A Roadmap to Create a Knowledge Graph for the Circular Factory for the Perpetual Product

Ratan Bahadur Thapa[1], Daniel Hernández[1], Nico Brandt[3], Jan-Felix Klein[5], Etienne Hoffmann[5], Steffen Staab[1,2], Michael Selzer[3] and Gisela Lanza[4]

[1]*Analytic Computing, Institute for Artificial Intelligence, University of Stuttgart, Stuttgart, Germany*
[2]*Electronics and Computer Science, University of Southampton, United Kingdom*
[3]*Institute for Applied Materials (IAM), Karlsruhe Institute of Technology (KIT), Karlsruhe, 76131, Germany*
[4]*wbk Institute of Production Science, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany*
[5]*Institute for Material Handling and Logistics (IFL), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany*

Abstract

New economic systems are needed to decouple resource consumption from wealth. The linear economic approach of “take-make-use-dispose” is not a recipe for success in the long term. Circular production offers a solution to this problem. Our Collaborative Research Center 1574 aims to enable integrated linear and circular production on an industrial scale. To this end, the Collaborative Research Center 1574 investigates how used products and their multiple generations can achieve the vision of the perpetual product. This research involves multiple scientific questions related to production technology, product development and materials technology, ergonomics, robotics, computer science, and knowledge modeling. The Collaborative Research Center 1574 involves eighteen subprojects. Each one studies a dimension of these multiple scientific questions. One of these subprojects, called INF, aims to provide an infrastructure and teach the other subprojects’ researchers how to operate and integrate all the data they produced into a unified knowledge graph. This paper describes the roadmap of the INF subproject, including the ongoing work, future steps, and vision of the INF subproject.

Keywords

Circular Factory, Ontology, Knowledge Graph, Sustainability

1. Introduction

The linear economic approach of “take-make-use-dispose” is not a recipe for success in the long term because it consumes high levels of energy and materials. Hence, new economic systems are needed to decouple resource consumption from wealth. A circular production that repairs products and minimizes material consumption offers a solution to this problem. The Collaborative Research Center 1574 (CRC1574), *Circular Factory for the Perpetual Product*, is a research center created in April of 2024 and funded by the German Research Foundation (DFG) to achieve the integrated linear and circular production on an industrial scale [1].

To make concrete the application of the study, the CRC1574 chooses the production of angle grinders as the object of study. Angle grinders have several comparative advantages for the study. First, their generations are relatively short. New angle grinder models appear each year. Also, generations include

KG4S 2025: The 3rd Knowledge Graphs for Sustainability Workshop, June 01–05, 2025, Portoroz, Slovenia
ratan.thapa@ki.uni-stuttgart.de (R. B. Thapa); daniel.hernandez@ki.uni-stuttgart.de (D. Hernández); nico.brandt@kit.edu (N. Brandt); jan-felix.klein@kit.edu (J. Klein); etienne.hoffmann@kit.edu (E. Hoffmann); steffen.staab@ki.uni-stuttgart.de (S. Staab); michael.selzer@kit.edu (M. Selzer); gisela.lanza@kit.edu (G. Lanza)
https://www.ki.uni-stuttgart.de/institute/team/Thapa/ (R. B. Thapa);
https://www.ki.uni-stuttgart.de/institute/team/Hernandez/ (D. Hernández);
https://www.iam.kit.edu/mms/Mitarbeiter_4328.php (N. Brandt); https://www.ifl.kit.edu/english/6069_jan_felix_klein.php (J. Klein); https://www.ifl.kit.edu/mitarbeiter_5539.php (E. Hoffmann);
https://www.ipvs.uni-stuttgart.de/de/institut/team/Staab-00001/ (S. Staab);
https://www.iam.kit.edu/mms/Mitarbeiter_selzer.php (M. Selzer); https://www.wbk.kit.edu/21_90.php (G. Lanza)
0009-0000-2368-5928 (R. B. Thapa); 0000-0002-7896-0875 (D. Hernández); 0000-0002-3860-1376 (N. Brandt); 0000-0002-3704-9567 (J. Klein); 0000-0002-0780-4154 (S. Staab); 0000-0002-9756-646X (M. Selzer); 0000-0002-0481-4613 (G. Lanza)

multiple angle grinder models, which can differ in their capacities (e.g., the maximum rotation speed can vary), their functionalities (e.g., some are powered with a plug, whereas an internal battery powers others), and their components (e.g., some components are specific to one angle grinder, whereas others can be common to multiple models). This diversity of the angle grinder designs challenges the operations of a circular factory. In particular, when transferring the knowledge about one angle grinder model to another angle grinder model.

Like a real circular factory, the CRC1574 encompasses various subprojects focused on production technology, product development, materials science, ergonomics, robotics, computer science, and knowledge modeling. A crucial element in managing the composition of the factory's modular yet interconnected and interdependent subprojects is the uniform knowledge sharing among all participants involved in the circular production process. To support this, the CRC1574 includes a special subproject, called INF, whose goal is to bring two services: a *shared data repository* and a *knowledge graph*, as well as a suite of interconnecting interfaces to synchronize, update, enlarge and query the aforementioned services directly from the shop floor.

In addition to these two services, the INF subproject aims to teach team members of the other subprojects how to contribute and use the shared data repository and the knowledge graph. Each subproject is responsible for publishing its data in the shared repository and some data into the shared knowledge graph. This data can then be accessed throughout the whole collaborative research center to ensure product- and process integrity, allow reuse of expensively procured research data and enable process improvement through learning from the past. To this end, the team of each subproject must define some specific metadata for the data they are sharing and an ontology to describe the data they publish in the knowledge graph. However, most do not have expertise in knowledge graphs or ontologies. Indeed, they are PhD students whose research is about production engineering, systems engineering, informatics, product development, production technology, labor science, logistics, and robotics. This lack of expertise justifies the aforementioned teaching responsibility for the INF subproject.

The main contributions of this paper to the workshop are sharing the ontology modeling experience of the CRC1574 with the workshop audience and presenting our vision for the data infrastructure for the circular factory.

This paper is structured as follows. In Section 2, we describe the structure of the CRC1574, including its subprojects and the interaction between them; in Section 3, we present the roadmap of the INF subproject; in Section 4 we present the roadmap status and our conclusions.

2. The structure of the CRC1574

The CRC1574 is divided into three project areas: Project Area A is responsible for the overarching design, modeling, and control of the circular factory through a highly connected product-production co-design. Project Area B studies the acquisition and modeling of the individual product instance and the multimodal acquisition of people during object interactions. Project area C enables the implementation of the autonomous and modular production system. Each project is associated with one of eight specialties: production engineering, systems engineering, informatics, product development, production technology, labor science, logistics, and robotics.

In addition to the project areas, the CRC1574 includes some special subprojects for the whole functioning of the project. One special subproject is the INF, titled "Data Infrastructure for the Circular Factory," whose team are the authors of this paper. The goals of the INF subproject are: (1) Enabling the knowledge exchange between the subprojects in the Collaborative Research Center. (2) Providing the information infrastructure for research data management to guarantee the fulfillment of high standards in research data management and sustainability in accordance with the FAIR Data principles (Findable, Accessible, Interoperable, and Reusable Data). (3) Providing the information and automation technology infrastructure for the operation of the circular factory, particularly the control of orders.

A relevant aspect that the INF project must consider is that compared to a linear factory, a circular factory requires a higher level of interaction between the organizational units of a factory. For example,

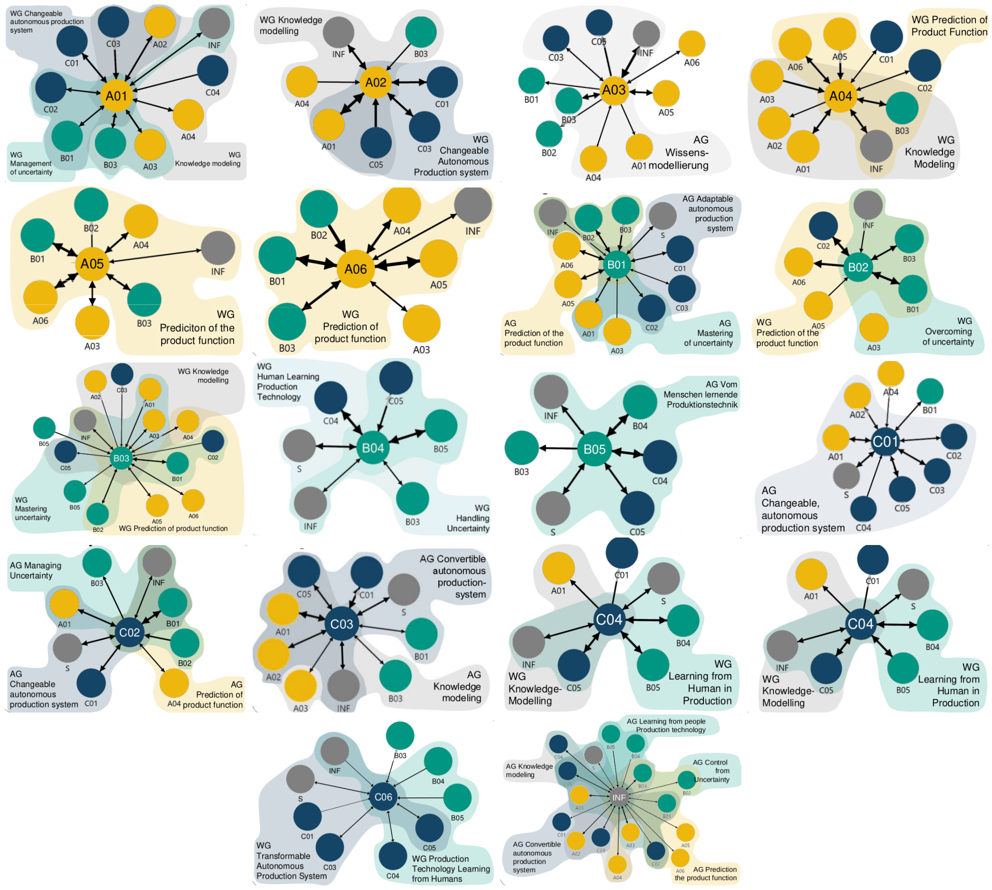

Figure 1: Expected networking between subprojects. The direction of the arrows indicate the direction of the knowledge transfer and the thickness indicate the relevance.

to facilitate the repair of products, the product designers must coordinate with those who will automate the product disassembly process. Similarly, the design of product generations must consider the reuse of components to simplify component stock management. The CRC1574 also expects a high interaction between the subprojects (see Figure 1).

3. INF subproject roadmap

This section outlines the INF subproject roadmap defined in October 2024 to be followed until December 2027 to achieve three key objectives. First, the *infrastructure goal* aims to have the two systems, namely a file repository and a triple store with a SPARQL endpoint (see Figure 2), fully operational and interoperable within the project specific runtime framework to support the functioning of the circular factory. Second, the *model goal* is to establish a unified model for data and metadata stored in these systems, focusing on the *ontology* that will structure the factory's knowledge. Finally, the *modeling know-how goal* specifies training components for researchers at the circular factory to define their knowledge models, adhering to existing *standard* ontologies, and to seamlessly integrate them with the models of other units for smooth data interoperability.

3.1. Infrastructure

We consider two systems: Kadi4Mat [2, 3], a virtual research environment for storing and managing data and metadata, and a SPARQL endpoint for managing knowledge graphs. To be able to keep these systems always up to date and make them accessible from the shopfloor, we plan to integrate them into a distributed framework, that allows for querying the knowledge graph via industrial protocols

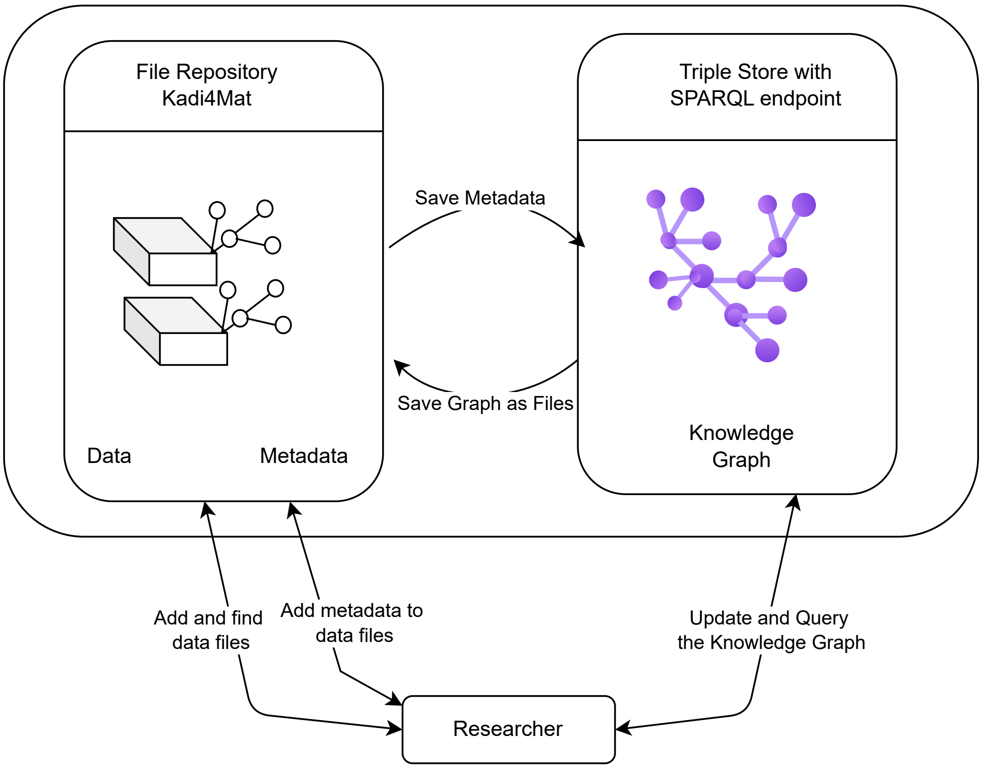

Figure 2: Circular Factory System: The triple store with SPARQL endpoint (right) allows researchers to update and query the knowledge graph — i.e., structured information about the circular factory. The file system (left) stores videos, images, and data formats that are not compatible with the structured data in the knowledge graph. In the file repository, researchers can add, locate, and edit files and their metadata.

such as OPC-UA or MQTT. Both systems will require a data governance framework to determine what researchers can do and their access levels. We envision that such data governance should be able to define or assign roles to working groups that represent the organizational units of a circular factory, preventing participants from accidentally deleting data belonging to other units.

Kadi4Mat already provides a means to organize users into groups, data and metadata into records, and multiple records into a hierarchy of collections. Each resource can be assigned differentiated access for individual users and groups. Knowledge graph management systems, on the other hand, provide different means for data governance. For example, the *GraphDB* engine [4] has an out-of-the-box web interface to define RDF datasets (which can contain multiple knowledge graphs) and assign users to them with different access roles. In contrast, the *Fuseki-Jena* engine [5] allows defining graph and RDF dataset control list but does not provide an out-of-the-box web interface to edit these control lists. We compared several engines and decided to use GraphDB for the initial phase of the CRC1574 because of the simplicity of the deployment and its out-of-the-box web interface to configure data repositories and users. In addition to that, the GraphDB Engine allows notifications on changes of predefined triples or named graphs, which is essential to make a knowledge graph the ground truth database for a fully automated circular factory, that is not dependent on polling.

Objectives

I. *Setting up Kadi4Mat.* Ensure that all researchers have accounts in Kadi4Mat, and each subproject has a dedicated collection for data uploads.

II. *Knowledge Graph Engine Selection.* Select a knowledge graph engine that supports the features needed to manage users and RDF datasets.

III. *Setting up knowledge Graph Engine.* Ensure that researchers can access the knowledge graph engine, create RDF datasets for their subprojects, and query and update data within the RDF datasets to which they have access.

IV. *Setting up RDF Data Integration.* Ensure that researchers can query an integration dataset that combines data from all subprojects.

V. *Setting up a Metadata RDF dataset* Ensure that all researchers can query the metadata of Kadi4Mat repositories and the available knowledge graphs using an RDF dataset.

Plan. We planned the work on Objectives I and II for November 2024, Objective III for November and December 2024, Objective IV for December 2024, and Objective V for January and December 2024.

3.2. Model

In this road map, we envision each subproject defining an ontology for a specific set of competency questions. With our guidance, the subprojects will align their ontologies. To this end, we will include a phase of detecting commonalities across the multiple subproject ontologies, and we will model these commonalities into a core ontology [6]. Similarly, we will analyze common patterns to encourage the uniformity of the model.

Objectives

I. Phase 1: First individual modeling cycle. At the end of this phase, we will document the existing resources and the resources each subproject will produce. Each subproject will have its initial ontology derived from its key competency questions. We will introduce standard ontologies and capture common patterns for the very beginning of this phase.

II. Phase 2: Alignment with the core ontology. At the end of this phase, each subproject should be aligned with the core ontology and follow similar ontology patterns.

III. Phase 3: Extensions and improvements. By the end of this phase, the ontologies will be more mature and thoroughly documented.

Plan. We planned the work of Objective I for November and December 2024, Objective II for January to August 2025, and Objective III for September 2025 to 2027.

3.3. Modeling know-how

Our teaching roadmap aims to equip researchers with foundational modeling skills and an understanding of the core ontology. We organize it with the milestones outlined below. The primary tool for achieving these milestones is a series of modeling workshops held every one to two months, based on the working group progress assessed by the principal investigators.

Objectives

I. After 2nd modeling workshop: Researchers should know how to define competency questions and understand designing ontologies.

II. After 2nd modeling workshop: Researchers should know how to populate and query data from the triple store.

III. After Kadi4Mat Demo: Researchers should know how to upload and access files from Kadi4Mat.

IV. After 3rd modeling workshop: Researchers should have a solid understanding of the core ontology.

V. After 3rd modeling workshop: Researchers should be familiar with *standard (domain-specific) ontologies* related to their subproject ontologies.

VI. After 4th modeling workshop: Researchers should be familiar with *Ontology Design Patterns*.

VII. After 5th modeling workshop: Researchers should know how to use *SHACL* to define schemas and validate data.

Plan. We planned Objectives I and II for November and December 2024, Objective III for November 2024, Objectives IV and V for January to April 2025, Objective VI for May to June 2025, and Objective VII for July to October 2025. We planned workshops for the first two years in August 2024, October 2024, January 2025, and April 2025.

4. Plan status and conclusions

We fulfilled the infrastructure objectives in the planned time. However, we observe additional needs for the project, like setting up code repositories for developing tools, a wiki for documentation, and a linked data service for making dereferenceable the URLs of resources described in the knowledge graph. We already identified the need of the following tools to facilitate the work for PIs and PHD Students working with the ontologies and the data infrastructure

1. a service, that automatically creates records in Kadi4Mat, that are cross referenced to the respective entities in the Graph Database

2. a guided way of creating instances of the modeled concepts, that helps the PHD students with correct modeling of the instances and that ensures completeness of the modeled data.

3. a runtime system, that allows to directly push status updates to these instances without human intervention

We will need to update our roadmap to consider these new needs.

The modeling part was more complex than expected because the participants faced difficulties describing their projects and understanding modeling concepts like the competency questions. So far, we have organized four full-day workshops that include concepts on modeling and modeling activities that the researchers have to practice with their own datasets. However, to improve and accelerate the modeling and data sharing, we started working one-on-one with the researchers of each subproject. This is also believed to keep the motivation of participating in Modeling high, since we can provide concrete examples at subproject level on how modeling can improve the usability and reusability of the developed systems/models/logics and how the increased interconnection between subprojects can increase the quality of the circular factory as a whole.

The first modeling tool we taught to the researchers was Protégé. However, we found that the visualization of the ontologies was not intuitive enough. Hence, we plan to use Chowlk [7], a graphical language and tool in what follows.

So far, the subprojects have generated fifteen preliminary ontologies describing resources, like the components of an angle grinder, disassembly processes, and logistics. These ontologies are motivated by five competency questions and are instantiated with SPARQL queries.

In addition to defining ontologies, the researchers must generate data in RDF format and upload it to the SPARQL endpoint. To facilitate this work, we avoided mapping languages like RML, which require an additional teaching effort and used scripting languages like Python to generate JSON-LD data.

Acknowledgments

This work is funded by the German Research Foundation (DFG) - SFB 1574 - 471687386.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to grammar and spell check, paraphrase, and improve the text readability. After using the tool, the authors reviewed and edited the content as needed to take full responsibility for the publication's content.

References

[1] SFB1574 - Circular Factory for the Perpetual Product, 2025. URL: https://www.sfb1574.kit.edu/english/index.php.

[2] N. Brandt, L. Griem, C. Herrmann, E. Schoof, G. Tosato, Y. Zhao, P. Zschumme, M. Selzer, Kadi4mat: A research data infrastructure for materials science, Data Sci. J. 20 (2021) 8. URL: https://doi.org/10.5334/dsj-2021-008. doi:10.5334/DSJ-2021-008.

[3] N. Brandt, N. T. Garabedian, E. Schoof, P. J. Schreiber, P. Zschumme, C. Greiner, M. Selzer, Managing FAIR tribological data using kadi4mat, Data 7 (2022) 15. URL: https://doi.org/10.3390/data7020015. doi:10.3390/DATA7020015.

[4] GraphDB Ontotext, 2025. URL: https://www.ontotext.com/products/graphdb/.

[5] Apache Jena, 2025. URL: https://jena.apache.org/.

[6] J. Klein, M. Wurster, N. Stricker, G. Lanza, K. Furmans, Towards ontology-based autonomous intralogistics for agile remanufacturing production systems, in: 26th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2021, Vasteras, Sweden, September 7-10, 2021, IEEE, 2021, pp. 1–7. URL: https://doi.org/10.1109/ETFA45728.2021.9613486. doi:10.1109/ETFA45728.2021.9613486.

[7] S. Chávez-Feria, R. García-Castro, M. Poveda-Villalón, Chowlk: from uml-based ontology conceptualizations to OWL, in: P. Groth, M. Vidal, F. M. Suchanek, P. A. Szekely, P. Kapanipathi, C. Pesquita, H. Skaf-Molli, M. Tamper (Eds.), The Semantic Web - 19th International Conference, ESWC 2022, Hersonissos, Crete, Greece, May 29 - June 2, 2022, Proceedings, volume 13261 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 338–352. URL: https://doi.org/10.1007/978-3-031-06981-9_20. doi:10.1007/978-3-031-06981-9_20.