ontology-based resource reconfiguration method for manufacturing cyber-physical systems, *IEEE/ASME Transactions on Mechatronics* **23**(6) (2018), 2537–2546. doi:[10.1109/TMECH.2018.2814784](https://doi.org/10.1109/TMECH.2018.2814784).
- [35] M. Wang, L. Qiu and X. Wang, A survey on knowledge graph embeddings for link prediction, *Symmetry* **13**(3) (2021), 485. doi:[10.3390/sym13030485](https://doi.org/10.3390/sym13030485).
- [36] S. Wang, J. Wan, D. Li and C. Zhang, Implementing smart factory of industrie 4.0: An outlook, *International journal of distributed sensor networks* **12**(1) (2016), 3159805. doi:[10.1155/2016/3159805](https://doi.org/10.1155/2016/3159805).
- [37] M. Yahya, J.G. Breslin and M.I. Ali, Semantic Web and Knowledge Graphs for Industry 4.0, *Applied Sciences* **11**(11) (2021), 5110. doi:[10.3390/app11115110](https://doi.org/10.3390/app11115110).
- [38] M. Yahya, B. Zhou, Z. Zheng, D. Zhou, J.G. Breslin, M.I. Ali and E. Kharlamov, *Towards Generalized Welding Ontology in Line with ISO and Knowledge Graph Construction*, ESWC (Posters & Demos), Springer, 2022.
- [39] X. Zou, *A Survey on Application of Knowledge Graph*, Journal of Physics: Conference Series, Vol. 1487, IOP Publishing, 2020, p. 012016.