

A Knowledge Graph for Industry 4.0

Sebastian R. Bader^{1,3}[0000–0003–1328–704X], Irlan Grangel-Gonzalez², Priyanka Nanjappa³, Maria-Esther Vidal⁴, and Maria Maleshkova³[0000–0003–3458–4748]

¹ Fraunhofer IAIS, Schloss Birlinghoven, 53757 Sankt Augustin, Germany

² Robert Bosch GmbH, Robert-Bosch-Campus 1, 71272 Renningen, Germany

³ University of Bonn, Endenicher Allee 19a, 53115 Bonn, Germany
sebastian.bader@iais.fraunhofer.de, Irlan.GrangelGonzalez@de.bosch.com,
priyanka.nanjappa@uni-bonn.de, maria.vidal@tib.eu,
maleshkova@cs.uni-bonn.de

Abstract. One of the most crucial tasks for today’s knowledge workers is to get and retain a thorough overview on the latest state of the art. Especially in dynamic and evolving domains, the amount of relevant sources is constantly increasing, updating and overruling previous methods and approaches. For instance, the digital transformation of manufacturing systems, called Industry 4.0, currently faces an overwhelming amount of standardization efforts and reference initiatives, resulting in a sophisticated information environment. We propose a structured dataset in the form of a semantically annotated knowledge graph for Industry 4.0 related standards, norms and reference frameworks. The graph provides a Linked Data-conform collection of annotated, classified reference guidelines supporting newcomers and experts alike in understanding how to implement Industry 4.0 systems. We illustrate the suitability of the graph for various use cases, its already existing applications, present the maintenance process and evaluate its quality.

Keywords: industry 4.0 · knowledge graph · standards · knowledge representation.

1 Introduction

Industrial processes are driven by norms and standards. While other domains and communities rely on common agreements and best practices, the specific reliability and safety requirements of industrial manufacturing demand strict and formal specifications. International institutions such as ISO, IEC, or ETSI together with national organizations such as NIST, DIN, or ANSI face this demand and form a network of highly recognised authorities, ensuring the quality of published standards and norms.

The rising popularity of digitizing processes, components, and complete production lines has consequently led to an increasing number of standards targeting the various related aspects. The so-called Industry 4.0 (I40) has drawn significant attention not only inside the manufacturing companies but also in academia and government. The result is an already overwhelming but further growing amount

Table 1. Resource overview

Resource type	RDF-based Knowledge graph
Location	https://github.com/i40-Tools/I40KG
Namespace	https://w3id.org/i40/sto#
Topic	Standards, norms and frameworks for Industry 4.0
License	Creative Common License 3

of relevant norms, standards, and specifications. The necessary effort for both domain experts and newcomers is also increased by the lack of suitable guidance and limited meta data. The interested reader can only evaluate the significance of a specific publication after examining the complete text – a substantial challenge regarding the amount of available specifications. Therefore, we identify a rising need for a structuring approach to better organize the relevant entities and to explicitly outline their interlinks and attributes.

We propose a publicly available knowledge graph containing the latest state of I40 specifications with respect to standards, reference frameworks as well as key requirements (cf. Tab. 1). The inter-linked nature of the content and its various relations to outside topics led to the design of an RDF-based knowledge graph for I40 standards and reference frameworks. Utilizing the information content of the proposed knowledge graph, the following types of relevant information can be retrieved:

1. Where can additional information about a certain topic be found?
2. Which specification is most appropriate for establishing a secure data exchange between Industry 4.0 devices?
3. What are the requirements related to a specific Industry 4.0 challenge and where can appropriate guidance to solving them be found?

A key feature of this work is the provisioning of relations to external data sources. Openly available information, for instance from DBpedia, enhances the understanding and points the user to further data sources in the Linked Open Data Cloud. The thereby accessible content makes the knowledge graph relevant for several potential consumer groups: *System architects* are interested in finding and learning about suitable design patterns, *I40 experts* working in standardization groups need to be aware of and observe related initiatives, *component developers* require best practices for interfaces and models, *system integrators* need to understand common data models and interaction patterns, *machine manufacturers* need to ensure the sustainability of their digital interfaces, and *I40 newcomers* want to reduce their onboarding time.

We contribute to the outlined challenges with the following: (1) present the Industry 4.0 Knowledge Graph (I40KG), (2) present its maintenance and curation processes, and (3) discuss its applicability as the basis for other resources and applications. The I40KG helps to overcome hindrances related to realizing the Industry 4.0 vision, which prerequisites not only comprehensive knowledge about distinct standards but needs to consider the semantics and relations between standards, standardization framework as well as their requirements.

The remainder of this paper is structured as follows: Section 2 gives an overview on the evolution of the resource and comparable approaches in the literature. The following section explains the I40KG principles, and how it is provisioned (Section 4). Section 5 presents intended use cases and evaluates the I40KG. We conclude the paper and outline future work in Section 6.

2 Application Domains and Impact of the Resource

This section explains the background of the proposed Industry 4.0 knowledge graph, portrays its development and compares it to similar approaches from the community.

2.1 State of the Art

The targeted challenge – to support newcomers, domain experts and any other stakeholder to establish and curate a proper overview on the published standards, frameworks, and concerns is one of the key obstacles hindering the wider adoption and successful fulfilment of the potential of I40 ideas. The hereby presented work extends previous efforts on creating an overall ontology for Industry 4.0 standards. Grangel-González *et al.* [7] introduced a first ontology for Industry 4.0 components, in particular for the Asset Administration Shell model. Extending this work, the basic structure and scheme of the graph has been developed, together with a first approach to structure the Industry 4.0-related standards and norms in terms of a unified landscape [6]. These publications introduced the initial definitions of the standard and standardization framework concepts. Further progress has been presented by Bader *et al.* [2], enhancing the graph with Industry 4.0 reference frameworks and new application patterns of Web-based visualization services and interactive views.

The I40KG is the first structured approach applying machine-readable data interlinking the textual, normative and informative resources containing the knowledge of I4.0 standardization. In comparison to the earlier evolution steps, the hereby presented I40KG has been significantly extended in terms of contained entities, from less than 80 as presented by Grangel-González *et al.* to more than 300 described instances. Furthermore, a vast number of Industry 4.0 affecting requirements has been introduced and implemented in order to allow use case-driven filtering and context-dependent discovery of relevant entities.

The I40KG constitutes a machine-readable resource of interrelated standards, reference frameworks, and concerns. It thereby comprises an extendable representation of the whole topic. In contrast to the more common format of literature reviews, the I40KG is a semantically enriched and openly accessible resource, which represents the state of the domain at its publication date and beyond. To the best of our knowledge, no comparable knowledge graph or similar resource is currently available.

The principles of Linked Data, especially of openly accessible data through established Web technologies, are at the core of the Semantic Web Community.

The proposed knowledge graph utilizes these practices and connects previously independent information sources with the Linked Open Data Cloud, in particular DBpedia. Thereby, the Semantic Web Community can use the knowledge graph to structure and extend the various related works in the context of I40. However, the targeted users of the I40KG are not limited to the Semantic Web community. As the major trend of digitization affects any domain, but in particular currently the manufacturing industry, multiple further communities can benefit from the proposed work as the insights gained in I40 radiate for instance into Smart Cities, new mobility solutions, Smart Homes and many more.

As the I40KG follows the principles for provisioning Linked Data, it also may serve as a way to spread semantic technologies to other communities. The recommendations and guidelines as for instance formulated by Noy *et al.* have been followed to ensure the quality of the graph [12]. The target groups are usually not too familiar with the Semantic Web in general and RDF-based knowledge graphs in particular, therefore the adaption of the I40KG can further support the dissemination of the mature practices of the Semantic Web and Linked Open Data.

2.2 Related Work

Overview works comparable to the one proposed in this paper usually appear in one of two forms. On the one hand, experts with an academic background collect relevant publications and comprise them in literature reviews. On the other hand, industry experts and consortia publish their views on the domain through reference frameworks and white papers. Both approaches require extensive efforts for the interested reader to discover, filter, and understand the provided content. Furthermore, the provided knowledge is only valid for a limited time around the publication date. Updates in terms of extensions and adjustments to recent developments are not common practice. Especially in the research community, updating survey papers – to reflect developments since the original publication – usually does not happen.

Still, a significant number of reviews on Industry 4.0 and the very much related IoT emerges each year. For instance, Xu *et al.* present a comprehensive overview on the major drivers and also standardization activities [16], mentioning the key developments and concerns. Martinez *et al.* outline the relations of Industry 4.0 with cyber-physical systems and (Industrial) IoT [14]. However, as typical for academic reviews, references to technical standards are omitted. This habit does not reflect the actual relevance of standards and norms for the engineering and implementation processes.

Searching for technical information in the internet is mainly executed through the established search engines. Even though more and more search queries can be answered directly returning related information, for instance by displaying Wikipedia abstracts, in general only collections of web sites are provided. The user then has to manually discover and examine the sources. Especially for technical information needs, this approach is highly inefficient as it is time-consuming and requires considerable prior knowledge. Lafia, Turner and Kuhn [10] show

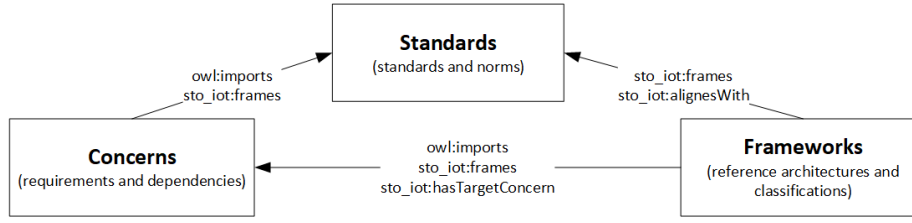


Fig. 1. The three partitions of the I40KG. The I40KG is designed in interconnected parts representing the I40 domain: Standards, Concerns, and Frameworks.

how semantic annotations and mappings on open data improves the discovery process. Nevertheless, the search for targeted, domain-specific information as regarded in this work, presents a significant burden.

Several works address the challenge of structuring the landscapes of industrial standards. For instance, Lu *et al.* [11] describe a landscape of Smart Manufacturing Systems. Similarly, Andreev *et al.* [1] provide several visual comparisons of radio connectivity standards and technologies. However, none of these surveys are published in an accessible data set as the contributions and insights are only represented written text and cannot be processed by further tools and applications.

3 Design and Technical Quality

The I40KG design follows best practices of publishing resources as Linked Data. As stated in Tab. 2, the resource conforms to the FAIR principles and is created, curated and accessible in an transparent and open manner. The required characteristics are listed in brackets using the notation of Wilkinson *et al.* [15]. The graph also reuses common RDF vocabularies wherever possible. Upper level ontologies, such as DUL or DCTERMS, support the understanding of classes and properties. Relations to DBpedia resources help to identify the intended entity but also provide valuable directions for further lookups.

3.1 Ontology Description

In this section, we present the relevant parts that form the I40KG. The I40KG is designed in a modular way in order to ensure the maintainability of the sources and increase the readability for the users. As recommended by Parent and Spaccapietra [13], each partition focuses on one of the mentioned sub-domains – *Standards*, *Concerns*, and *Reference Frameworks* (cf. Fig. 1), published in respective Turtle files. The partitions themselves depend on each other utilizing *owl:imports* statements.

The original standards ontology has been extended but still serves as the foundation for the other modules. It is focused on the description of a *standard* as a logical concept, defines attributes and relations, and contains all standard

Table 2. I40KG details. Relevant aspects of the I40KG and related resources.

General	Name	Industry 4.0 Knowledge Graph (I40KG)
	DL Expressivity	SHOIF(D)
	Licence (R1.1)	Creative Commons 3
	Size	44 classes, 35 object properties, 22 data properties, 1335 individuals
	Standards and Norms (R1.2)	338 standards and standard parts, 49 ISO standards, 67 IEC standards, 11 DIN standards
	Frameworks	18 reference frameworks divided into 138 classification sections
	Concerns	160 interrelated Industry 4.0 concerns in 6 categories
	External Links (F3, I3)	286 to DBpedia resources, 271 to Wikipedia pages
	Reasoning	4.257 derived triples
	Total size	16.447 unique triples without derived ones
Reuse	Reused Ontologies (I2, R1)	DCTERMS, DCELEMS, PROV, DUL, FOAF, OM, etc.
	Reused ODPs	Componency ODP
Documentation	Element description (F2, R1)	By means of <code>rdfs:label</code> , <code>rdfs:comment</code> , <code>skos:prefLabel</code> and <code>rdfs:isDefinedBy</code>
	Ontology Documentation	http://i40.semantic-interoperability.org/sto/
Conventions	Naming pattern	CamelCase notation for the schema and Ada for instances
	Linked Data (R1.3)	5 Star Linked Data
Multilinguality	English labels for all terms	<code>rdfs:label</code> and <code>rdfs:comment</code> with the @en notation
Availability	Persistent URI (F1)	https://w3id.org/i40/sto
	Serialisations (I1)	Turtle, RDF/XML
	GitHub (A1)	https://github.com/i40-Tools/I40KG/
	LOV (F4)	http://lov.okfn.org/dataset/lov/vocabs/sto
	OntoPortal (A2)	http://iofportal.ncor.buffalo.edu/ontologies/STO
	Licence	Creative Commons 3.0
	VoCol Instance (A2)	http://vocol.iais.fraunhofer.de/sto/

instances (cf. Fig. 2). Concerns, as defined in ISO 42010 [9], can be understood as domain requirements, motives or issues, which a stakeholder can have about an IT system in general and – in the context of this paper – an Industry 4.0 setting. To increase readability, we further use the terms ‘concern’ and ‘requirement’ synonymously, even though the definitions in ISO 42010 slightly differ.

While ISO 42010 defines the terminology of a concern itself, it lacks an approach to supply a set of usable instances. The I40KG therefore contains a taxonomy for I40-related concerns, which is intended as a first outline undergoing further refinements. Starting with six top-level concerns (*Data Sovereignty*, *Internet of Things*, *Trustworthiness*, *Data Analytics*, *Interoperability*, *Business Context*), cycle-free dependencies of sub-concerns are formed. Further details about the concerns themselves have been presented also by Bader *et al.* [2].

3.2 I40KG Example Instances

Figure 3 shows a set of I40KG instances. The *IEC 62714* about AutomationML has various links (*sto:uses*, *sto:isComponentOf*, *sto:relatedTo*) to other standards. In addition, annotations (green, values yellow) explain the entity itself, containing among others the official location of the source document. For IEC standards,

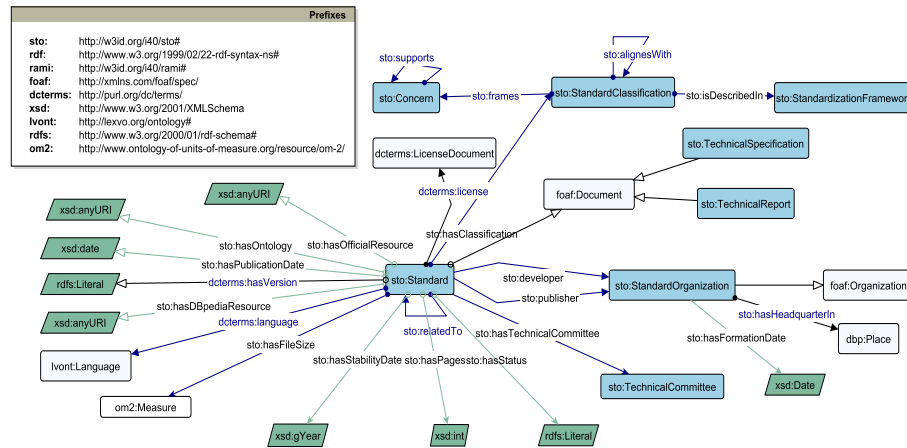


Fig. 2. Core classes and properties of the Standards Module. I40KG-specific classes (light blue), imported properties (blue) and classes (white) from FOAF, DC-TERMS, and RAMI4.0 ontologies.

this is usually the IEC webstore site of the respective standard. More relations to external resources are also supplied, mainly to Wikipedia/DBpedia.

As depicted in Figure 3, IEC 62714 is classified as relevant for the *RAMI Control Device*, a *Standard Classification* scheme related to the *RAMI4.0 Standardization Framework*. A user can traverse these links and discover another *Standard Classification* instance of RAMI4.0 frames *Trustworthiness*, the *Concern* also presented in Figure 4. In this way, further information can be accessed and the user is able to further explore the I40KG.

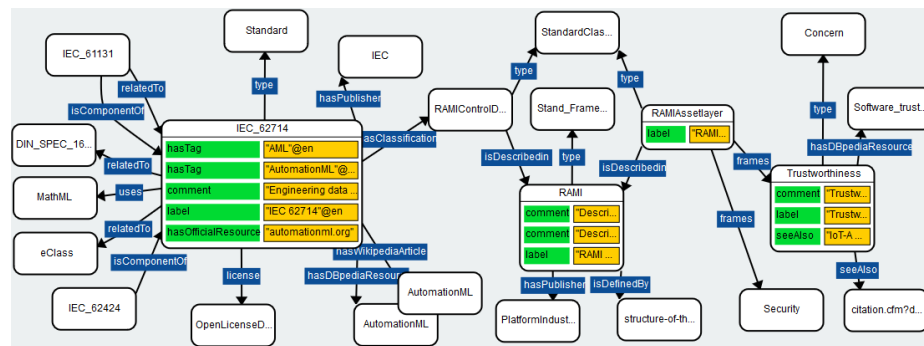


Fig. 3. Contained entities: Standards (IEC 62714) link to standard classifications (RAMI Asset Layer) with frameworks (RAMI) and requirements (Trustworthiness).

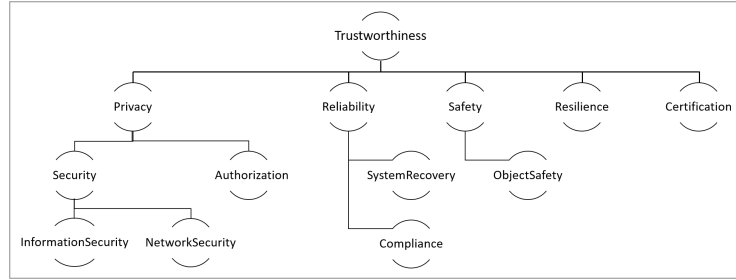


Fig. 4. Concern hierarchy. Illustration for the “Trustworthiness” of a system and the underlying concern taxonomy.

3.3 Updating Process

The knowledge graph is maintained following three different insertion processes. As depicted in Figure 5, one process for the selection, examination and annotation for standards (top) and reference frameworks (bottom) have been established. Details about the selection criteria have already been explained [6] [3] and are therefore omitted here. Both approaches are transparently executed using the GitHub repository and its commit history.

In addition to the manual extensions, an automated update process has been introduced (cf. Fig. 5). As the frequency of new standards and updates of already published ones is too high, a bot searches for such events, maps the metadata to RDF, filters relevant standards and norms, and proposes the resulting entities for insertion into the I40KG. Currently, only IEC standards are monitored but a further generalization is intended. The automated proposals require a manual approval, usually together with additional annotations to external resources, for instance to DBpedia resources.

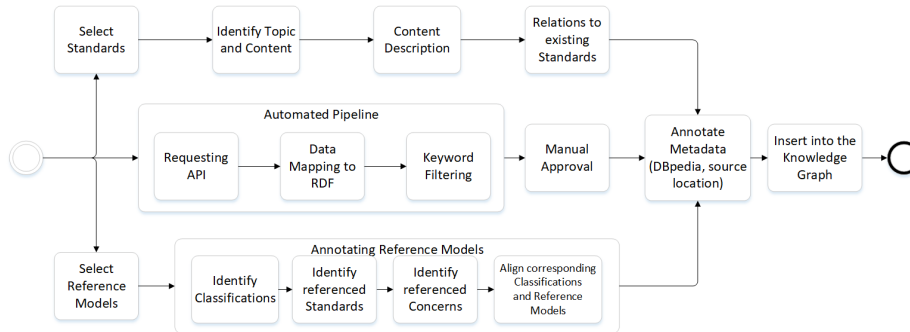


Fig. 5. Insertion Process: Three different sub-processes to create the I40KG content.

4 Availability of I40KG

The I40KG is documented following the established best practices for ontologies and Linked Data resources. We supply a human readable documentation page for all classes, properties and instances⁴. Furthermore, several serializations, e.g., RDF/XML, Turtle, N-Triples, etc. are provided, where the Turtle-files act as the single source of truth. Redirects and content negotiation is in place to supply each client with the most appropriate serialisation.

The I40KG and its entities are defined in the STO namespace, using W3IDs for long-term accessibility. STO was the original acronym for “standards ontology” and is retained for sustainability reasons. The knowledge graph is available under the **Creative Commons 3.0 license** and can be reused by anyone and for any purpose. Extensions to the original graph in terms of A- and T-Box are possible but require approval of the graph creators in order to ensure the consistency and quality of the content. Change requests can be placed at its official location, a publicly available **GitHub repository** (cf. Tab. 1).

The maintenance and further development of the knowledge graph is organized in the mentioned GitHub repository, in particular through GitHub issues. The issue system is also the main communication channel in order to propose changes, document errors and outline extensions. The complete sources are accessible and all changes and updates are executed in a publicly visible and transparent manner. Following best practices of the Semantic Web, each entity is annotated with well-known annotation properties, i.e., `rdfs:label`, `rdfs:comment` and is linked to DBpedia resources, wherever a suitable entry exists.

5 Reusability of the Graph Content

The described knowledge graph is used in several projects. In the context of the International Data Spaces (IDS)⁵, it is used in its data model but also as a reference resource for the I40 domain in general and the most up-to-date reference frameworks and architectures. We use knowledge graph embeddings on top of the I40KG to automatically exploiting the meaning of the relationships between standards⁶. We then employed unsupervised Machine Learning methods, e.g., Clustering, to unveil existing relations of standards in the I40KG. A visualisation tool has been developed in order to support and outline the use of the provided information content⁷. The various preconfigured views allow the interactive selection and comparison of I40KG’s entities. The website provides a hierarchical overview of the contained standards, a timeline, network views visualizing the various inter-relations and a comparison tools utilizing Venn diagrams and co-occurrence matrices. Figure 6 and Figure 7 show the capabilities of this tool. Furthermore, a public SPARQL endpoint⁸ provides the latest ver-

⁴ <http://i40.semantic-interoperability.org/sto/>

⁵ <https://www.internationaldataspaces.org/>

⁶ <https://github.com/i40-Tools/I40KG-Embeddings/>

⁷ <https://i40-tools.github.io/StandardOntologyVisualization/>

⁸ <https://dydra.com/mtasnim/stoviz/>

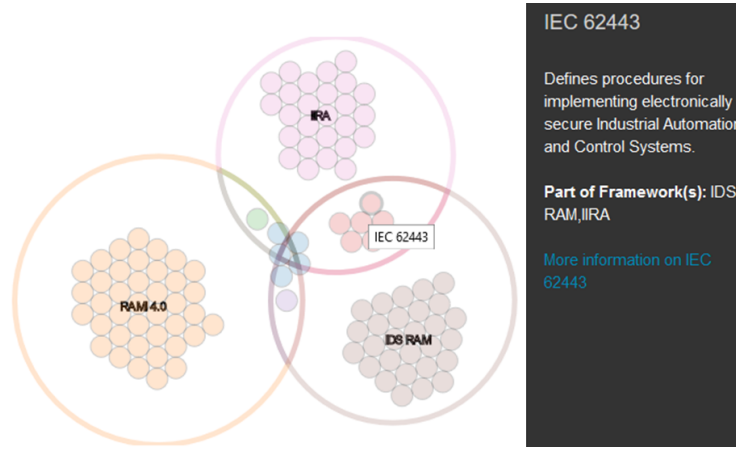


Fig.6. Venn diagrams for reference frameworks and standards. The Venn diagrams localise the standards (e.g. IEC 62443) in regard to the reference frameworks, for instance to recognize the overlaps but also uniquely covered areas.

sion of I40KG, also hosted at a VoCol instance⁹ [8] for additional documentation purposes.

All generally available RDF tools can work with the I40KG and its source files. Its core classes are, wherever suitable, linked to upper level ontologies. In particular, the linking to commonly-known DBpedia resources allows its direct integration with other knowledge graphs and especially the Linked Open Data Cloud. However, the I40KG does not intend to fully cover the domain, nor represent or judge the internal quality of the referred standards, norms and frameworks. It is – and always has to be – in the responsibility of the user to finally decide on the suitability of a certain standard or norm regarding the specific context or use case. The I40KG can support the user to effectively gain an overview and discover unknown resources. While we constantly extend and update the graph, a perfect coverage is neither possible nor intended. Nevertheless, a sufficient completeness of the domain is necessary and has been examined by Bader *et al.* [2]. The presented selection criteria show how academic and industry impact have been examined to optimally discover and filter the I40KG entities.

Nevertheless, a comprehensive overview with as much content as possible is desirable. The supplied content must comply to best practices and meet the expectations of potential users in order to provide value. We therefore evaluated the knowledge graph by two approaches. Section 5.1 explains potential use cases, shows which tasks can be solved and how the I40KG is capable of supporting the target groups. Section 5.2 describes the executed tests and quality metrics.

⁹ <https://vocol.iais.fraunhofer.de/sto/>

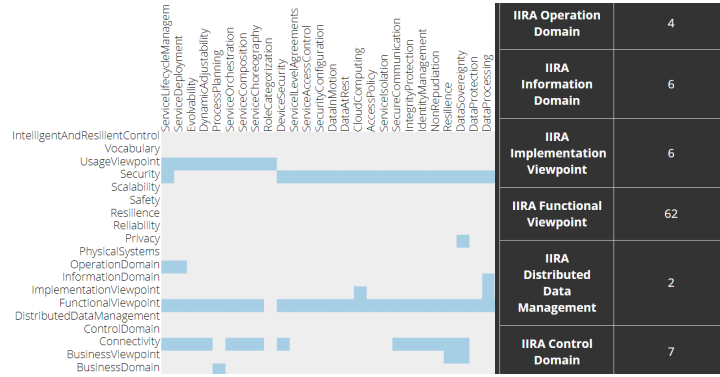


Fig. 7. Co-occurrence matrix between concerns and classification categories. The co-occurrence matrix enables insights which concerns are targeted by which classification categories of the presented frameworks.

5.1 User Stories

The outlined information content is without comparison regarding its relations to the Linked Open Data Cloud resources and the amount of described technical standards and architectural propositions. The knowledge graph can be used in various ways. We further give adoption examples by describing several user stories. Alice, Bob and Charlie represent typical users, each with a different background and information need in the context of Industry 4.0.

Alice, who is just starting with Industry 4.0 applications, needs to quickly gain an overview of the most influential reference frameworks. She has to communicate with consultants, suppliers and developers using the correct terminology and concepts in order to effectively manage the project resources. Alice looks through the hierarchy view of the mentioned web service, learning which frameworks contain which categories and standards. She gains a quick overview of which standards are the most prevalent in almost all the frameworks. She follows the relations between the classifications and traverses the links to standards and other publications but also to new reference frameworks. This process gives her a general understanding of the structure of the domain, the relevant technical standards and the their relations. Alice also executes unsupervised Machine Learning algorithms on top of the I40KG. The output of those algorithms provide knowledge about non existing relations of standards that can be used to improve the classification that the frameworks provides w.r.t standards. This also enables the enrichment of the current landscape of Industry 4.0 standards. Thus, enhancing the understanding Alice of this complex domain.

Another user of the I40KG is Bob, an industry expert working in a standardization council. He is aware of all the details of the group's works and ideas, and knows which arguments led to the proposed solution of this council. For further iterations of their guidelines, Bob would like to know about the focus and state

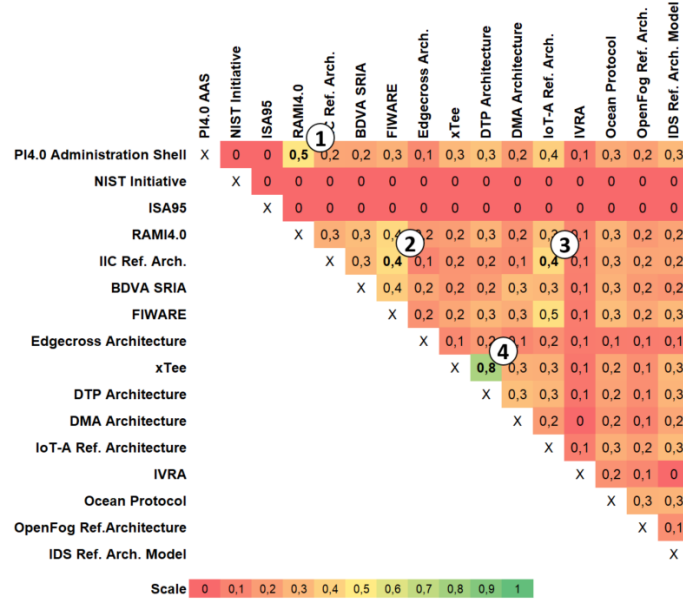


Fig. 8. Overlap of reference frameworks. Symmetric matrix displaying similar frameworks based on the amount of targeted Industry 4.0 concerns.

of complementary but also competing approaches. Furthermore, Bob searches for good ideas for his own standardization work.

Bob uses the I40KG to create the analysis shown in Figure 8. A quick look at the results tells him (cf. Fig. 8 (1)), that for instance the concepts defined in the Plattform Industrie 4.0 Asset Administration Shell model are closely related to the Reference Architecture Model Industry 4.0 (RAMI4.0). This quite obvious discovery is due to the fact that both models are published by the same organization, which Bob quickly recognizes by following the relations between the two entities in the knowledge graph. In addition, Bob also identifies a significant overlap between the Reference Architecture of the Industrial Internet Consortium (IIC) and the FIWARE platform specification¹⁰ and IoT-A Reference Architecture [4] (cf. Fig. 8 (2) and (3)). He is already familiar with the work of the IIC, therefore he decides to also examine the publications from FIWARE and IoT-A, as they might provide further suitable insights.

Charlie, a senior system architect, is aware of the concerns and requirements that his customer will face in his next project. With the aim to ensure the data security and protection of his customer’s data, he searches for best practices for implementing upcoming technologies. The co-occurrence matrix of the already mentioned web service depicts which reference frameworks and which respective classifications frame Charlie’s concerns.

¹⁰ <https://www.fiware.org/>

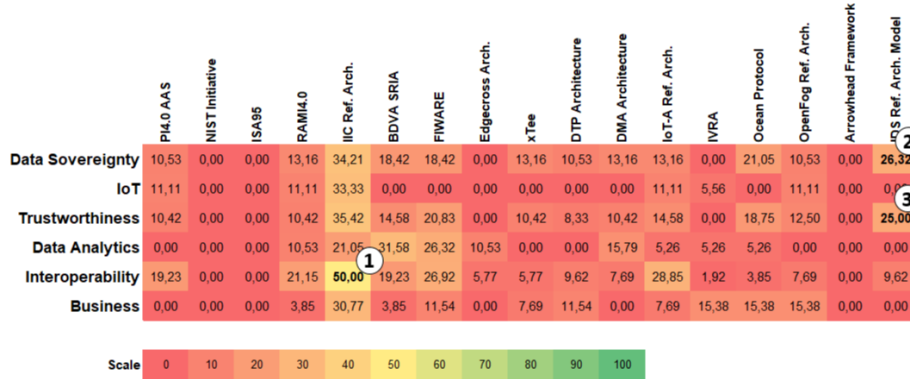


Fig. 9. Focus comparison. Calculated total coverage of Industry 4.0 requirements by reference frameworks. Higher scores do not indicate higher quality but broader coverage of a topic.

Furthermore, he uses the concern hierarchy to aggregate the information of the I40KG (cf. Fig. 9). With this query, Charlie is able to see that the IIC Reference Architecture surpasses the others in terms of its interoperability references (cf. Fig. 9 (1)). However, as data protection is his major target, the IDS Reference Architecture Model seems like a valuable information source (cf. Fig. 9 (2) and (3)).

5.2 Technical Evaluation

The syntactic quality has been checked by commonly used tools such as the Ontology Pitfall Checker¹¹ and RDF-TripleChecker¹². These tools indicate that the I40KG is consistent and correct in terms of common RDF and ontology pitfalls. Wherever the mentioned tools indicated potential for improvement, the respective sections have undergone an intense manual evaluation. The reports are also hosted in the GitHub repository.

The reports, for instance, mention two issues. Several properties miss domain and/or range attributes and sometimes the disjointness of classes is not sufficiently declared. However, it has been explicitly decided to not set the range and domain to all properties, as their implications for reasoning on the I40KG can easily result in inconsistencies. Complete disjointness statements, on the other hand, are rather uncommon, adding only limited added value to the graph itself but requiring extensive maintenance.

Furthermore, we evaluated the quality of the I40KG by using metrics as proposed by Färber *et al.* [5]. Table 3 contains all metrics grouped by the categories from Färber *et al.* in order to provide as much information as possible.

¹¹ <http://oops.linkeddata.es/>

¹² <http://graphite.ecs.soton.ac.uk/checker/>

Table 3. I40KG evaluation results.

Metric	Result	Explanation
Accuracy		
Synt. validity of RDF doc.	$m_{synRDF}(I40KG) = 1$	RDF documents are syntactically valid.
Synt. validity of literals	$m_{synLit}(I40KG) = 1$	Literals conform to their datatype.
Semant. validity of triples	$m_{sem}(I40KG) = 1$	No gold standard available. References to original information sources applied.
Trustworthiness		
KG level	$m_{graph}(I40KG) \geq 0.5$	Manual data curation but also automated process in place.
Statement level	$m_{fact}(I40KG) = 0.5$	Provenance information provided on resource level.
Unknown/empty values	$m_{NoVal}(I40KG) = 0$	Unknown values are not indicated
Consistency		
Schema restr. at insertion	$m_{checkRestr}(I40KG) = 1$	Schema restrictions are (partly) checked.
Class constraints	$m_{conClass}(I40KG) = 1$	Empty set of class constraints.
Relation constraints	$m_{conRelat}(I40KG) = 1$	Domain and range are consistent.
Relevancy		
Ranking of statements	$m_{Ranking}(I40KG) = 0$	Ranking of statements is not feasible.
Completeness		
	–	No gold standard available
Timeliness		
Frequency of the KG	$m_{Freq}(I40KG) = 0.5$	Discrete periodic updates, also through the automated pipeline.
Validity period of stmts.	$m_{Validity}(I40KG) = 0$	Provisioning of validity statements is not intended.
Modification date of stmts.	$m_{Change}(I40KG) = 0$	Modification dates are only supplied on knowledge graph level.
Ease of understanding		
Description of resources	$m_{Descr}(I40KG) = 1$	All resources have a label and comment.
Labels in multiple lang.	$m_{Lang}(I40KG) = 0$	Only some resources have multi-language annotations.
RDF serialization	$m_{uSer}(I40KG) = 1$	Serializations in Turtle and RDF/XML
Self-describing URIs	$m_{uURI}(I40KG) = 1$	Self-describing URIs are always used.
Interoperability		
Blank nodes & RDF reification	$m_{Reif}(I40KG) = 1$	No blank nodes or RDF reification.
Serialization formats	$m_{iSerial}(I40KG) = 1$	RDF/XML and Turtle are supplied when dereferencing URIs.
Using external vocabulary	$m_{extVoc}(I40KG) = 0.65$	Ratio of external properties
Used proprietary vocab.	$m_{propVoc}(I40KG) = 0.63$	34 classes and 23 proprietary properties without relations to external definitions out of 66 overall classes and 88 properties
Accessibility		
	$m_{access}(I40KG) = 1$	see Tab. 2
License		
	$m_{macLicense}(I40KG) = 1$	Machine-readable licensing available
Interlinking		
Interlinking via owl:sameAs	$m_{Inst}(I40KG) = 0$	owl:sameAs not appropriate for external linking. <i>sto:hasDBpediaResource</i> used wherever possible (for instance).
Validity of external URIs	$m_{URIs}(I40KG) = 1$	External URIs are resolvable.

Nevertheless, the expressiveness of several of the suggest criteria is certainly limited. One reason is that the I40KG covers a new domain for structured or open data, therefore no gold standard exists (cf. *Completeness*). In addition, it has been explicitly decided to avoid certain statements and relations. For instance, the validity time of standards is not determined by the publishers, making any inserted information wrong by default (cf. *Validity period*). Regarding the suggestions for interlinking resources, *owl:sameAs* would result in wrong inferences, leading to the introduction of, for instance, *sto:hasDBpediaResource* and *sto:hasWikipediaResource*.

In summary, we are confident that the I40KG meets the expectations and standards of the community, even though some metrics could not be met. We argue that the outlined characteristics support the potential user to better estimate the strengths and limits of the I40KG. Best practices and recommendations have been implemented wherever feasible. Deviations have been analyzed and consciously addressed in order to retain the best possible quality of the overall resource and to support the adoption by the community.

6 Conclusion and Future Work

In this paper, we present the Industry 4.0 Knowledge Graph depicting the latest status of standards, reference frameworks and concerns. The resource describes, connects, and outlines the most relevant information sources. We have explained the characteristics of I40KG, presented its content and outlined its various applications. The I40KG has been created following best practices, conforms by design to the Linked Data principles and is enhanced with a set of supporting tools, documentation and hosting services. It is transparently maintained and open to the community.

We identify the cumbersome search and structuring of the information resources for each involved participant as one of the most crucial obstacles for efficiently realizing Industry 4.0 use cases. The presented approach addresses precisely this challenge. The benefits of the Semantic Web technology stack can support the industrial community and furthermore reach new application areas. We have outlined how the I40KG can solve some of these issues and create added value for various target groups.

The knowledge graph will be further maintained and extended. After having reached a certain maturity level, the next steps focus on the application of I40KG in higher-level applications. The formalized knowledge of the graph can, for instance, be used to improve the performance of ML-based recommender systems. The main target was and will remain the support of the modern knowledge worker in the manufacturing industry. The faced obstacles and efforts are still too high and prevent the easy and wide implementation of Industry 4.0.

Acknowledgements

This work has been supported by the German Federal Ministry of Education and Research through the research project “Industrial Data Space Plus” (grant no. 01IS17031) and the EU H2020 project “BOOST4.0” (grant no. 780732).

References

1. Andreev, S., Galinina, O., Pyattaev, A., Gerasimenko, M., Tirronen, T., Torsner, J., et al.: Understanding the IoT Connectivity Landscape: A Contemporary M2M Radio Technology Roadmap. *IEEE Communications Magazine* **53**(9), 32–40 (2015)
2. Bader, S.R., Grangel-González, I., Tasnim, M., Lohmann, S.: Structuring the Industry 4.0 Landscape. In: *International Conference on Emerging Technologies and Factory Automation (ETFA)*. pp. 224–231. IEEE (2019)
3. Bader, S.R., Maleshkova, M., Lohmann, S.: Structuring Reference Architectures for the Industrial Internet of Things. *Future Internet* **11**(7), 151 (2019)
4. Bassi, A., Bauer, M., Fiedler, M., Kramp, T., Kranenburg, R.v., Lange, S., Meissner, S.: *Enabling Things to Talk: Designing IoT Solutions with the IoT Architectural Reference Model*. SpringerOpen, Springer, Heidelberg (2013)
5. Färber, M., Bartscherer, F., Menne, C., Rettinger, A.: Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web* **9**(1) (2018)
6. Grangel-González, I., Baptista, P., Halilaj, L., Lohmann, S., Vidal, M.E., Mader, C., Auer, S.: The Industry 4.0 Standards Landscape from a Semantic Integration Perspective. In: *ETFA 2017*. pp. 1–8. IEEE (2017)
7. Grangel-González, I., Halilaj, L., Auer, S., Lohmann, S., Lange, C., Collarana, D.: An RDF-based Approach for Implementing Industry 4.0 Components with Administration Shells. In: *ETFA 2016*. pp. 1–8. IEEE (2016)
8. Halilaj, L., Petersen, N., Grangel-González, I., Lange, C., Auer, S., Coskun, G., Lohmann, S.: Vocol: An Integrated Environment to Support Version-Controlled Vocabulary Development. In: *EKAU 2016*. vol. 10024, pp. 303–319. Springer (2016)
9. Hilliard, R., et al.: *ISO/IEC/IEEE 42010: Systems and Software Engineering - Architecture Description* (Dec 2011)
10. Lafia, S., Turner, A., Kuhn, W.: Improving Discovery of Open Civic Data. In: *10th International Conference on Geographic Information Science, GIScience*. LIPIcs, vol. 114, pp. 9:1–9:15. Schloss Dagstuhl (2018)
11. Lu, Y., Morris, K.C., Frechette, S.: Current Standards Landscape for Smart Manufacturing Systems. *NIST* **8107**, 39 (2016)
12. Noy, N.F., McGuinness, D.L., et al.: *Ontology Development 101: A Guide to Creating Your First Ontology* (2001)
13. Parent, C., Spaccapietra, S.: An Overview of Modularity. In: *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*, pp. 5–23. Springer (2009)
14. Saucedo-Martínez, J.A., Pérez-Lara, M., Marmolejo-Saucedo, J.A., Salais-Fierro, T.E.: Industry 4.0 Framework for Management and Operations: A Review. *Journal of Ambient Intelligence and Humanized Computing* **9**(3), 789–801 (Jun 2018)
15. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., et al.: The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* **3** (Mar 2016)
16. Xu, L.D., Xu, E.L., Li, L.: Industry 4.0: State of the Art and Future Trends. *International Journal of Production Research* **56**(8), 2941–2962 (2018)