Research Article

Constantin Hofmann*, Steffen Staab, Michael Selzer, Gerhard Neumann, Kai Furmans, Michael Heizmann, Jürgen Beyerer, Gisela Lanza, Julius Pfrommer, Tobias Düser, and Jan-Felix Klein

The Role of an Ontology-based Knowledge Backbone in a Circular Factory

https://doi.org/10.1515/sample- YYYY-XXXX
Received Month DD, YYYY; revised Month DD, YYYY; accepted Month DD, YYYY

Abstract: In a circular factory, new products are produced reusing parts from used products, as well as newly manufactured parts. The production system consists of disassembly, testing as well as assembly steps. Due to the unforeseeable conditions of the used parts, the complexity of such a circular factory is challenging. This paper contributes a concept of an ontology-based knowledge backbone to master the challenges of such a circular factory. The concept addresses the representation of knowledge especially taking into account uncertainty, how to design queries and means to detect similarities and analogies. Furthermore, the role of research data management with automatized workflows as a supplier for FAIR data is elaborated.

Keywords: Circular factory, ontology, research data management

1 Introduction

In the current linear economy, economical growth is intertwined with high resource utilization. One of the consequences is that more resources are used than regenerated naturally [1]. This economical model is built on exploitation of natural resources and thus not sustainable in the long run. For sustainable growth in the sense of social, ecological and economical well-fare, new approaches are required. Industrial production plays a key role in the transition to a more sustainable future [2]. One approach to reduce resource consumption whilst maintaining economical growth is the transition to a circular economy. Reusing resources for several product life cycles, for example in remanufacturing improves sustainability [3]. The ability to produce new innovative goods building on second-life materials and components in one circular factory would lead the way into sustainable industrial production. Besides the manifold advantages regarding resource utilization, such a circular economy comes at a cost. On the downside, the complexity of such a circular system increases exponentially compared to its linear counterpart. In contrast to a linear economy with its directed and well-specified material and information flows, a circular factory has to deal with reversed material flows of unknown quantities and properties at all stages of the production process. Being able to reuse these materials and components at a high rate, to learn and understand their properties and their influence on production processes are key success factors for a efficient circular production. This paper introduces approaches how to build, use and maintain knowledge in order to successfully manage the complexity of a circular factory.

For the remainder of the paper, the term circular factory describes a flexible and changeable production system where new products are produced using components from already used products of the current or even former product generations among newly manufactured components. The used products, so called cores, arrive in unknown conditions, have to be disassembled, analysed, eventually altered and later reassembled to new products. The products leaving the circular factory fulfill the quality standard of new products and are of the most recent product generation.

2 Challenges

In such a circular factory the following challenges have to be addressed regarding the process of acquiring, maintaining and accessing knowledge:

1. **C1 - Semantic representation of unique products, resources and processes with their individual conditions and properties**: Representations

***Corresponding author: Constantin Hofmann, Michael Selzer, Gerhard Neumann, Kai Furmans, Michael Heizmann, Jürgen Beyerer, Gisela Lanza, Julius Pfrommer, Tobias Düser, Jan-Felix Klein,** Karlsruhe Institute of Technology, Karlsruhe, Germany, e-mail: constantin.hofmann@kit.edu
Steffen Staab, University of Stuttgart, Institute for Artificial Intelligence Analytic Computing, Stuttgart, Germany, e-mail: steffen.staab@ipvs.uni-stuttgart.de

have to be semantic such that computers which encounter a semantic representation can draw inferences that are somewhat comparable to the ones that a human would draw. Each product leaving the circular factory consists of unique subsystems, which are partly originating from cores and have unique conditions and characteristics. Each product and its subsystems have to be described precisely. This includes data from the production and testing processes but also models and simulations.

2. **C2 - Semantic representation of uncertain knowledge**: The knowledge of production processes as well as products is uncertain. With an increased number of observations, the knowledge evolves. This knowledge has to be represented and the uncertainty expressed.
3. **C3 - Leverage experience and similarities**: Each product leaving the circular factory is unique in the sense that it consists of an individual set of components regarding their origins and degree of reprocessing. Nonetheless, there are similarities between products and processes that should be identified and used to advantage.
4. **C4 - Interoperability**: In a circular factory, different specialized subsystems work together ranging from formal models for prediction of functions to disassembly and assembly as well as measurement systems. Sharing information between these systems at different levels of aggregation is key for efficiency and consistency.
5. **C5 - Collecting data for research during the production process**: As each production order makes its way through the circular factory, valuable data is generated that has to be collected to make it exploitable for research.

3 Approach

To respond to the challenges of a circular factory, the following concept of an ontology-based knowledge backbone is proposed, see Figure 1. First, the structure of the approach is presented, which is detailed in the subsequent sections.

The resources of the circular factory performing disassembly, manufacturing, testing, inspection and assembly steps communicate with a manufacturing execution system (MES). This system provides scheduling capabilities optimizing the material flow of each order, collects data of the resources and provides various applications for deeper insights in the current state or on historic data. Whilst planning and scheduling at a higher level of aggregation has more time available for calculation, the scheduling component of the manufacturing execution system must react to deviations almost immediately and guarantee feasibility of the alternative solution. One frequently chosen strategy is to follow a top-down computed plan, but once a deviation occurs, local plan healing mechanisms are applied, giving the central planning a predefined time span and a known state to take over again. Previous works propose a declarative approach how to define process sequences without actual resource assignment. The resource assignment is done close to real-time using decentralized priority heuristics. The planning decisions together with required parameters are then transferred to the resources using Open Platform Communications Unified Architecture (OPC UA) [4]. This manufacturing execution system relies on a common knowledge representation with powerful possibilities to access knowledge. Additionally new data is automatically gathered and stored using a sophisticated research data management system.

In the described approach, ontologies are the foundation of semantic **knowledge representation** [5]. These are structured in core and sub ontologies. The questions addressed are: How to describe unique products semantically? How to semantically describe diverse data ranging from measurements, image and video data to parameter sets each with individual degrees of uncertainty? How to include the uncertainty appropriately? The second pillar addresses the ways of **accessing knowledge**. The main focus is how uncertainty can be expressed in queries and how similarities and analogies can be recognized.

For the **data storage**, triple storages such as GraphDB, RDFox, Amazon Neptune or Blazegraph, just to name a few, are used to store ontology-based knowledge graphs.

The area of **research data management** summarizes the efforts to collect data in accordance to FAIR principles. Workflow automation proposes a means to automatically collect data including a wide range of meta information [6].

3.1 Knowledge representation

Building on previous research in the context of remanufacturing [7] core and sub-ontologies are developed. The ontology uses products, processes and resources as top-level entities, see Figure 2. Whilst the core ontology defines common entities and relations, specific sub ontologies enlarge and detail the terms and relations needed for specific

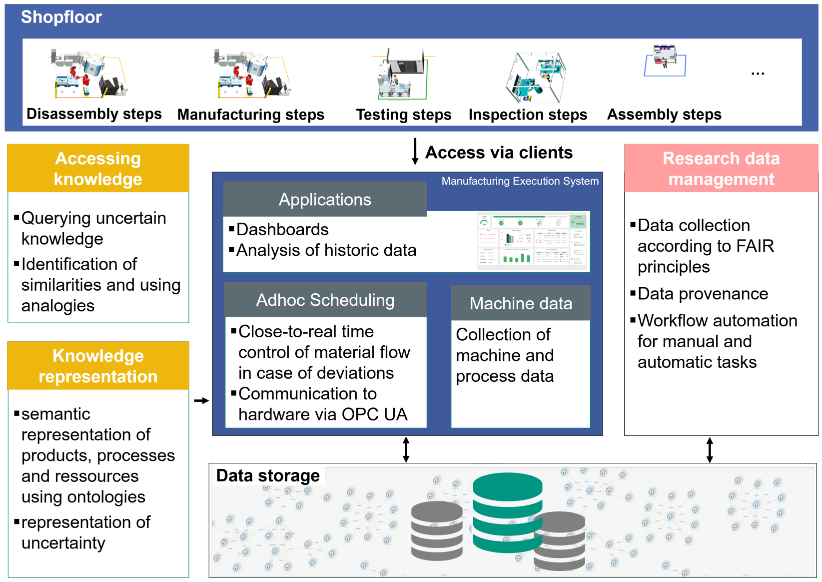

Fig. 1: Building blocks of an ontology-based knowledge backbone for circular factories.

areas of the circular factory in accordance to the core ontology, e. g. simulations or measurements. The ontologies for circular factories extend current remanufacturing ontologies by adding semantic representations to capture diverse measurement and quality data, including point clouds, image and video data as well as parameters. In addition to the representation, algebraic operations such as intersection and union operators have to be defined. Existing ontologies regarding multimedia data [8] and sensor data [9] are taken into account. Multimedia data or time series data can still be stored using dedicated and suited database technologies. However, at least the reference to the storage location has to be represented in the ontology.

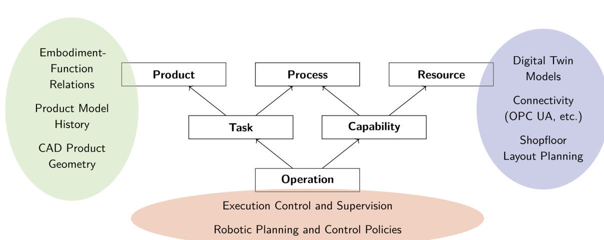

Fig. 2: Remanufacturing ontology based on [7]

Regarding uncertainty in knowledge, three uncertainty types are commonly discerned. *Attribute-based* uncertainty stands for uncertainty regarding a value. In the context of the circular factory all measurements include measurement uncertainty which is influenced by the entire chain of data acquisition and processing. The second relevant source of uncertainty is *tuple uncertainty* which refers to the probability of occurrence of a specific triple in a knowledge graph. Lastly, the *group-based* uncertainty represents statistical knowledge. In the context of the circular factory, this type will be referred to as terminology uncertainty.

Attribute-based uncertainty can be described using measurement uncertainty [10]. Thus, measurements are not simply quantities but observations of a random process which follows a known parametric distribution, e. g. Gaussian mixture models, or non-parametric distributions represented by histograms. In the context at hand, tuple-based uncertainty is used to represent the relation between parts and the whole, as it is for example unclear which parts are included in a finished good.

The terminology-based uncertainty is very important when it comes to uncertain knowledge. For example, it models the uncertainty that a certain product is not only the instance of product type A but also of product type B. The particular challenge lies in finding a flexible representation that supports discrete as well as continuous distributions but which remains easy to compute. This topic is addressed in different communities, e. g. the database research field, where each triple can be associated with a probability [11] or the knowledge representation research area [12]. Promising recent research shows how probabilistic circuits can be used to model discrete as well as continuous probability distributions. The low polynomial computing time is one of the main advantages of this approach. The first step is to represent attribute uncertainty as over-parametrized probabilistic circuits. For this task, a syntax and semantic have to be developed using existing approaches, e. g. probabilistic RDF (Resource Description Framework) [13] to express tuple and attribute based uncertainty.

3.2 Accessing knowledge

To query multi-modal measurements, the standard query language SPARQL is extended in a way that semantic and data type specific operations can be integrated seamlessly.

To request measurements taking into account attribute uncertainty, the ability to design and execute complex queries has to be integrated into SPARQL. Queries needed to retrieve, for example, products that have a probability of 95% or higher to fulfil a certain criteria require mathematical operations on the triples in the data store such as summation or integration of probabilistic distributions. Statistical EL is an extension of the EL++ part of OWL that is capable of making statistical statements about terms [14]. Currently, Statistical EL has a complexity of EXPTIME which hinders a brought application. The first step is to develop a shared syntax for probabilistic RDF, Statistical EL and Probabilistic Conjunctive Queries. The syntactical definition of probabilistic conjunctive queries also defines which queries are possible in a circular factory, e. g. map operations or conditional queries. To improve the computational effort, the requests on Statistical EL terminologies have to be approximated. Statistical EL defines the relationships between terms with intervals that represent lower and upper statistical bounds. An alternative approach is to model the meaning of terms as boxes in a high-dimensional euclidean space, see [15] where the volume represents the probability.

To query for uncertain triples probabilistic circuit requests have to be integrated into SPARQL. Probabilistic Circuit requests can be modelled using a semantics based on semi-rings [16]. At the same time, SPARQL can also be associated to an answer semantic relying on spm semirings. An interesting field of research is the approach to use an integrated semiring semantic to transform propositional logical connections on RDF graphs in a first-order representation.

Understanding similarities and analogies are difficult tasks, as these relations are not expressed in an explicit way. Early approaches project logical structures in geometric spaces to encode similarity. Both, the types of links and nodes can be encoded in such a form [17]. Further improvements include the content of the attributes along with the links and nodes to arrive at a more detailed perception of similarity [18]. More recent work use Relational Graph Convolutional Networks (Relational GCN) to encode graph structures. Then, a decoder can predict links and nodes in that latent space [19]. As an extension of the current state of the art, the relational GCN HeterogeneousEdgeGATEncoder [20] can be adapted to the requirements of a circular factory. In particular, new modules are needed to cope with the data types of a circular factory, e. g. regarding measurements. The encoded structures are then decoded by two different decoders. One focusing on the similarities and one predicting additional links.

3.3 Research data management

The circular factory produces data permanently which can be used to improve the understanding of products, resources and processes. All this data must be stored in or linked to ontology-based knowledge graphs. A key success factor is to store the data according to the FAIR principles: Data must be **F**indable, **A**ccessible, **I**nteroperable and **R**eusable. The ontologies and subontologies, see 3.1 in combination with the new ways of accessing knowledge in a targeted and intuitive way, see 3.2, play a major role in achieving this FAIR-principles. However, the usefulness of data is predetermined during the collection process. Especially meta data is important to put data in context and to pave the way for future reuse.

To that end, a system for research data management in a circular factory is developed based on proven research data management systems from the context of material science [21]. This system stores the data in ontology-based knowledge graphs and thus works seamlessly with other subsystems of the circular factory. This is archived by a REST-API which can be used along the workflows of the circular manufacturing processes. In contrast to the manual web-based upload mechanism this enables (semi-) automated capturing and documentation of the data provenance of each data set and corresponding meta data independent of its origin and used data formats.

The circular factory consists of manual and automatized work stations. Both are important sources of information. One key asset is workflow automation for data collection according to the FAIR principles. In the context of the circular factory, workflows define the conditional sequence of required steps. In each step, data with its associated meta data is collected and automatically stored in the knowledge graph. First applications in a lab environment demonstrate the positive effects on data availability and quality [22]. In the current state, workflows are based mainly on manually defined steps. In case of manual interaction, users are prompted to supply the information which is then automatically stored in the knowledge graph, see figure 3. In the future, the workflows will use more and more automatized steps to reduce the amount of user interactions and to improve the quality of the collected data by reducing errors due to manually transferring information between different systems and by following the structure of the knowledge graph more closely.

In the domain of mechanics, the research data management system has been used to speed up the development of surrogate models, mainly by using multiple data sources in parallel and framing the process in workflows [23]. In the context of the circular factory, simulation and models also play a key role and guiding this process would be a key aspect to ensure the availability of FAIR data.

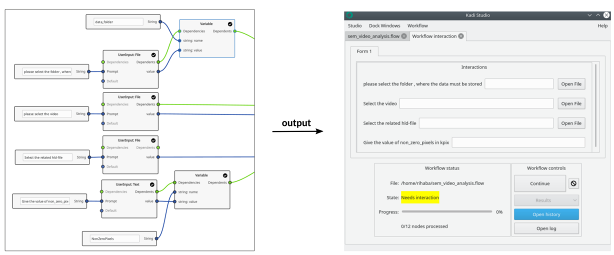

Fig. 3: A workflow sequence and its automatically generated, corresponding user prompt [22]

4 Conclusion

Five challenges regarding building, maintaining and accessing knowledge in a circular factory have been discussed. The presented concept consists of the three pillars **representation of knowledge**, **access of knowledge** and **research data management system**. The three building blocks enable the full extent of the understanding and of the control of the circular factory.

The representation of knowledge is based on core and sub ontologies. These are tailored to the needs of the ciruclar factory and extend current approaches by providing new ways to represent diverse measurement data as well as uncertain knowledge in terms of attribute uncertainty, tuple uncertainty and terminology-based uncertainty.

To access this knowledge the well-proven language SPARQL is extended to query multi-modal measurements, form requests including uncertainty and even to request uncertain terms. In the presented concept, an approach is included how to detect similarities and analogies using Relational Graph Convolutional Networks that are based on links, nodes as well as properties and capable of evaluation similarities and predicting new links.

The research data management system continues the work currently tested in lab environments and proposes automatic workflows to collect data during the process and store it automatically in ontology-based knowledge graphs.

5 Outlook

Whilst this contribution is at the stage of a sound coherent concept, future research will focus on the validation of the building blocks.

Funding: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 471687386 - SFB 1574

References

[1] D. Lin, L. Wambersie, and M. Wackernagel, "Estimating the date of earth overshoot day 2022," *Nowcasting the World's Footprint & Biocapacity for*, pp. 1–8, 2021.

[2] IEA, "Global co2 emissions by sector, 2019-2022," *IEA, Paris, https://www.iea.org/data-and-statistics/charts/global-co2-emissions-by-sector-2019-2022*, 2022.

[3] A. Benoy, L. Owen, and M. Folkerson, "Triple win-the social, economic and environmental case for remanufacturing," *All-Party Parliamentary Sustainable Resource Group & All-Party Parliamentary Manufacturing Group, London*, 2014.

[4] D. Lünsch, P. Detzner, A. Ebner, and S. Kerner, "Swap-it: A scalable and lightweight industry 4.0 architecture for cyber-physical production systems," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, pp. 312–318, IEEE, 2022.

[5] S. Staab and R. Studer, *Handbook on ontologies*. Springer Science & Business Media, 2010.

[6] L. Griem, P. Zschumme, M. Laqua, N. Brandt, E. Schoof, P. Altschuh, and M. Selzer, "Kadistudio: Fair modelling of scientific research processes," *Data Science Journal*, vol. 21, pp. 16–16, 2022.

[7] J. Pfrommer, J.-F. Klein, M. Wurster, S. Rapp, P. Grauberger, G. Lanza, A. Albers, S. Matthiesen, and J. Beyerer, "An ontology for remanufacturing systems," *at - Automatisierungstechnik*, vol. 70, no. 6, pp. 534–541, 2022.

[8] R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura, "Comm: designing a well-founded multimedia ontology for the web," in *The Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007+ ASWC 2007, Busan, Korea, November 11-15, 2007. Proceedings*, pp. 30–43, Springer, 2007.

[9] A. Haller, K. Janowicz, S. J. Cox, M. Lefrançois, K. Taylor, D. Le Phuoc, J. Lieberman, R. García-Castro, R. Atkinson,

and C. Stadler, "The modular ssn ontology: A joint w3c and ogc standard specifying the semantics of sensors, observations, sampling, and actuation," *Semantic Web*, vol. 10, no. 1, pp. 9–32, 2019.

[10] I. O. for Standardization, *Uncertainty of measurement-Part 3: Guide to the expression of uncertainty in measurement (GUM: 1995)*. ISO, 2008.

[11] G. Gaur, A. Dang, A. Bhattacharya, and S. Bedathur, "Computing and maintaining provenance of query result probabilities in uncertain knowledge graphs," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 545–554, 2021.

[12] T. Lukasiewicz and U. Straccia, "Managing uncertainty and vagueness in description logics for the semantic web," *Journal of Web Semantics*, vol. 6, no. 4, pp. 291–308, 2008.

[13] X. Lian and L. Chen, "Efficient query answering in probabilistic rdf graphs," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pp. 157–168, 2011.

[14] R. Peñaloza and N. Potyka, "Towards statistical reasoning in description logics over finite domains," in *International Conference on Scalable Uncertainty Management*, pp. 280–294, Springer, 2017.

[15] B. Xiong, N. Potyka, T.-K. Tran, M. Nayyeri, and S. Staab, "Faithful embeddings for el++ knowledge bases," in *International Semantic Web Conference*, pp. 22–38, Springer, 2022.

[16] V. Belle and L. De Raedt, "Semiring programming: A semantic framework for generalized sum product problems," *International Journal of Approximate Reasoning*, vol. 126, pp. 181–201, 2020.

[17] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," *Advances in neural information processing systems*, vol. 26, 2013.

[18] G. A. Gesese, R. Biswas, M. Alam, and H. Sack, "A survey on knowledge graph embeddings with literals: Which model links better literal-ly?," *Semantic Web*, vol. 12, no. 4, pp. 617–647, 2021.

[19] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*, pp. 593–607, Springer, 2018.

[20] T. Monninger, J. Schmidt, J. Rupprecht, D. Raba, J. Jordan, D. Frank, S. Staab, and K. Dietmayer, "Scene: Reasoning about traffic scenes using heterogeneous graph neural networks," *IEEE Robotics and Automation Letters*, vol. 8, no. 3, pp. 1531–1538, 2023.

[21] N. Brandt, L. Griem, C. Herrmann, E. Schoof, G. Tosato, Y. Zhao, P. Zschumme, and M. Selzer, "Kadi4mat: A research data infrastructure for materials science," *Data Science Journal*, Feb 2021.

[22] R. Al-Salman, C. A. Teixeira, P. Zschumme, S. Lee, L. Griem, J. Aghassi-Hagmann, C. Kirchlechner, and M. Selzer, "Kadistudio use-case workflow: Automation of data processing for in situ micropillar compression tests," *Data Science Journal*, vol. 22, pp. Art.–Nr.: 21, 2023. 43.31.01; LK 01.

[23] A. Koeppe, F. Bamer, M. Selzer, B. Nestler, and B. Markert, "Workflow concepts to model nonlinear mechanics with computational intelligence," *PAMM*, vol. 21, no. 1, p. e202100238, 2021.