

Organizing Scientific Knowledge from Engineering Sciences Using the Open Research Knowledge Graph: The Tailored Forming Process Chain Use Case



OLIVER KARRAS

LAURA BUDDE

PAULINA MERKEL

JÖRG HERMSDORF

MALTE STONIS

LUDGER OVERMEYER

BERND-ARNO BEHRENS

SÖREN AUER

*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

Background: Engineering sciences are essential for addressing contemporary technical, environmental, and economic challenges. Despite its data-intensive and interdisciplinary nature, the organization of Findable, Accessible, Interoperable, and Reusable (FAIR) scientific knowledge and data in this research field remains understudied. Engineers need infrastructures with services that support them in organizing FAIR scientific knowledge and data for communication and (re-)use.

Aim: We explore the use of the Open Research Knowledge Graph (ORKG) as such an infrastructure by demonstrating how engineers can utilize the ORKG in innovative ways for communication and (re-)use.

Method: For a use case from the Collaborative Research Center 1153 “Tailored Forming”, we collect, extract, and analyze scientific knowledge on 10 Tailored Forming Process Chains (TFPCs) from five publications in the ORKG. In particular, we semantically describe the TFPCs, i.e., regarding their steps, manufacturing methods, measurements, and results. The usefulness of the data extraction topics, their organization, and the relevance of the knowledge described is examined by an expert consultation with 21 experts.

Results: Based on the described knowledge, we build and publish an ORKG comparison as a detailed overview for communication. Furthermore, we (re-)use the knowledge and answer eight competency questions asked by two domain experts. The validation shows a clear agreement of the 21 experts regarding the examined usefulness and relevance.

Conclusions: Our use case shows that the ORKG as a ready-to-use infrastructure with services supports researchers, including engineers, in sustainably organizing FAIR scientific knowledge. The direct use of the ORKG by engineers is feasible, so the ORKG is a promising infrastructure for innovative ways of communicating and (re-)using FAIR scientific knowledge in engineering sciences, thus advancing this research field.

CORRESPONDING AUTHOR:

Oliver Karras

TIB – Leibniz Information
Centre for Science and
Technology, Hannover,
Germany

oliver.karras@tib.eu

KEYWORDS:

Tailored forming; scientific
knowledge; communication;
(re-)use; infrastructure;
knowledge graph

TO CITE THIS ARTICLE:

Karras, O., Budde, L., Merkel, P.,
Hermsdorf, J., Stonis, M.,
Overmeyer, L., Behrens, B.-A.
and Auer, S. (2024) Organizing
Scientific Knowledge from
Engineering Sciences Using
the Open Research Knowledge
Graph: The Tailored Forming
Process Chain Use Case.
Data Science Journal, 23: 52,
pp. 1–36. DOI: <https://doi.org/10.5334/dsj-2024-052>

1 INTRODUCTION

Engineering sciences are essential in developing solutions to our modern society's technical, environmental, and economic challenges. For example, engineers (re-)use knowledge and data from mechanical processes to develop new or revised processes for manufacturing components that meet these challenges. For this purpose, they must be able to understand and reproduce all steps of engineering (research) processes to guarantee the trustworthiness of published results, to prevent redundancies, and to ensure their social acceptance (Schmitt et al., 2020). However, engineering projects are interdisciplinary and typically comprise several sub-projects that must collaborate closely (Sheveleva et al., 2020). Consequently, engineers must communicate and (re-)use heterogeneous knowledge and data (Prakash and Sandfeld, 2018). According to various researchers (Kapogiannis and Sherratt, 2018; Schmitt et al., 2020; Sheveleva et al., 2020), engineers need infrastructures with services to support them in the organizing of scientific knowledge and data for communication and (re-)use. Although the importance of Findable, Accessible, Interoperable, and Reusable (FAIR) (Wilkinson et al., 2016) scientific knowledge and data is increasing in engineering sciences, this topic remains understudied despite the data-intensive and interdisciplinary nature of this research field (Schmitt et al., 2020).

Historically, scientific knowledge and data are published in document-based artifacts such as publications, software, and datasets. Nowadays, these artifacts are published as digitized files in various formats, e.g., PDFs and ZIP archives, hosted on numerous repositories (Auer et al., 2020; Karras et al., 2023c). While humans can easily process these digitized document-based artifacts, machines cannot easily access and interpret them, as these artifacts are not FAIR and thus not machine-actionable (van de Sompel and Lagoze, 2009). Sustainable (re-)use of scientific knowledge and data requires a more flexible, fine-grained, semantic, and context-sensitive representation that can be understood, processed, and (re-)used by humans and machines (Auer et al., 2020; Karras et al., 2023c). Over the last decade, Knowledge Graphs (KGs) have become an emerging technology in industry and academia as they enable this versatile representation (Auer et al., 2018; Hussein et al., 2022; Zou, 2020). Besides well-known KGs for encyclopedic and factual knowledge and data, such as DBpedia (Auer et al., 2007) and WikiData (Vrandečić and Krötzsch, 2014), using so-called Research Knowledge Graphs (RKGs) for scientific knowledge and data is a rather new approach (Ammar et al., 2018; Auer et al., 2018; Dessì et al., 2022). RKGs generally include bibliographic metadata, e.g., titles, authors, and venues, as well as scientific knowledge and data, e.g., processes, methods, measurements, and results (Gkatzelis et al., 2021; Jaradeh et al., 2019; Jeschke et al., 2020; Papers With Code, 2020; Penev et al., 2019; Spadaro et al., 2022). They are a promising technology to sustainably organize scientific knowledge and data so that all information is openly accessible in the long term (Auer et al., 2023b; Stocker et al., 2022). This potential led to an increasing amount of research and approaches working on solutions for infrastructures with services using RKGs (Zou, 2020).

In 2020, the consortium *NFDI for Engineering Sciences*¹ (NFDI4Ing), part of the *National Research Data Infrastructure*² (NFDI), in Germany, began developing and deploying infrastructures with services for engineers to organize FAIR scientific knowledge and data (Schmitt et al., 2020). One of the infrastructures is the *Open Research Knowledge Graph*³ (ORKG). The ORKG is a cross-domain and cross-topic RKG for the production, curation, and (re-)use of machine-actionable FAIR scientific knowledge originally published in articles (Auer et al., 2023b; Stocker et al., 2023). For this purpose, the ORKG offers accompanying services that combine manual crowdsourcing and (semi-)automated approaches to organize scientific knowledge (Karras et al., 2021). Beyond NFDI4Ing, the ORKG became an integral part of the NFDI (Rossenova et al., 2023) and three other engineering consortia—*NFDI for Interdisciplinary Energy System Research*⁴ (NFDI4Energy) (Nieße et al., 2022), *NFDI for Data Science and Artificial Intelligence*⁵

1 <https://nfdi4ing.de/>.

2 <https://www.nfdi.de/?lang=en>.

3 <https://orkg.org>.

4 <https://nfdi4energy.uol.de/>.

5 <https://www.nfdi4datascience.de/>.

Given the central role of the ORKG in the NFDI and especially for the four engineering consortia, we explore how engineers can benefit from the ORKG infrastructure and services by organizing FAIR scientific knowledge for communication and (re-)use. For this purpose, we organize scientific knowledge of a use case on *Tailored Forming Process Chains* (TFPCs) from the *Collaborative Research Center* (CRC) 1153 “*Tailored Forming*”⁷ (cf. section 2.3). This use case is the first attempt to describe TFPCs semantically in the ORKG and show how engineers can organize scientific knowledge beyond publications. We focus on a smaller but complete set of five publications (Behrens et al., 2019; Budde et al., 2022; Coors et al., 2020; Kruse et al., 2019; Pape et al., 2018) from the CRC 1153 reporting 10 TFPCs on the topic of “*Tailored Forming Process Chains for the Manufacturing of Hybrid Components with Bearing Raceways Using Different Material Combinations*.” From these publications, we semantically describe scientific knowledge on 10 reported TFPCs regarding the individual steps, their sequences, their incoming and outgoing components, the manufacturing methods, and the qualities investigated, including the measurement methods used and the measurement results. Based on the described knowledge, we demonstrate how engineers can utilize the ORKG infrastructure in innovative ways for communication and (re-)use of scientific knowledge in the field of engineering sciences. First, we build and publish an ORKG comparison (Karras et al., 2023a) to provide a detailed overview of the described knowledge about the 10 TFPCs for communication. Second, we answer eight competency questions related to the described knowledge, which were asked by two domain experts who did not know the actual content of the ORKG comparison (Karras et al., 2023a). In this way, we show how the scientific knowledge from individual publications can be jointly (re-)used to gain new overviews and insights. We validate the results by consulting experts in engineering sciences to assess the usefulness of the data extraction topics, their organization, and the relevance of the competency questions with their answers for engineering sciences. In addition, we apply our approach for describing manufacturing process chains to publications outside CRC 1153 to demonstrate its applicability. Overall, we provide the following contributions:

Contribution:

- 1) A use case on the sustainable organization of FAIR scientific knowledge for communication and (re-)use in engineering sciences.
- 2) A reusable, expandable, and validated ORKG template for the semantic description of scientific knowledge on TFPCs.
- 3) A citable and versionable ORKG comparison on TFPCs for the manufacturing of hybrid components with bearing raceways using different material combinations, including all supplementary materials (Karras et al., 2023a).
- 4) A reproducible data analysis of the described scientific knowledge to answer eight competency questions, showing how the (re-)use of the knowledge can lead to new overviews and insight, that experts assessed as relevant for engineering sciences (Karras, 2024).

Article structure: Section 2 presents background and related work. While section 3 describes the research approach, section 4 and section 5 reports the results. Section 6 addresses threats to validity, and section 7 discusses the findings. Section 8 concludes the paper.

2 BACKGROUND & RELATED WORK

Below, we first present RKGs in general and give a brief overview of some exemplary RKGs, including the ORKG, as an initial basic orientation for readers. We also explain in more detail why we decided to use the ORKG. We then present some use cases in which the ORKG has

⁶ <https://nfdixcs.org/>.

⁷ <https://www.sfb1153.uni-hannover.de/en/crc-1153/profile>.

been successfully used for organizing scientific knowledge in different research fields, which substantiate our decision to use the ORKG. Finally, we introduce the CRC 1153 and its use case on the topic of TFPCs.

2.1 RESEARCH KNOWLEDGE GRAPHS

A research knowledge graph (RKG) represents scientific knowledge semantically, i.e., explicitly and formally, by linking existing (meta-)data of scientific artifacts (publications, software, and datasets) and entities (persons and organizations), which offers several benefits (Auer et al., 2018; Karras et al., 2021). According to Auer et al., 2020; the structured and semantic representation of scientific knowledge leads to *better identification, traceability, and reduced ambiguity* of concepts and relationships of scientific knowledge through terminological and conceptual clarity across disciplinary boundaries. These improvements, in turn, result in *easier reuse* of scientific knowledge and thus *less redundancy and duplication*. In addition, access to scientific knowledge is not only made easier for humans, but also for machines (artificial intelligence) by enabling *machine actionability*, as machines can grasp and understand the structure and semantics of scientific knowledge. This improved access enables far-reaching possibilities for the *development of novel digital services* in science, such as visualizations (Wiens et al., 2020) and question answering systems (Auer et al., 2023a; Jaradeh et al., 2020).

2.1.1 Exemplary RKGs

In the following, we present some exemplary RKGs. We do not claim that this overview is complete. Instead, it is primarily intended to provide readers with an initial basic orientation in the field of RKGs. Basically, RKGs are a relatively new research area and therefore their landscape in terms of underlying data models and technologies used is still diverse, varied, and non-standardized (Kejriwal, 2022). For this reason, there is no agreed common form for a RKG. However, there are ongoing efforts, such as the Research Data Alliance Open Science Graphs for FAIR Data Interest Group, to standardize the landscape of RKGs.⁸

Overall, there are two types of RKGs: *generic* and *specific* RKGs (Stocker et al., 2022). While generic RKGs are mainly based on bibliographic metadata, e.g., titles, authors, affiliations, citations, venues, and keywords, specific RKGs are mainly based on content data, e.g., scientific theories, claims, processes, measurements, and results.

Generic RKGs focus on bibliographic metadata of scientific artifacts and entities. There are several well-known generic RKGs, such as *Microsoft Academic Knowledge Graph*⁹ (Färber, 2019), *OpenAlex*¹⁰ (Priem et al., 2022), *Springer Nature SciGraph*¹¹ (Hammond et al., 2017), *Semantic Scholar Literature Graph*¹² (Ammar et al., 2018), *OpenAIRE Research Graph*¹³ (Manghi et al., 2019; Schirrwagen et al., 2013), *Research Graph*¹⁴ (Aryani and Wang, 2017), and *Scholarly Link Exchange (Scholix)*¹⁵ (Burton et al., 2017). All these RKGs have in common that they use bibliographic metadata to organize scientific artifacts, entities, and their relationships to enable, for example, their search, visualization, and processing (Brack et al., 2022; Stocker et al., 2022).

Specific RKGs focus mainly on content data combined with bibliographic metadata to describe and link scientific knowledge of the corresponding artifacts and entities. Specific RKGs are either specific to certain topics or more general to certain research fields. Some well-known examples of topic-specific RKGs are *CovidGraph*¹⁶ (Domingo-Fernández et al., 2020), *COVID-19 Air Quality Data Collection*¹⁷ (COVID-19 Air Quality Data Collection, 2021; Gkatzelis et al., 2021),

⁸ <https://www.rd-alliance.org/groups/open-science-graphs-fair-data-ig/>.

⁹ <https://makg.org/>.

¹⁰ <https://openalex.org/>.

¹¹ <https://scigraph.springernature.com/>.

¹² <https://semanticscholar.org>.

¹³ <https://graph.openaire.eu/>.

¹⁴ <https://researchgraph.org/>.

¹⁵ <https://scholia.toolforge.org/>.

¹⁶ <https://covidgraph.org/>.

¹⁷ <https://covid-aqs.fz-juelich.de/>.

and *SoftwareKG*¹⁸ (Schindler et al., 2020; Schindler et al., 2021; Schindler et al., 2022). The first two RKGs address COVID-19. However, the *CovidGraph* looks more generally at scientific knowledge on COVID-19 to explore publications, patents, existing treatments, and drugs around the coronavirus family (Domingo-Fernández et al., 2020). In contrast, the *COVID-19 Air Quality Data Collection* is even more fine-grained, focusing only on scientific content from publications about the impacts of COVID-19 lockdowns on air quality (COVID-19 Air Quality Data Collection, 2021; Gkatzelis et al., 2021). The *SoftwareKG* deals with the topic of software that is mentioned in scientific publications. This RKG is designed to enable users to explore and understand the role of software in science (Schindler et al., 2020; Schindler et al., 2021; Schindler et al., 2022). In addition to the topic-specific RKGs, there are several examples for domain-specific RKGs (Stocker et al., 2022), such as *Computer Science Knowledge Graph* (CS-KG)¹⁹ (Dessi et al., 2022), *Papers-with-Code*²⁰ (Papers With Code, 2020), *Hi Knowledge*²¹ (Jeschke et al., 2020), *Cooperation Databank* (CoDa)²² (Spadaro et al., 2022), and *OpenBiodiv*²³ (Penev et al., 2019). Each of these RKGs is dedicated to organize scientific knowledge from a specific research field such as computer science (Dessi et al., 2022), machine learning (Papers With Code, 2020), invasion biology (Jeschke et al., 2020), social sciences (Spadaro et al., 2022), and biodiversity (Penev et al., 2019).

2.1.2 Open Research Knowledge Graph

The previous examples give a first impression of the wide range of possible uses of RKGs in terms of research fields, topics, and contents. Due to this diversity and the lack of a comprehensive overview of available RKGs, it was difficult for us to find a suitable domain-specific or even topic-specific RKG for our use case in the engineering sciences. The ORKG appeared to be an interesting and promising solution, which we explain in more detail below.

The ORKG is continuously developed and maintained as one of the main services by the TIB - Leibniz Information Centre for Science and Technology—which has committed itself to long-term provision and maintenance of the ORKG. In contrast to the RKGs presented above, the ORKG²⁴ is both a cross-domain and cross-topic RKG. The ORKG provides an openly accessible, ready-to-use, sustainably governed infrastructure with services for the organization of any topic-specific scientific knowledge across all research fields following the FAIR data principles (Stocker et al., 2023). In this way, the ORKG offers all researchers the opportunity to organize, curate, and maintain any topic-specific scientific knowledge from any research field in the ORKG.

In particular, the ORKG organizes scientific knowledge provided by a publication as a collection of so-called ORKG contributions (D’Souza et al., 2024). An ORKG contribution consists of a semantic description of the scientific knowledge. Selected ORKG contributions can be compared in so-called ORKG comparisons (D’Souza et al., 2024). An ORKG comparison is a table where the columns denote the selected ORKG contributions by publication and the rows denote the semantically described scientific knowledge. ORKG comparisons can be enriched with so-called related resources and related figures. While related resources are supplementary materials such as Jupyter notebooks, related figures are additional visualizations to supplement an ORKG comparison. Figure 1 shows a snippet of an ORKG comparison of studies on Germany’s energy supply in 2050 (Karras et al., 2024; Kullmann et al., 2021). The snippet compares the targeted CO₂ reduction until the year 2050 of three publications.

In addition to the organizational and technical benefits described above, the ORKG is an integral part of the NFDI (Rossenova et al., 2023) in Germany. In particular, the four engineering consortia NFDI4Ing, NFDI4Energy, NFDI4DataScience, and NFDI4CS have committed themselves to use the ORKG for organizing scientific knowledge. In addition, all 26 consortia of the NFDI have

18 <http://softwarekg.git.gesis.org/softwarekg-site/>.

19 <https://scholkg.kmi.open.ac.uk/>.

20 <https://paperswithcode.com/>.

21 <https://hi-knowledge.org/>.

22 <https://cooperationdatabank.org/>.

23 <http://openbiodiv.net/>.

24 <https://orkg.org>.

Properties	Klimaneutrales Deutschland 2020 - Contribution	Wasserstoff-Roadmap Nordrhein-Westfalen 2020 - Contribution	Wege zu einem klimaneutralen Energiesystem 2020 - Contribution
hasgoal/goal			
↳ has_description*	100% CO2 reduction until 2050	95% CO2 reduction until 2050	95% CO2 reduction until 2050
↳ has_value*	100	95	95
↳ has_unit*	percent	percent	percent
↳ has_type*	CO2 reduction	CO2 reduction	CO2 reduction
↳ has_time_frame*	2050	2050	2050

Figure 1 Snippet of an ORKG comparison of targeted CO₂ reduction until 2050 (Karras et al., 2024; Kullmann et al., 2021).

launched a joint initiative called *Base Services for NFDI*²⁵ (Base4NFDI) to ensure the integration of services from individual consortia into the overall infrastructure and thus make them available in the NFDI and beyond. Part of Base4NFDI is a project called *Knowledge Graph Infrastructure* (KGI4NFDI). KGI4NFDI advocates for a central and reusable KG infrastructure to enhance the interoperability of (R)KGs. It aims to provide essential components, including a (R)KG registry and a service for accessing (R)KGs across the NFDI consortia. The ORKG is one of the selected RKGs considered by KGI4NFDI to be integrated into the overall KG infrastructure.

Due to all these reasons and especially the central role of the ORKG in the NFDI, we decided to use of the ORKG for our use case.

2.2 USE CASES OF THE ORKG

Several researchers have already reported on the successful use of the ORKG to organize scientific knowledge in specific research fields. In total, we found five publications reporting on nine use cases of the ORKG in the research fields: Numerical analysis and computation, software engineering, artificial intelligence, databases/information systems, virology, inorganic chemistry, molecular biology, and control theory.

In their recent publication on the ORKG, Auer et al. (2020) presented four different use cases of the ORKG in the research fields of artificial intelligence,²⁶ virology²⁷ (Oelen et al., 2020), inorganic chemistry,²⁸ and databases/information systems.²⁹ Each use case includes a published ORKG comparison consisting of six to 21 publications to illustrate how the ORKG can be used to organize scientific knowledge and provide an overview of the state-of-the-art in these research fields on a specific topic. These ORKG comparisons address different contents, such as the performance evaluation results of question answering systems, the reported basic reproduction numbers of Covid-19, the process parameters for the generation of silicon surface structures, and the properties of various RKGs. The use case on basic reproduction numbers of Covid-19 also shows how data in the ORKG can be further processed to gain new insights. In particular, Auer et al. (2020) show how the basic reproduction number of Covid-19 can be estimated using the reported values of individual international studies.

Runnwerth et al. (2020) used the ORKG to organize scientific knowledge about operational research in the research field of numerical analysis and computation. They organized scientific knowledge by adding six publications to the ORKG, using a developed and refined semantic structure to describe the operational research process. Runnwerth et al. (2020) described several publications, but they did not publish an ORKG comparison or further processed the scientific knowledge.

Anteghini et al. (2020) present an ongoing work to develop a semi-automated workflow to organize scientific knowledge from the research field of molecular biology in the ORKG. They

²⁵ <https://www.nfdi.de/base4nfdi/?lang=en>.

²⁶ <https://orkg.org/comparison/R6757/>.

²⁷ <https://orkg.org/comparison/R44930/>.

²⁸ <https://orkg.org/comparison/R41148/>.

²⁹ <https://orkg.org/comparison/R78023/>.

use the BioAssay Ontology (BAO) as a semantification model for representing bioassays. Using three manually semantified bioassays, they demonstrated their idea and created an ORKG comparison.³⁰ However, this ORKG comparison was not published, and the scientific knowledge was not further processed.

Karras et al. (2021) used the ORKG in a software engineering context. Based on two published systematic literature reviews (Khan et al., 2019; Santos et al., 2019), they published two ORKG comparisons (Karras and Groen, 2021; Karras and Khan, 2021). The first ORKG comparison focuses on qualitative scientific knowledge from 27 publications on the topic of crowd intelligence in requirements engineering (Karras and Khan, 2021). The second ORKG comparison deals with quantitative scientific knowledge from 19 publications on the topic of user feedback classification approaches (Karras and Groen, 2021). Both ORKG comparisons were created with the goal of organizing the detailed results of the systematic literature reviews so that the results are preserved for the long term by making them openly accessible and reusable.

Knoll (2022) examined the ORKG as a medium to facilitate the knowledge transfer in the research field of control theory. In particular, he investigated the current state of scientific knowledge about control theory in the ORKG and report on the process of extending this knowledge by adding new publications. For this purpose, Knoll (2022) organizes scientific knowledge from five publications in the ORKG. However, he has neither publish an ORKG comparison nor processed the data further.

All five publications demonstrate the applicability of the ORKG for organizing scientific knowledge in various research fields. However, only one use case by Auer et al. (2020) on the basic reproduction number of Covid-19 (Oelen et al., 2020) covered the entire process of communicating and (re-)using scientific knowledge. In agreement, the different authors conclude that the ORKG has the potential to become the next-generation digital library for fine-grained, semantic, and thus FAIR scientific knowledge. These successful use cases have further encouraged us to analyze the ORKG for organizing scientific knowledge in engineering sciences, a hitherto understudied but data-intensive and interdisciplinary research field where the importance of FAIR scientific knowledge is increasingly growing (Schmitt et al., 2020).

2.3 CRC 1153

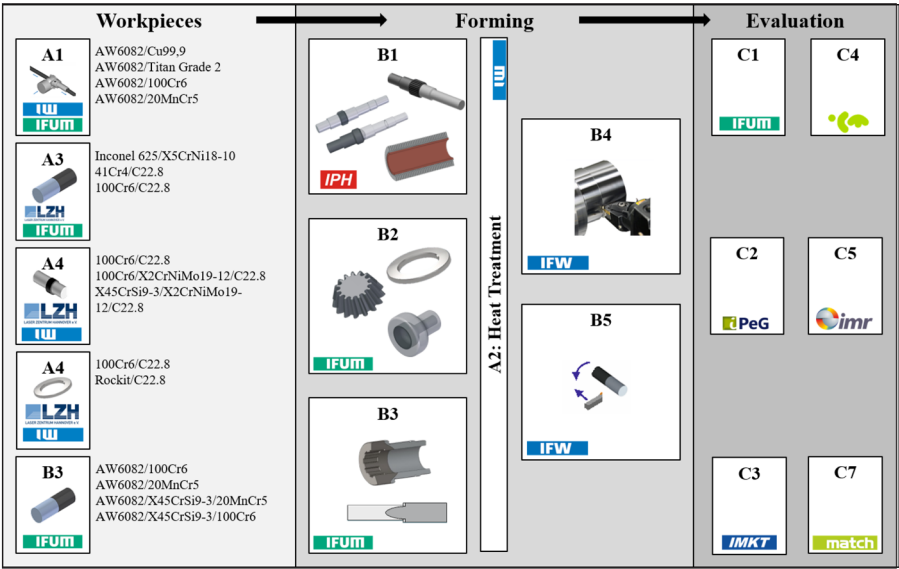
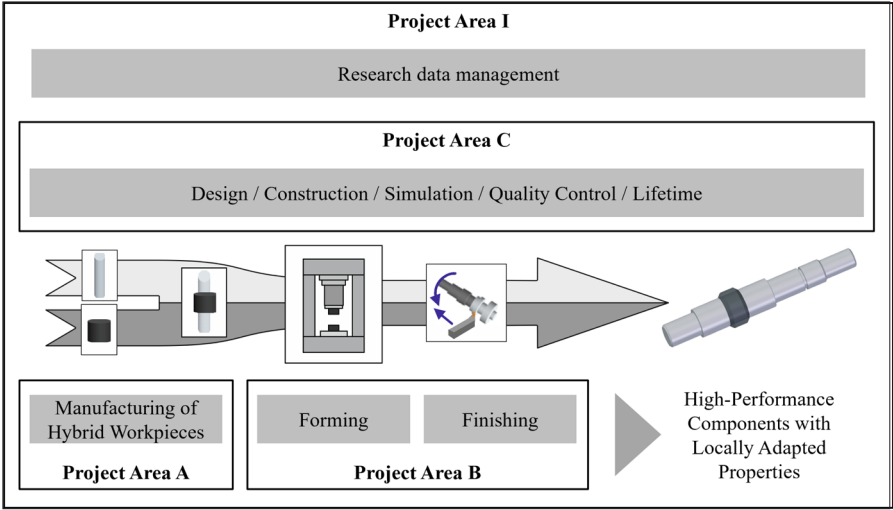
The Collaborative Research Center (CRC) 1153 is a long-term project that explores an innovative type of process chain called “Tailored Forming” for the manufacturing of hybrid solid components (Brockmüller et al., 2020). The demand for hybrid components is steadily increasing due to rising energy costs and growing material scarcity. In addition, stricter environmental regulations and higher requirements for strength, functional integration, and resource consumption can only be solved by combining different materials (Uhe and Behrens, 2019).

The CRC 1153 consists of four project areas A, B, C, and I with 20 sub-projects involving twelve different research institutes and 50 employees (see Figure 2). In project area A “Manufacturing of Hybrid Workpieces”, hybrid semi-finished workpieces are manufactured by means of different joining processes and heat treatments. The focus of the project area is on the selection of suitable material combinations and the ongoing technical development of joining methods. Project area B “Forming” is responsible for further processing the hybrid semi-finished workpieces. Major challenges are the precise shaping of complex geometries using materials with different flow properties and the joining zone quality. Project area C “Design and Evaluation” focuses on process design, extension, and development for the evaluation of the process chains, e.g., by failure predictions, modeling of joining zones, or rolling contact fatigue strength. In project area I “Research Data Management,” the research data management is realized by developing a technology-specific process model with the inclusion of corresponding parameters. Each project area, in turn, comprises several sub-projects which are worked on by employees from different research institutes, who have to cooperate closely. An overview of the individual sub-projects, the workpieces, and the process chains is given in Figure 3. In the sub-projects A1, A3, A4, and B3, hybrid workpieces are manufactured from different material combinations. These workpieces are then formed in one of the three sub-projects B1, B2, or B3, machined in one of sub-projects B4 and B5, and heat treated in sub-project A2. The evaluation of the process results is carried out in sub-projects C1, C2, C3, C4, and C5. Sub-project C7

30 <https://www.orkg.org/orkg/comparison?contributions=R48195,R48179,R48147>.

Figure 2 General process chain of CRC 1153 and its project areas.

Figure 3 General process chain of CRC 1153 and its sub-projects.



supports the process automation of the various sub-projects. The CRC 1153 is a typical project for the engineering sciences, as it is confronted with the challenges of communicating and (re-) using extensive and heterogeneous knowledge and data.

The employees of the CRC 1153 already published over 150 scientific publications, resulting from the long-term nature of the projects and its complexity due to the development of eleven demonstrators whose TFPCs have numerous adjustable parameters. Each adjustment leads to a new TFPC that may result in a new publication. The employees face the challenges of communicating and (re-)using the scientific knowledge of these publications as they lack a solid overview to track what has been done and what still needs to be done. For this reason, the CRC 1153, with its need for an overview of its own publications, provides a suitable use case for organizing knowledge in engineering sciences with the ORKG.

3 RESEARCH APPROACH

We first define the research goal and research question to ensure that the scope of the research approach is clearly defined before presenting its details. We defined the goal in detail using the goal definition template (Basili et al., 1994):

Goal definition:

We analyze the ORKG for the purpose of sustainable organization of FAIR scientific knowledge in engineering sciences with respect to communication and (re-)use from the point of view of engineers as ORKG users in the context of the use case on TFPCs from the CRC 1153.

Based on this goal, we ask the following research question:

Research question:

Based on the use case of TFPCs, how can engineers use the ORKG for the sustainable organization of FAIR scientific knowledge for communication and (re-)use in engineering sciences?

We frame our research approach using the design science paradigm (Runeson et al., 2020) (see Figure 4).

For the *solution design*, we collect publications from the field of mechanical process engineering and extract knowledge from the publications on the reported TFPCs. In this way, we build and publish an ORKG comparison for providing a detailed overview of the described knowledge, which we analyze to demonstrate how this knowledge can be (re-)used to gain new overviews and insights.

For the validation, we perform two activities: 1) We ask experts to assess the usefulness of the data extraction topics, their organization, and the relevance of the competency questions with their answers for engineering sciences; 2) We apply our solution design to publications in engineering sciences outside the CRC 1153 to show its applicability.

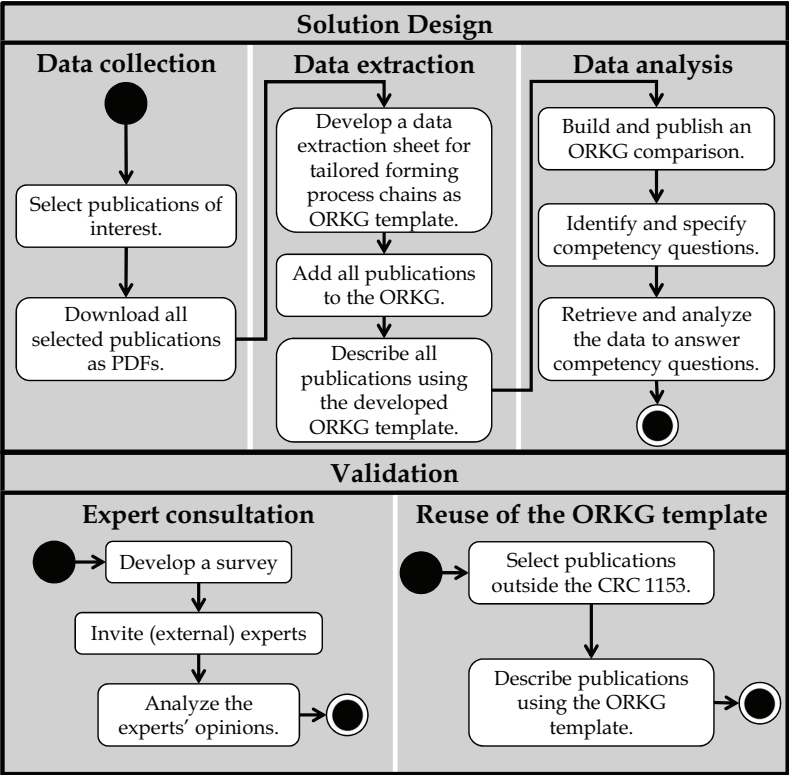


Figure 4 Activity diagram of the research approach.

3.1 SOLUTION DESIGN

The solution design of our research approach consists of the three steps data collection, data extraction, and data analysis (see Figure 4), which we explain in more detail below.

3.1.1 Data collection

Our use case is the first attempt to describe TFPCs semantically in the ORKG. For this reason, we focus on a smaller but complete set of five publications (Behrens et al., 2019; Budde et al., 2022; Coors et al., 2020; Kruse et al., 2019; Pape et al., 2018), all dealing with the specific topic of “Tailored Forming Process Chains for the Manufacturing of Hybrid Components with Bearing Raceways Using Different Material Combinations.” In particular, these five publications report on 10 TFPCs. We consciously decided to limit the number of publications to end up with a

more homogeneous set of them,³¹ which simplifies data extraction and semantic description while ensuring consistency and comparability of the described scientific knowledge. For data collection, we downloaded the selected publications as PDF files from the websites of the respective publishers.

3.1.2 Data extraction

Data extraction is the fundamental step in organizing scientific knowledge in the ORKG to build and publish an ORKG comparison and further analyze the knowledge. Based on the given use case on TFPCs and the selected publications, we had a clear idea about the topics for data extraction, including individual steps, their sequences, their incoming and outgoing components, the manufacturing methods, and the qualities investigated including the measurement methods used and the measurement results. Instead of a spreadsheet, we implemented the data extraction sheet as a so-called ORKG template (Hussein et al., 2023) to organize the scientific knowledge. In the following, we first explain the technical details of ORKG templates, followed by the development and application process of our ORKG template.

ORKG template system

The ORKG provides a template system for the development and application of ORKG templates. ORKG templates are based on the Shapes Constraint Language (SHACL) (Knublauch and Kontokostas, 2017) by implementing and adopting a subset for defining and enforcing data structures, formats, and constraints in the form of graph patterns with placeholders of specified types. Once an ORKG template is developed, it can be applied on ORKG contributions to define their graph structure. The ORKG parses the ORKG template and generates a corresponding input form including data validation based on the graph patterns. In this way, a publication can be described manually in the ORKG. However, it is also possible to use ORKG templates without the generated input form to programmatically import large amounts of data into the ORKG. During manual and programmatic data entry, the ORKG automatically validates the entries to ensure that the data structures, formats, and constraints of the ORKG template are met. This approach ensures data quality, integrity, and interoperability. As a result, the knowledge to be extracted can be specified, and it is ensured that all the publications described are comparable as they have the same structure.

Development

Based on a proposed initial version of a domain-specific ontology for Tailored Forming technology (Sheveleva et al., 2020), the first three authors of this article developed the ORKG template³² in three iterations over a total period of two weeks using the ORKG template system. When developing the ORKG template, we focused on a generic design to ensure its reusability. Starting from an initial draft, we applied the (revised) ORKG template to one randomly selected TFPC of a publication from data collection. Based on our experiences in data extraction, we continuously adapted the ORKG template and constantly revised the descriptions of all TFPCs described so far. After three iterations, there were no more changes. All remaining authors reviewed the final version of the ORKG template and confirmed its suitability for data extraction. Figure 5 shows an excerpt from the ORKG template to illustrate the structure for describing a TFPC reported in a publication. The ORKG template does not include bibliographic metadata, as the ORKG (semi-)automatically compiles the bibliographic metadata of a publication when the publication is added to the ORKG. For a complete overview of the ORKG template, refer to our supplementary materials³³ (Karras, 2024).

Application

For data extraction, the first three authors added the publications from the data collection to the ORKG using their Digital Object Identifiers (DOIs). In this way, the ORKG automatically

³¹ We want to emphasize that the selected publications are only a starting point for our work. We first develop an initial structure for the semantic description of TFPCs that is as generic as possible and then successively apply it to other CRC 1153 publications to expand it in the future.

³² <https://www.orkg.org/orkg/template/R150595>.

³³ <https://github.com/okarras/TailoredForming-Analysis/blob/master/Supplementary%20materials/Detailed%20ORKG%20template%20structure.png>.

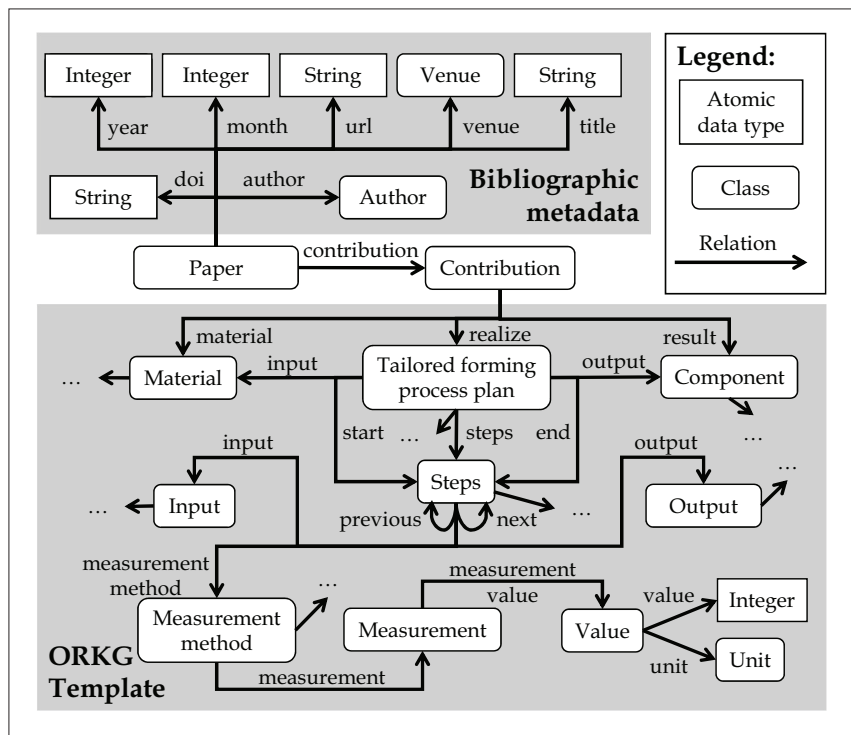


Figure 5 Excerpt from the ORKG template for data extraction.

compiles the bibliographic metadata of the publications. For each publication, they created an ORKG contribution for each TFPC and applied the developed ORKG template to it. Each TFPC was described jointly by two of the first three authors. If the information was available, they extracted it manually from the publication and entered it into the ORKG using the generated input form (see Figure 6). They always adhered to the terminology used in the publications to ensure an accurate and consistent description of the corresponding knowledge. The remaining author of the three reviewed each description by comparing the described knowledge with the respective publication. In case of inconsistencies or ambiguities, all three authors discussed and resolved the issues identified. Figure 7 shows a concrete example for one TFPC of a selected publication by Budde et al. (2022) described in the ORKG.³⁴ Overall, this publication (Budde et al., 2022) reports on four TFPCs for manufacturing hybrid solid components, that are all described in the ORKG based on the ORKG template.

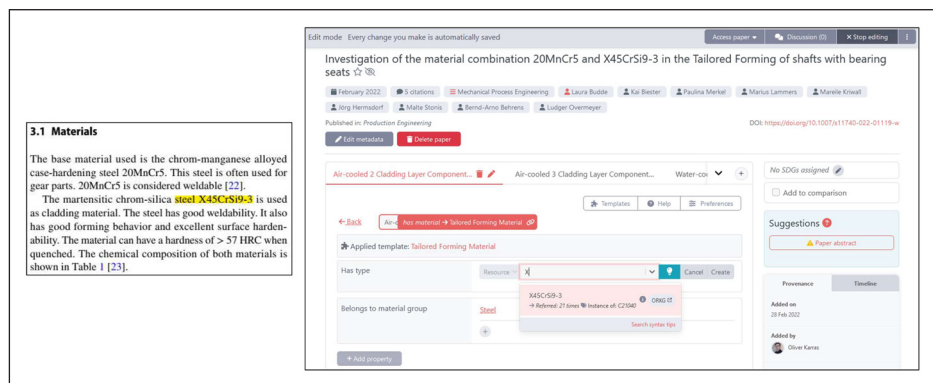


Figure 6 Data extraction example: On the left, a paragraph of the publication by Budde et al. (2022) about the materials used in the TFPCs. We highlighted the material to be described in yellow. On the right, we show the input form of the ORKG generated based on our ORKG template for describing materials.

3.1.3 Data analysis

The data analysis serves to demonstrate how engineers can utilize the ORKG in innovative ways for *communication* and *(re-)use* of scientific knowledge in the field of engineering sciences. First, we build and publish an ORKG comparison to provide a detailed overview of the described knowledge about the 10 TFPCs for *communication*. Second, we analyze the scientific knowledge from the publications to show how its *(re-)use* can lead to new overviews and insights.

³⁴ <https://orkg.org/paper/R171846/R171849>.

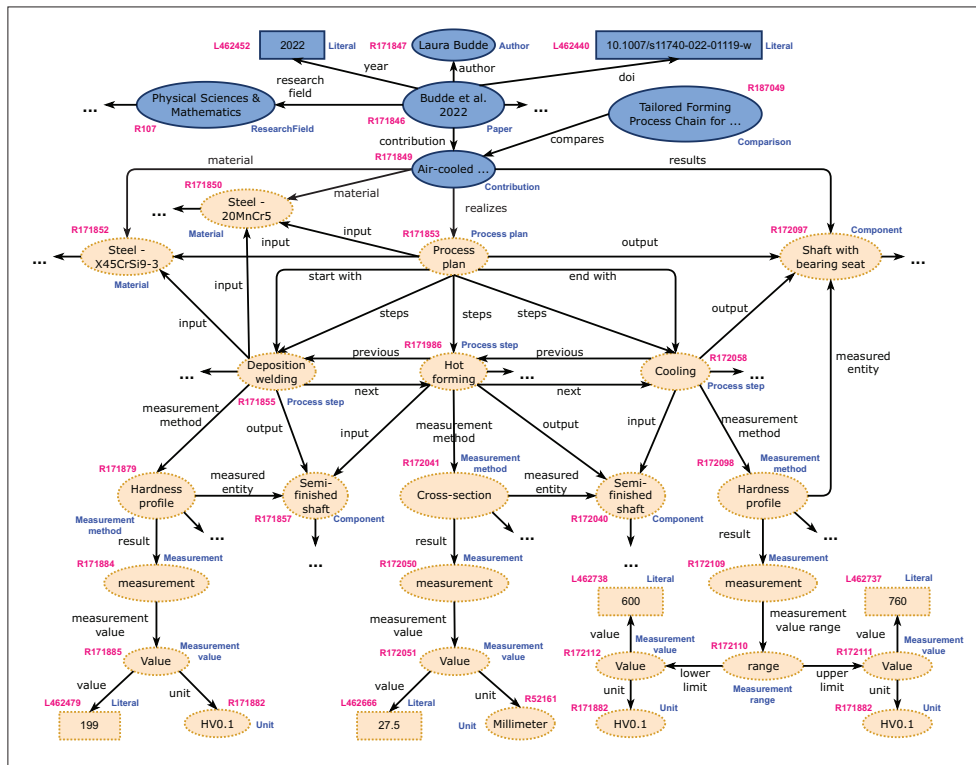


Figure 7 Example from selected publications (Budde et al., 2022) in the ORKG: Blue shapes represent ORKG resources and predicates. Yellow shapes are user-generated. Class labels are in blue, and resource IDs are in magenta.

Competency questions are an established method for analyzing and evaluating KGs and their contents (Grüniger and Fox, 1995). A competency question is a natural language question that represents an information need related to the content of a KG and for which a KG must provide relevant information (Hogan et al., 2021). Based on the collected publications, we asked two domain experts, who only knew the topics of the data extraction, to ask competency questions. The two domain experts came up with eight competency questions that we should be able to answer using the described knowledge in the ORKG (see Table 1).

ID	COMPETENCY QUESTION
1	How often are which qualities of hybrid components examined in the individual steps of the Tailored Forming Process Chain?
2	Which material combinations are used for the manufacturing of hybrid components?
3	How do the steps of the Tailored Forming Process Chain affect the hardness of the cladding layer of hybrid components?
4	In which material combinations and steps of the Tailored Forming Process Chain can defects occur in hybrid components?
5	What service lives can be achieved for hybrid components in the Tailored Forming Process Chain?
6	Which manufacturing methods are used in the individual steps of the Tailored Forming Process Chain to manufacture the respective hybrid component?
7	How often are certain manufacturing methods used in total in all described Tailored Forming Process Chains?
8	Which microstructures are present in hybrid components after the steps of the Tailored Forming Process Chain?

Table 1 The eight competency questions asked by the two domain experts.

For data analysis, we retrieve the scientific knowledge from the ORKG to answer the competency questions using a query specified with the query language SPARQL (Seaborne and Harris, 2013). We used SPARQL as the ORKG provides a SPARQL endpoint³⁵ for accessing all its contents. In Listing 1, we provide an exemplary SPARQL query to retrieve the data for answering the fifth competence question (cf. Table 1). For the SPARQL query used to retrieve all data, refer to our supplementary material (Karras, 2024). We implemented the analysis

35 <https://www.orkg.org/sparql>.

```

1 PREFIX orkg: <http://orkg.org/orkg/resource/>
2 PREFIX orkgc: <http://orkg.org/orkg/class/>
3 PREFIX orkgp: <http://orkg.org/orkg/predicate/>
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
6
7 #This query can be executed in the visual SPARQL editor of ORKG: https://www.orkg.org/sparql.
8 #You must delete line 9 if you only want to see the results table rather than the bar chart.
9 #defaultView:BarChart
10 SELECT (STR(?orkg_contribution) AS ?paper) ?measurement_value ?paper
11 WHERE {
12
13   #Get all ORKG contributions from the CRC 1153 papers that use our ORKG template
14   ?orkg_contribution a orkgc:C21027;
15                     orkgp:P32 ?research_problem.
16   FILTER(?research_problem = orkg:R720788)
17
18   #Get all "Fatigue test" measurements of each step of all tailored forming process chains
19   ?orkg_contribution orkgp:P7108 ?tfpc.
20   ?tfpc orkgp:P43070 ?step.
21   ?step orkgp:P44012 ?measurement_method.
22   ?measurement_method orkgp:P7111 ?measurement_method_name;
23                     orkgp:HAS_RESULTS ?measurement_result.
24   ?measurement_method_name rdfs:label ?measurement_method_name_label.
25   FILTER(?measurement_method_name_label = "Fatigue_test"^^xsd:string)
26
27   #Get all measurement values and units of all "Fatigue test" measurements
28   ?measurement_result orkgp:HAS_VALUE ?measurement.
29   ?measurement orkgp:P44011 ?measurement_value;
30                     orkgp:P42000 ?measurement_unit.
31   ?measurement_unit rdfs:label ?measurement_unit_label.
32 }

```

Listing 1 Exemplary SPARQL query to retrieve the data for answering the fifth competence question on the service lives achieved for hybrid components in the Tailored Forming Process Chain (cf. Table 1).

using a Jupyter notebook with Python, which is published on GitHub³⁶ (Karras, 2024). We also host the analysis on mybinder.org.³⁷ In this way, the analysis is always accessible to anyone for interactive reproduction and (re-)use, retrieving the latest data from the ORKG. Due to the uniform structure provided by the developed ORKG template, we can always retrieve newly added publications that use the ORKG template and include them in the analysis by simply running the script again.

3.2 VALIDATION

The validation consists of an expert consultation and the reuse of the ORKG template on publications outside the CRC 1153 (see Figure 4). Below, we present the two parts of the validation in more detail.

3.2.1 Expert consultation

The expert consultation serves to examine the usefulness of the data extraction topics, their organization in the ORKG template, and the relevance of data analysis results (competency questions and answers with visualizations) for engineering sciences.

Survey design

For the expert consultation, we developed a survey following the process for conducting a questionnaire survey by Robson and McCartan (2016). A survey is a suitable research method for collecting information to analyze opinions (Kitchenham and Pfleeger, 2008).

We developed the survey instrument using Google Forms.³⁸ The initial questionnaire consisted of four parts to collect demographics and to get the respondents' opinions on the data extraction topics, their organization, and the data analysis results. For the demographics, we asked the respondents in which subject area of engineering sciences they are a professional, how many years of professional experience they have in engineering sciences, and what their highest level of education in engineering sciences is. For the assessment of the data extraction topics, we used a 5-point Likert scale for each of the eleven topics extracted to assess the respondents' level of agreement with the statement "The topic [...] is useful for describing process chains in engineering sciences." For the assessment of the topic organization, we have provided a simplified overview of the structure of the ORKG template and have formulated the structure as a statement for each topic (see Table 2). Using 5-point Likert scales, the respondents indicated

³⁶ <https://github.com/okarras/TailoredForming-Analysis>.

³⁷ [https://mybinder.org/v2/gh/okarras/TailoredForming-Analysis/HEAD?labpath=%2Ftailored`forming`analysis.ipynb](https://mybinder.org/v2/gh/okarras/TailoredForming-Analysis/HEAD?labpath=%2Ftailored%20forming%20analysis.ipynb).

³⁸ <https://www.google.de/intl/de/forms/about/>.

TOPIC	STATEMENT
Material	A material has one or more qualities that are measured.
Process chain	A process chain has two or more process steps, a start process step, an end process step, an input, and an output.
Process step	A process step can have a previous process step and a next process step, and it has an input, an output, one or more manufacturing methods, and one or more measurement methods.
Input	An input is either one or more materials or the output of the previous process step.
Output	An output is a manufactured component.
Manufacturing method	A manufacturing method is the method used in a process step to manufacture the output of the process step.
Manufactured component	A manufactured component has one or more layers.
Layer	A layer consists of a material.
Quality	A quality is measured by one or more measurements.
Measurement method	A measurement method has an output of a process step as the object of investigation and one or more measurements.
Measurement	A measurement has a value and a unit.

Table 2 Statements describing the organization of the data extraction topics assessed by the respondents.

their level of agreement with each of the 11 formulated statements. For the assessment of the data analysis results, we presented each competency question and its answer with a visualization. For the competency question and the answer (including the visualization), the respondents indicated their level of agreement with the statement “[...] is relevant for Tailored Forming Process Chains in engineering sciences” using a 5-point Likert scale.

Pre-tests

We conducted three rounds of pre-tests with three experts from engineering sciences who are not members of the CRC 1153. In each pre-test, one expert completed the survey, and we discussed how the questionnaire could be improved. Besides improvements to wording and presentation, the main finding was that the experts found it difficult to evaluate the data analysis results. All three experts attributed this difficulty to a lack of domain knowledge regarding Tailored Forming. For this reason, we created two versions of the questionnaire. The first version contains all four parts of the improved questionnaire and is aimed at experts in the CRC 1153 who are not authors of this article. These experts have the required domain knowledge to assess the results of the data analysis. The second version is aimed at external experts in the engineering sciences and therefore includes all parts except for the assessment of the data analysis results. We provide the two final questionnaires and the collected raw data in our supplementary materials on GitHub³⁹ (Karras, 2024).

Sampling

For the sampling, we applied purposive sampling, which is a non-probabilistic sampling technique that is most effective when experts are needed as subjects (Tongco, 2007). We contacted 15 external experts in engineering sciences and 10 experts from the CRC 1153. These 25 experts were invited to participate in the survey. The survey started on August 6th, 2024, and ended on August 31st, 2024.

3.2.2 Reuse of the ORKG template

The reuse of the ORKG template is intended to show the applicability of our solution design to publications outside the CRC 1153 that deal with manufacturing process chains.

For the reuse, we focused on Additive Manufacturing Process Chains (AMPCs). AMPCs encompass all manufacturing process chains in which a workpiece is built up layer by layer by adding one or more materials. AMPCs are an interesting area of application for the ORKG template, as they have a number of similarities with TFPCs. Both types of process chains consist of several

³⁹ <https://github.com/okarras/TailoredForming-Analysis>.

steps and aim to optimize the manufacturing of customized components made of one or more materials with specific qualities. We randomly selected three publications on AMPCs (Papke et al., 2018; Papke et al., 2023; Schröcker et al., 2022) outside the CRC 1153 which report a total of 20 AMPCs. These publications examine the effects of process steps on additively manufactured components, such as the hardness in the outermost layer.

For the data extraction, we applied the same procedure as for the TFPCs (cf. Application.). The first three authors added the publications to the ORKG and applied the ORKG template for each of the 20 AMPCs. While two authors jointly described the AMPCs, the remaining author reviewed the descriptions. All three authors discussed and resolved any issues identified.

After explaining the underlying research approach, we first present the results of the solution design (cf. section 4) and then the results of the validation (cf. section 5).

4 RESULTS: SOLUTION DESIGN

Below, we show the results of the data analysis from the solution design. First, we present our published ORKG comparison as an innovative way to communicate the described knowledge about the 10 TFPCs. Second, we *analyze* the scientific knowledge from the five publications to show how its (re-)use can lead to new overviews and insights by answering the competency questions.

4.1 ORKG COMPARISON

ORKG comparisons are one of the core artifacts of the ORKG. They provide a detailed overview for communicating scientific knowledge on the state-of-the-art for a particular research topic, problem, or question. In our case, we built and published an ORKG comparison of the 10 TFPCs from the five publications (Karras et al., 2023a). We also enriched the ORKG comparison with supplementary materials, i.e, the Jupyter notebook from our data analysis and the visualizations created with it, to provide the employees of the CRC 1153, but also every ORKG user, central access to all materials. Figure 8 shows the published ORKG comparison.

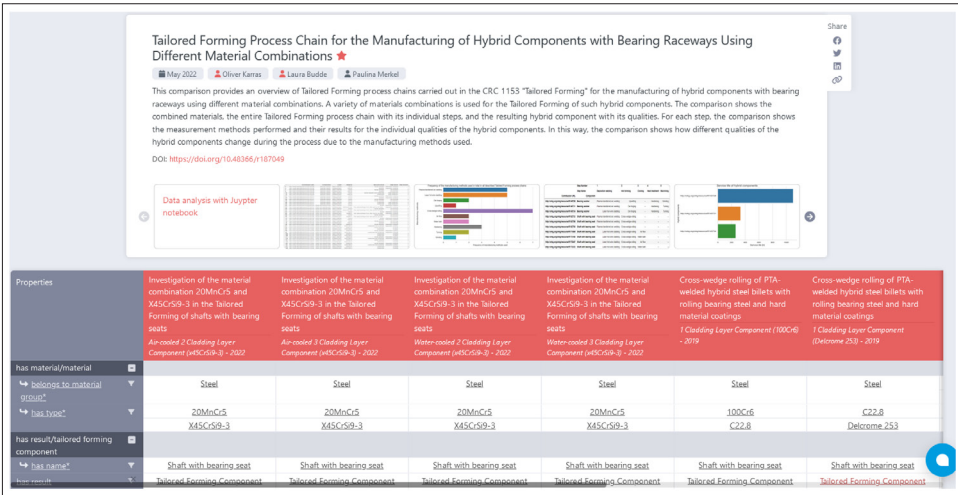


Figure 8 Comparison of TFPCs for the manufacturing of hybrid components (Karras et al., 2023a).

The ORKG comparison consists of typical metadata, e.g., title, publication month and year, authors, a description, a DOI, related resources and related figures, as well as the scientific knowledge of the 10 described TFPCs in tabular form (cf. section 2.1). This tabular form is interactive, allowing navigation and filtering of its content for exploration and more detailed consideration. For example, the first two rows show the materials used regarding their material groups and specific types across all 10 TFPCs. If we filter the second row for the material type “C22.8,” we obtain an updated tabular form containing only all described TFPCs that use this specific material type. Everyone can also (re-)use the underlying data of ORKG comparisons, e.g., in a Jupyter notebook, as the ORKG provides various interfaces such as a REST API, a Python package, and a SPARQL endpoint to programmatically access all its contents. In our case, we used the SPARQL endpoint for data analysis to answer the eight competency questions. For a better understanding of the contents of the ORKG comparison, we added visualizations as supplementary materials resulting

from our data analysis. We also added the Jupyter notebook, that is hosted on mybinder.org,⁴⁰ as supplementary material. While our published ORKG comparison is stable and accessible over the long term through its DOI, ORKG comparisons are versionable so that they can be continuously (re-)used, updated, and expanded. When employees of CRC 1153 publish a new publication regarding the use case of this article, they can easily update the ORKG comparison. They only need to describe the new TFPC using the developed ORKG template, add the new ORKG contribution to the ORKG comparison, publish the updated ORKG comparison as a new version, and execute the Jupyter notebook to update the visualizations.

4.1.1 Analysis of the use cases

We have used all features of ORKG comparisons to achieve the maximum benefit for successful and good communication of the described scientific knowledge about the 10 TFPCs from the five publications. Despite our conscious decision to limit the number of publications to obtain a more homogeneous set of publications, the question arises of how comparable our use case is with other use cases. For this reason, we analyze our use case and the nine use cases from the related work (cf. section 2.2) regarding their complexity in terms of the number of statements, resources, and literals. In addition, we also collect information about the number of analyzed papers, described ORKG contributions, related resources, and related figures. However, these results are limited due to the low number of use cases. Therefore, we extend the analysis by also including all 507 ORKG comparisons published with a DOI, as these ORKG comparisons are intended to be cited as use cases in corresponding publications. We provide the entire data analysis of the complexity as an interactive Jupyter notebook⁴¹ (Karras, 2024).

4.1.2 Findings on the use cases

Table 3 provides an overview of the descriptive statistics of our analysis, which are partially visualized in Figure 9. While the nine use cases from related work include on average 13.33 ± 8.04 , [3–27] papers and 14.44 ± 9.57 , [3–31] contributions, the 507 comparisons with DOI include on average 9.94 ± 12.43 , [1–86] papers and 12.85 ± 19.11 , [2–157] contributions. When we compare these numbers with our use case (5 papers and 10 contributions), we see that our use case is smaller in both regards, but close to the median. In Figure 9, we show boxplots of the complexity of the 507 ORKG comparisons with DOI, the nine use case of related work (×-symbols), and our use case (✖-symbol). We observe that the nine use cases of related work are distributed across the entire range of the respective boxplot for the number of statements, resources, and literals. While our use case is in the upper 75% percentile for the number of literals, it is even a potential outlier with a significantly higher number of statements and resources. These results are positive indicators for a comprehensive description of the scientific knowledge resulting from the detailed structure of our ORKG template. We have many statements that consist mainly of resources and some literals. Basically, a more extensive use of resources instead of literals is desirable, as more precise and fine-grained semantic descriptions are possible.

Despite the smaller number of papers and contributions, our use case already has a higher level of complexity and depth of description than the nine use cases from related work and the majority of the 507 ORKG comparisons with DOI. Our use case is sufficiently extensive and complex for a first attempt to describe TFPCs semantically in the ORKG. For this reason, we conclude that our use case represents a comprehensive example, which is not only comparable with other use cases, but already exceeds most of them in complexity.

4.2 COMPETENCY QUESTIONS

In the following, we report on our data analysis to answer the eight competency questions asked by two domain experts. The purpose of this analysis is to demonstrate how engineers can utilize the ORKG with its services for the (re-)use scientific knowledge. For this reason, we do not present the answers to all competency questions in detail. Instead, we focus overall on the analysis and coverage of the competency questions, which we illustrate with four examples. In this way, we want to encourage other engineers and empower them to apply our approach to their research.

40 <https://mybinder.org/v2/gh/okarras/Jupyter-Notebooks/HEAD?labpath=%2FTailoredFormingAnalysis%2Ftforkg.ipynb>.

41 <https://mybinder.org/v2/gh/okarras/TailoredForming-Analysis/HEAD?labpath=%2Fcomparison%2Fanalysis.ipynb>.

PUBLICATION	#PAPERS	#CONTRIB.	#STATEMENTS	#RESOURCES	#LITERALS	#RELATED RESOURCES	#RELATED FIGURES
Our use case							
This article	5	10	1918	3662	174	1	8
Nine use case from related work							
Runnwerth <i>et al.</i> (2020)	6	6	50	100	0	0	0
Karras <i>et al.</i> (2021)	19	19	609	820	398	0	0
Karras <i>et al.</i> (2021)	27	27	503	683	323	0	0
Auer <i>et al.</i> (2020)	17	17	202	382	22	0	0
Auer <i>et al.</i> (2020)	21	31	415	568	262	2	2
Auer <i>et al.</i> (2020)	16	16	127	159	95	0	0
Auer <i>et al.</i> (2020)	6	6	54	67	41	0	0
Anteghini <i>et al.</i> (2020)	3	3	100	199	1	0	0
Knoll (2022)	5	5	46	76	16	0	0
Mean	13.33	14.44	234	339.33	128.67	0.22	0.22
Std. deviation	8.04	9.57	204.88	270.03	146.75	0.63	0.63
Median	16	16	127	199	41	0	0
All 507 ORKG comparisons with DOI							
Mean	9.94	12.85	343.11	514.48	171.74	0.01	0.02
Std. deviation	12.43	19.11	589.59	882.18	352.87	0.11	0.38
Minimum	1	2	6	12	0	0	0
Median	6	7	157	234	80	0	0
Maximum	86	157	6089	9503	4842	2	8

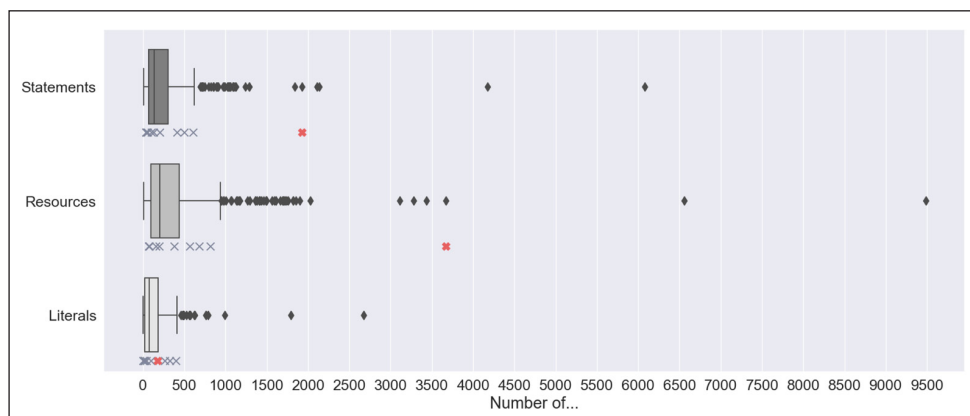


Table 3 Complexity of our use case, of nine use cases from related work, and of 507 ORKG comparisons with DOI.

Figure 9 Boxplots of the complexity of 507 ORKG comparisons with DOI, the nine use cases from related work and our use case. Remark: The $\times$ -symbol marks use cases from related work. The $\times$ -symbol marks the use case of this article. The $\blacklozenge$ -symbol marks potential outliers that are more than 1.5 times the interquartile range above the upper quartile.

4.2.1 Coverage

A competency question is a natural language question that represents an information need related to the content of a KG and for which a KG must provide relevant information (cf. section 3.1.3). In this context, no statement is deliberately made about the completeness of the relevant information provided, as the assessment of completeness is a complex, challenging, and non-trivial task that requires knowledge about the hypothetical ideal KG which usually does not exist (Darari *et al.*, 2018). Therefore, answering a competency question is fundamentally considered binary, so in terms of coverage, we want to know whether an answer can be given by providing relevant information that meets the corresponding information need.

We provide the entire data analysis for all eight competency questions with their answers and corresponding visualization as an interactive Jupyter notebook⁴² (Karras, 2024). Table 4

⁴² <https://github.com/okarras/TailoredForming-Analysis>.

COMPETENCY QUESTION ID	QUESTION ANSWERED?	RELEVANT INFORMATION FROM	
		#TFPCs	#PUBLICATIONS
1	Yes	10	5
2	Yes	10	5
3	Yes	10	5
4	Yes	6	3
5	Yes	3	3
6	Yes	10	5
7	Yes	10	5
8	Yes	9	4

Table 4 Coverage of the eight competency questions, with details on the number of TFPCs and publications that provide relevant information to answer the questions and thus meet the information needs.

summarizes the results for the coverage of the eight competency questions, including details on the number of TFPCs and publications that provide relevant information to answer the questions and thus meet the corresponding information needs. The results show that we answer all eight competency questions and thus achieve a complete coverage. We presented the answers with the associated visualizations from our data analysis to the two domain experts to verify our results. Both domain experts confirmed that our answers and visualizations provide relevant information to the respective questions and thus meet the corresponding information needs.

Overall, we answer the majority of the questions using information from all 10 TFPCs of all five publications and, on average, from 8.5 TFPCs (median: 10) of 4.4 publications (median: 5). In particular, the answers to five questions (1, 2, 3, 6, and 7) use information from all 10 described TFPCs of all five publications. While one question (8) is answered using information from nine TFPCs of four publications, only two questions (4 and 5) are answered using information from six and three TFPCs of three publications.

A closer look at the competency questions reveals that the questions 1, 2, 6, and 7, which use information from all the TFPCs described, are more general as they cover broader topics such as qualities, material combinations, and manufacturing methods. In contrast, the questions 3, 4, 5, and 8 are more specific covering specific qualities such as hardness, defects, service lives, and microstructures. While all publications always report on the general topics, the reporting of specific topics varies. Thus, the publications overlap significantly in terms of information on general topics of TFPCs, but may differ in terms of details on specific topics. In the following section, we present four examples of the answers to two general competency questions (1 and 6) and two specific competency questions (3 and 5).

4.2.2 Examples

The use case of this article is the first attempt to describe TFPCs semantically in the ORKG. For this reason, we cannot provide an analysis with comprehensive and general results, findings, and insights on TFPCs from the CRC 1153. Nevertheless, we show some results for four exemplary competency questions. These examples illustrate the previous observations regarding difference in the reporting on general and specific topics, as well as how the (re-)use of scientific knowledge can lead to new overviews and insights.

CQ1: How often are which qualities of hybrid components examined in the individual steps of the Tailored Forming Process Chain?

Figure 10 shows the frequency of the examined qualities in all 10 described TFPCs per step. In total, the 10 TFPCs report 14 different qualities. The two qualities *hardness* and *microstructure* are examined most frequently in four out of five steps. The machining step is the only step where no hardness measurements and microstructure analyses are reported, as no relevant changes in these qualities are expected in this step. Furthermore, we observe that several qualities, such as Weibull shape parameter, experimental life, coefficient of friction, etc., are only measured in the machining step. These qualities are related to the evaluation of the characteristics of the bearing running surface, which can only be examined after the last step, as this step leads to the final running surface of the rolling elements.

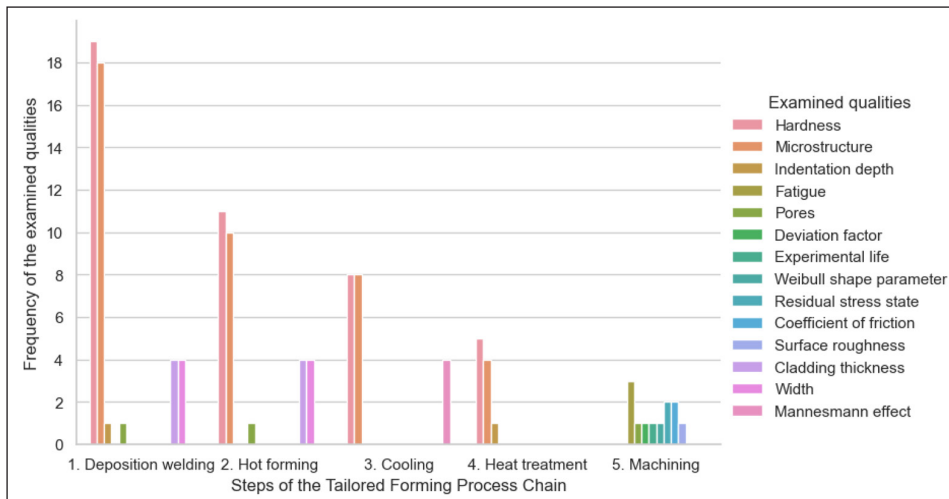


Figure 10 Frequency of examined qualities of hybrid components per step of the TFPC.

CQ3: How do the steps of the Tailored Forming Process Chain affect the hardness in the cladding layer of hybrid components?

In Figure 11, we show how the hardness of the cladding layer changes in the steps due to the manufacturing methods used. While the publications report the individual values in various places, none of the publications contain a diagram of the change in the hardness along the TFPC steps similar to the one we created. Figure 11 makes it possible to compare the different TFPCs and see how they affect the hardness. Based on Figure 11, we observe that none of the publications reports the hardness in all steps of the TFPC. While all publications report the hardness in the deposition welding step and six publications in the hot forming step, only four publications report the hardness in the cooling step and three publications in the heat treatment step. No publication reports the hardness after the machining step, as machining does not influence the hardness. In the selected publications, the authors usually report the hardness in the unit Vickers Pyramid Number (HV). Only in one publication, the authors used the unit Gigapascal (GPa). The heat treatment led to an increase in the hardness in all reported cases. The influence of hot forming on hardness depends on the cladding layer material and the specific manufacturing method used for hot forming. Hot forming can lead to an increase or a decrease in hardness compared to the state after deposition welding.

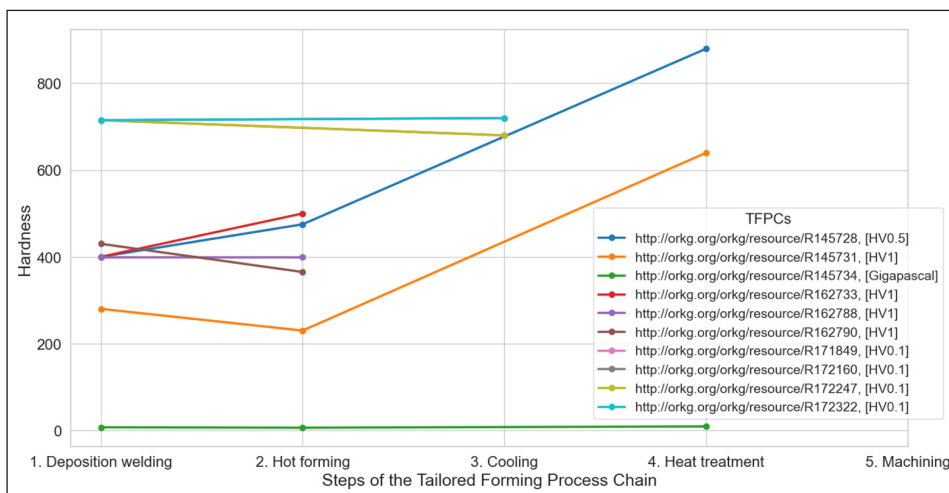
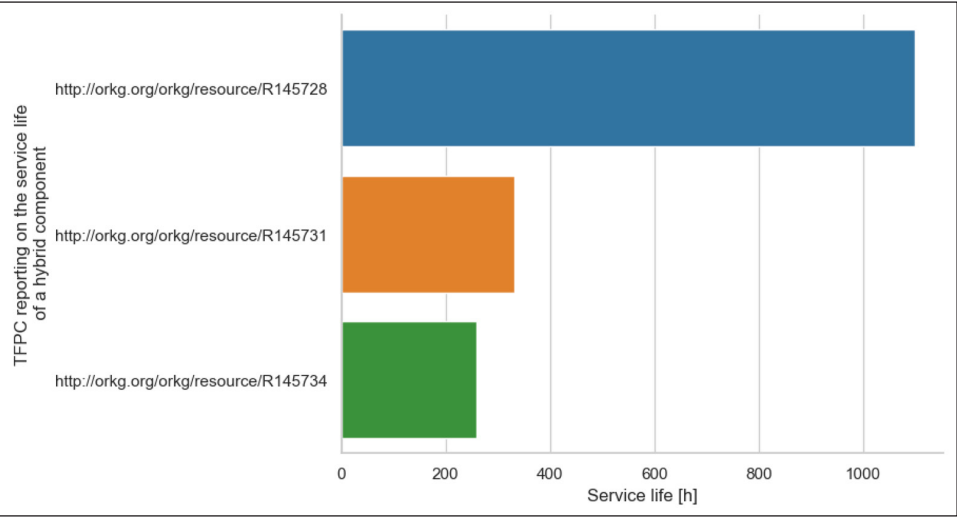


Figure 11 Hardness of the cladding layer of hybrid components per step of the TFPC.

CQ5: What service lives can be achieved for hybrid components in the Tailored Forming Process Chain?

Figure 12 presents the reported service lives of different hybrid components. In total, only three out of the 10 TFPCs report a service life for the manufactured hybrid component, which substantiates that this information and question are specific. We observe that there is one TFPC that has led to more than three times higher service life for the manufactured hybrid component than the other two TFPCs. This finding highlights the usefulness of this type of

Figure 12 Service lives of hybrid components manufactured in TFPCs.



analysis, as it helps not only to compare TFPCs with each other, but also to find TFPCs with specific results, such as a manufactured hybrid component with a long service life. Based on this finding, researchers can take a closer look at the associated publication to better understand how this TFPC could lead to this long service life.

CQ6: Which manufacturing methods are used in the individual steps of the Tailored Forming Process Chain to manufacture the respective hybrid component?

For each of the 10 TFPCs, Figure 13 shows the hybrid component and the manufacturing methods used in each step of the respective TFPC. While all publications report some details on the manufacturing methods used, none of them provide details on all steps of the TFPC. First, we observe that a manufacturing method is always reported for the deposition welding and hot forming step. In particular, plasma-transferred arc welding and laser hot-wire cladding are each used five times as a manufacturing method for deposition welding. Three different manufacturing methods are used for hot forming, with cross-wedge rolling being used most frequently (seven times). For the subsequent steps of the TFPCs, information about the manufacturing method used is often missing in the associated publications. After consulting with the domain experts who asked the competency questions, we found that they interpreted the missing data as follows: 1) A missing manufacturing method for the cooling step is interpreted as air cooling, and 2) A missing manufacturing method for the subsequent heat treatment and machining steps are interpreted to mean that the steps were not performed. This interpretation is critical as it can lead to incorrect conclusions. As the data is stored in the ORKG, we follow the open-world assumption, which means that missing information is interpreted as unknown information and not as negative information, such as steps not performed (Keet, 2013). For this reason, the domain experts’ interpretation of the missing data is not well-founded, as there is no evidence in the publications and therefore may lead to incorrect conclusions. This way of interpreting missing data shows how important it is for engineers to document all steps of an engineering (research) process in detail to guarantee the trustworthiness of published results and their reproducibility.

	Step number	1	2	3	4	5
	Step name	Deposition welding	Hot forming	Cooling	Heat treatment	Machining
Tailored Forming Process Chain (TFPC)	Hybird component					
http://orkg.org/orkg/resource/R145728	Bearing washer	Plasma-transferred arc welding	Upsetting	--	Hardening	Grinding
http://orkg.org/orkg/resource/R145731	Bearing washer	Plasma-transferred arc welding	Die forging	--	Hardening	Turning
http://orkg.org/orkg/resource/R145734	Bearing washer	Laser hot-wire cladding	Die forging	--	Hardening	Turning
http://orkg.org/orkg/resource/R162733	Shaft with bearing seat	Plasma-transferred arc welding	Cross-wedge-rolling	--	--	--
http://orkg.org/orkg/resource/R162788	Shaft with bearing seat	Plasma-transferred arc welding	Cross-wedge-rolling	--	--	--
http://orkg.org/orkg/resource/R162790	Shaft with bearing seat	Plasma-transferred arc welding	Cross-wedge-rolling	--	--	--
http://orkg.org/orkg/resource/R171849	Shaft with bearing seat	Laser hot-wire cladding	Cross-wedge-rolling	Air flow	--	--
http://orkg.org/orkg/resource/R172160	Shaft with bearing seat	Laser hot-wire cladding	Cross-wedge-rolling	Water bath	--	--
http://orkg.org/orkg/resource/R172247	Shaft with bearing seat	Laser hot-wire cladding	Cross-wedge-rolling	Air flow	--	--
http://orkg.org/orkg/resource/R172322	Shaft with bearing seat	Laser hot-wire cladding	Cross-wedge-rolling	Water bath	--	--

Figure 13 Manufacturing methods per step of the TFPCs.

As mentioned before, these four examples serve to illustrate the previously presented observations from section 4.2.1. We provide the entire analysis of all eight competency questions with visualizations and explanations as an interactive Jupyter notebook, that is hosted on mybinder.org.⁴³

5 RESULTS: VALIDATION

Below, we show the results of the validation. First, we present the results of the expert consultation, followed by the results of the reuse of the ORKG template on publications outside CRC 1153.

5.1 EXPERTS’ OPINION

We report the results according to the structure of the survey, starting with the demographics of the sample and followed by the three aspects investigated: Data extraction topics, their organization, and data analysis results.

5.1.1 Sample

In total, 21 out of 25 invited experts responded to the survey. Of the 21 respondents, 12 were external experts and 9 were experts from CRC1153. In the following, we report on the demographics of all experts summarized, as they represent a group of experts in the engineering sciences and their subdivision was only necessary to ensure the validity of the assessment of the data analysis results. However, we provide all details of the individual groups in Figure 14. The 21 respondents are experts in the engineering sciences from the three subject areas: Mechanical and industrial engineering (10 experts), computer science (9 experts), and material sciences and engineering (2 experts). All experts have at least a master’s degree in engineering sciences and at least 2 (on average 8.9 ± 7.8 , [2–30]) years of professional experience in engineering sciences. Therefore, the respondents have the necessary expertise to assess our results, as they have extensive experience and education backgrounds in engineering sciences, that include both content-related and technical expertise.

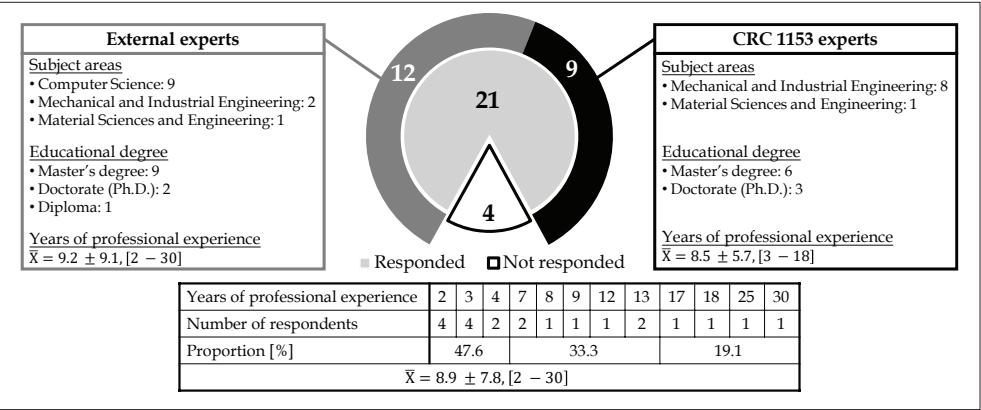


Figure 14 Demographics of the experts surveyed.

5.1.2 Data extraction topics

Despite the clear idea of the topics for data extraction due to TFPC use case, the question arises whether these topics are also generally useful for describing process chains in the engineering sciences. For this reason, we asked all 21 experts to indicate their level of agreement with the usefulness of each data extraction topic.

In Figure 15, we present the experts’ level of agreement, showing a positive trend overall. It is noticeable that no expert disagrees or even strongly disagrees with the usefulness of the data extraction topics, and on average only $11.8\% \pm 9.1$, [0.0%–28.6%] (2.5 respondents) of the respondents have a neutral level of agreement. In contrast, there is an overall high level of agreement. On average, $88.8\% \pm 9.1$, [71.4%–100.0%] (18.7 respondents) of the respondents

⁴³ <https://mybinder.org/v2/gh/okarras/Jupyter-Notebooks/HEAD?labpath=%2FTailoredFormingAnalysis%2Ftforkg.ipynb>.

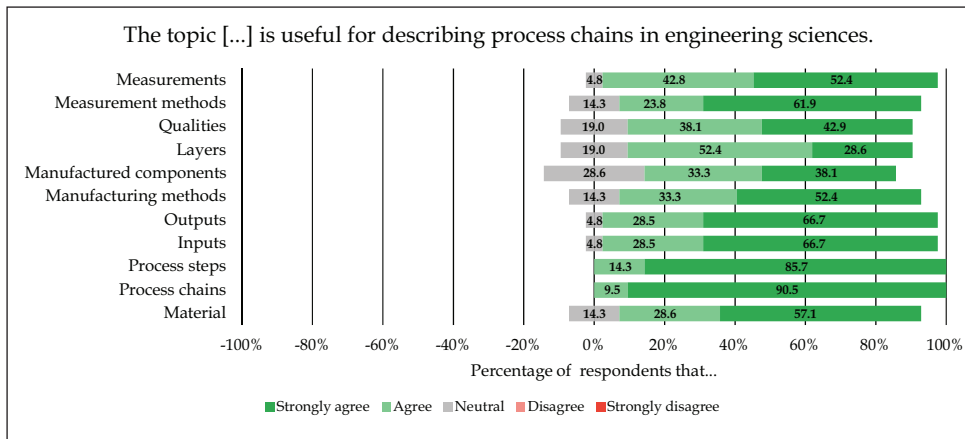


Figure 15 Experts' level of agreement with the usefulness of the data extraction topics.

agree or strongly agree with the usefulness of the data extraction topics. The experts show a high level of agreement with core process chain elements, such as the process chain, process steps, inputs, outputs, and measurements. The methods used and the component-specific topics, such as the manufactured component, the materials, qualities, and layers were assessed more neutrally, but still with a high level of agreement. In total, the results show that the majority of the experts assess the data extraction topics to be useful for describing process chains in engineering sciences.

5.1.3 Organization of the topics

We used an ORKG template to organize the data extraction topics. However, some design decisions had to be made when structuring and modeling the topics. Even though the ORKG template was developed with a clear strategy and internally reviewed and confirmed (cf. Development.), an external validation through experts is necessary to verify its applicability and effectiveness across various subject areas within engineering sciences. For this reason, we asked all 21 experts to indicate their level of agreement with the statements describing the organization of the data extraction topics (cf. Table 2).

Figure 16 presents the experts' level of agreement with these statements. Overall, the results show a positive trend, but there are some minor disagreements and larger neutral level of agreement. However, no expert strongly disagrees with any statement and only 6.5% $\pm$ 5.7, [0.0%–14.3%] (1.4 respondents) of the respondents disagree with some of the statement. On average, 17.3% $\pm$ 14.7, [0.0%–42.9%] (3.6 respondents) of the respondents have a neutral level of agreement, which means that disagreement and neutral level of agreement together are still below 25% on average. The majority of the experts, with an average of 75.6% $\pm$ 18.8, [37.6%–100.0%] (15.9 respondents), agree or strongly agree with statements. A closer look into the data reveals that external experts mostly disagree with various statements or have a neutral level of agreement (14 disagreements and 22 neutral assessments in total) compared to CRC 1153 experts (1 disagreement and 11 neutral assessments in total) (Karras, 2024). These results indicate that the organization of the topics in the ORKG template is well-suited for the area of TFPCs, while other sub-disciplines likely require refinements of topics and their organization or even additional ones to comprehensively and adequately describe their process chains. Nevertheless, the results show that the majority of the experts agree with the organization of the data extraction topics.

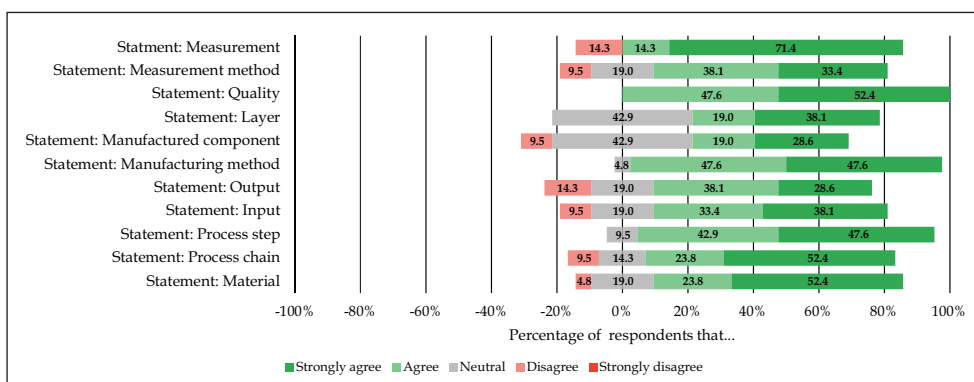


Figure 16 Experts' level of agreement with the organization of the data extraction topics.

5.1.4 Data analysis results

As a part of the solution design results (cf. section 4.2.1), two domain experts confirmed that the answers with the visualizations provide relevant information to the competency questions. However, a more comprehensive assessment of the relevance of the competency questions and answers with visualizations is required to ensure their validity for TFPCs in engineering sciences. For this reason, we asked the nine CRC 1153 experts to assess their level of agreement with the relevance of the competency questions and answers with their visualizations.

In Figure 17, we present the experts' level of agreement, showing a positive trend overall. No expert strongly disagrees, and only in two cases regarding competency questions 7 and its answer, one expert disagrees. On average, only $1.4\% \pm 3.8$, [0.0%–11.1%] (0.1 respondents) of the respondents disagree with the relevance of the competency questions and answers for TFPCs in engineering sciences. We also observe a rather low neutral level of agreement with an average of $12.5\% \pm 11.4$, [0.0%–33.4%] (1.1 respondents) of the respondents. The majority of the experts, with an average of $84.9\% \pm 12.8$, [55.5%–100.0%] (7.6 respondents), agree or strongly agree with the relevance of the competency questions and answers with visualizations for TFPCs. In particular, the results show that the experts assess competence question 7 as less relevant, but nevertheless agree to a large extent with the relevance of the answer and the visualization. This result indicates that even if the competency question is less relevant, the answer and visualization are still helpful. For competency question 5, we observe the opposite. In this case, the experts assess the competency question as very relevant, but the answer and visualization are less satisfactory, which is reflected by the higher neutral level of agreement. In this case, the answer and visualization is only based on three data points, as the specific information on the service life of the hybrid components was only reported for three of the 10 TFPCs (see section 4.2.2 and Figure 12). Therefore, the higher neutral level of assessment seems to be less a problem of the answer and the visualization itself, but rather the underlying limited amount of data. In summary, the results show that the majority of the experts assess the competency questions and answers with visualizations to be relevant for TFPCs in engineering sciences.

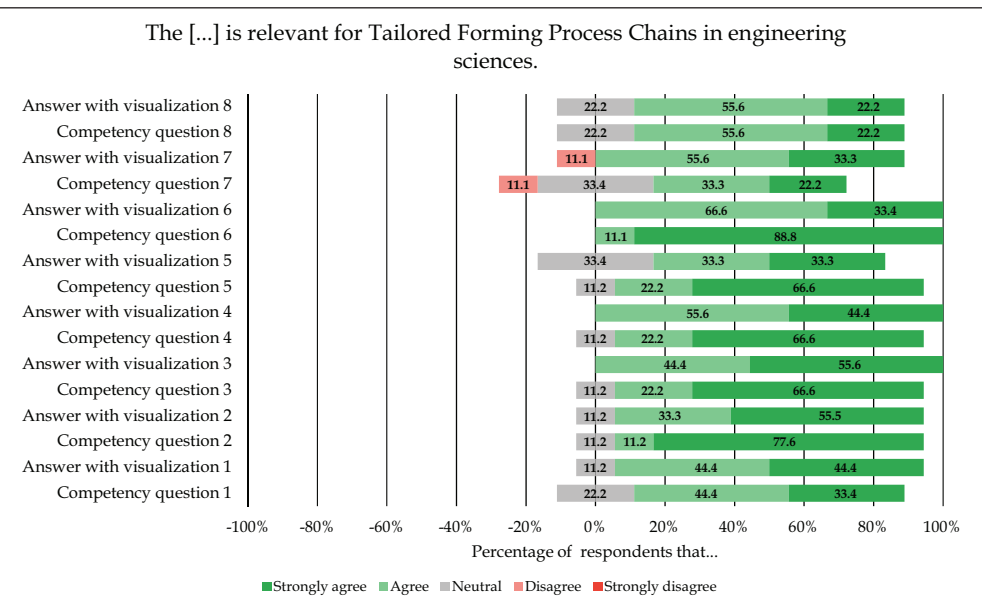


Figure 17 Experts' level of agreement with the relevance of the competency questions and answers with visualizations for Tailored Forming Process Chains in engineering sciences.

In summary, the survey results show that the 21 experts are generally of the opinion that the topics of data extraction are useful for describing process chains in engineering sciences and that their organization in the ORKG template is appropriate, but that there is still potential for improvement and expansion. In this way, the results support the validity of our research findings. However, comprehensive validation requires the reuse of the ORKG template to publications outside the CRC 1153 to demonstrate its applicability. We present the results of this reuse of the ORKG template in the next section.

5.2 APPLICABILITY BEYOND CRC 1153

Below, we report on the reuse of the developed ORKG template on three publications outside the CRC 1153 that deal with a total of 20 AMPCs.

Overall, we were able to fully apply the ORKG template to all AMPCs and thus organize the corresponding scientific knowledge from the associated publications. This result basically shows the applicability of the ORKG template beyond CRC 1153 for other manufacturing process chains.

Similar to the TFPC use case (cf. section 4.1.2), we analyze the 20 described AMPCs regarding their complexity in terms of number of statements, resources, and literals. We compare these results with the results from the 10 described TFPCs to better understand how well the ORKG template could be used to describe the AMPCs. We expect that the numbers of statements, resources, and literals will be similar across all 30 process chains, but differences may occur depending on the reported information in terms of the number of process steps performed, qualities investigated, measurements conducted, etc.

Regarding the complexity, we found that each description of the 20 AMPCs consists, on average, of 125.95 ± 57.08 , [62–295] statements, 237.90 ± 103.84 , [119–547] resources, and $14.00\% \pm 10.60$, [5–43] literals. These numbers are comparable to the results of the 10 described TFPCs with, on average, 191.80 ± 55.23 , [132–313] statements, $336.2\% \pm 104.69$, [259–591] resources, and $17.40\% \pm 7.23$, [5–35] literals (cf. Table 3). In Figure 18, we show boxplots of the complexity of all 30 process chains divided into AMPCs (x-symbols) and TFPCs (x-symbol). Figure 18 shows a good distribution of the 30 process chains across the respective boxplots, whereby the TFPCs are usually described with more statements, resources, and literals than the AMPCs. A closer look at the TFPCs and AMPCs shows that the process chains differ in their number of reported process steps. While the TFPCs report on median 3 process steps, the AMPCs report on median 2 process steps. Consequently, the descriptions of the TFPCs require more statements, resources, and literals on average. In principle, however, the descriptions of the AMPCs are similarly comprehensive as the descriptions of the TFPCs. The extensive use of resources compared to literals is a positive indicator for precise and fine-grained semantic descriptions resulting from the detailed structure of the ORKG template. Thus, the reuse of the ORKG template on the 3 publications outside the CRC 1153 with the 20 AMPCs represents another comprehensive example which underlines the applicability of the developed ORKG template beyond the CRC 1153.

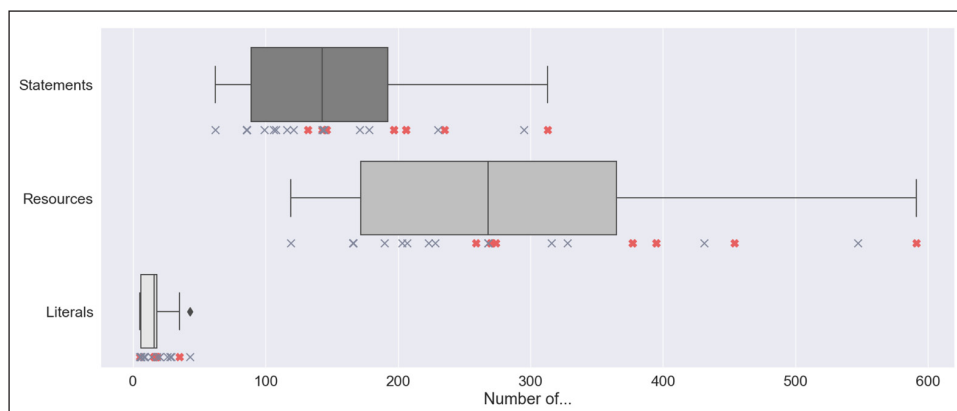


Figure 18 Boxplots of the complexity of 30 process chains described.

Remark: The x-symbol marks the 20 AMPCs from the three external publications. The x-symbol marks the 10 TFPCs from the five publications of the CRC 1153. The ♦-symbol marks potential outliers that are more than 1.5 times the interquartile range above the upper quartile.

In the following, we illustrate how scientific knowledge from all 30 process chains described can be jointly (re-)used. Based on CQ3 (cf. Table 1), we developed a new SPARQL query (see Listing 2) that queries the reported hardness in the outermost layer of the final manufactured component, if available, and visualizes the results as a bar chart. Figure 19 shows that 27 out of the 30 process chains report a hardness in the outermost layer of the final manufactured component. These results are relevant for engineers who look for a process chain that can be used to manufacture a component that can withstand high mechanical or tribological stress. Overall, we have shown that the ORKG template is applicable for comprehensive descriptions of manufacturing process chains beyond the CRC 1153. The consistently structured descriptions enable detailed searches that allow an easy comparison of different process chains.

Our results, observations, and findings are subject to several threats to validity that must be considered. The discussion of these threats is important to highlight areas where caution should be used in interpreting the results. In the following, we discuss the threats to validity in more detail.


```

1 PREFIX orkg: <http://orkg.org/orkg/resource/>
2 PREFIX orkgc: <http://orkg.org/orkg/class/>
3 PREFIX orkgp: <http://orkg.org/orkg/predicate/>
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
6
7 #This query can be executed in the visual SPARQL editor of ORKG: https://www.orkg.org/sparql.
8 #You must delete line 9 if you only want to see the results table rather than the bar chart.
9 #defaultView:BarChart
10 SELECT (STR(?orkg_contribution) AS ?paper), ?final_value, (STR(?final_unit) AS ?Unit)
11 WHERE {
12   #Get final components of all ORKG contributions being an instance of our ORKG template
13   ?orkg_contribution a orkgc:C21027;
14     orkgp:HAS_RESULTS ?final_component.
15
16   #Get all measurements of the final manufactured components
17   ?final_component orkgp:P28005 ?layer.
18   ?layer orkgp:P3 ?layer_type;
19     orkgp:P7000 ?material.
20   ?material orkgp:P7113 ?quality.
21   ?quality orkgp:P7111 ?quality_type;
22     orkgp:P44022 ?measurement.
23
24   #For range measurements, get the lower and upper measured value and their unit
25   OPTIONAL{?measurement orkgp:P44010 ?range.
26     ?range orkgp:P43248 ?lower_value;
27       orkgp:P43247 ?upper_value.
28     ?lower_value orkgp:P44011 ?lower_num_value.
29     ?upper_value orkgp:P44011 ?upper_num_value.
30     ?upper_value orkgp:P42000 ?range_unit.
31     ?range_unit rdfs:label ?range_unit_name.
32   }
33
34   #For single value measurements, get the measured value and its unit
35   OPTIONAL{?measurement orkgp:HAS_VALUE ?measurement_value.
36     ?measurement_value orkgp:P44011 ?value;
37       orkgp:P42000 ?unit.
38     ?unit rdfs:label ?unit_name.
39   }
40
41   #If no single value is reported, calculate the average based on the range measurement
42   BIND(IF (BOUND(?value), "", ((xsd:integer(?lower_num_value) + xsd:integer(?upper_num_value)
43     ↪ ) / 2)) AS ?avg_range).
44   BIND(IF (BOUND(?value), ?value, ?avg_range) AS ?final_value).
45
46   #If no single value is reported, get the unit from the range measurement
47   BIND(IF (BOUND(?value), ?unit_name, ?range_unit_name) AS ?final_unit).
48
49   #Filter for the top layers respective single layers with a hardness measurement
50   ?layer_type rdfs:label ?layer_type_name.
51   ?quality_type rdfs:label ?quality_name.
52   FILTER(?layer_type_name = "top_layer"^^xsd:string || ?layer_type_name = "single_layer"^^
53     ↪ xsd:string )
54   FILTER(?quality_name = "Hardness"^^xsd:string)
55 } Order BY DESC(?final_unit), DESC(?final_value)

```

Listing 2 SPARQL query to retrieve the hardness in the outermost layer of all final components manufactured in the 30 described process chains, if reported.

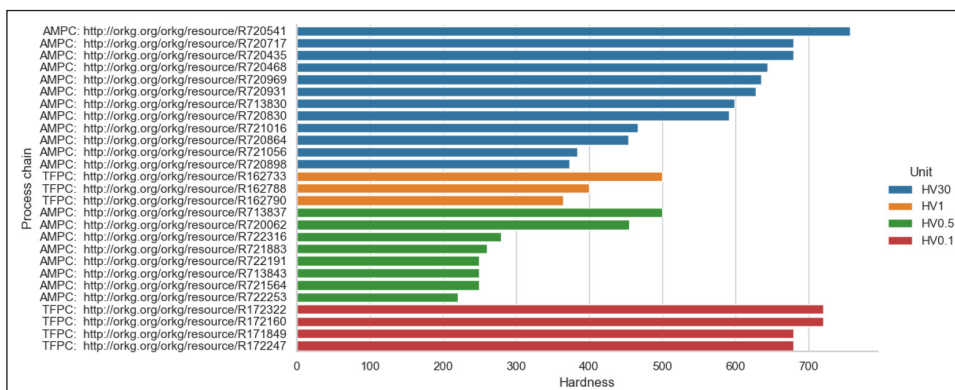


Figure 19 Hardness in the outermost layer of all final manufactured components, reported in 27 of the 30 process chains described.

6 THREATS TO VALIDITY

We discuss threats to *study selection*, *data*, and *research validity* based on the guideline for managing threats to validity of secondary studies by Ampatzoglou et al. (2020).

Study selection validity includes threats to the search process and study filtering. Regarding the *accessibility of the publications*, the majority (four out of five publications) is open access, increasing the transparency and replicability of our research approach. We ensured the *adequacy of relevant publication identification* by having co-authors from the CRC 1153 who selected the *included publications* for our use case. The familiarity of the co-authors from the CRC 1153

with the included publications guaranteed their suitability for our use case, the accuracy of the extracted knowledge, and its correct semantic description. However, the number of selected publication is small, which inherently limits the generalizability of the results. While a smaller set of publications mitigates the risk of variation due to random heterogeneity, it also leads to a narrow scope as it does not represent the real world with all its available and potentially relevant data. For the CRC 1153, it was more important to obtain an initial coherent starting point for the description of TFPCs than a holistically comprehensive generic approach. For this reason, we *limited the number of publications* so that we obtain a more homogeneous but complete set of publications to simplify the data extraction and semantic description, as our use case is the first attempt to describe TFPCs in the ORKG. Even if the developed ORKG template and thus the topics covered are not yet comprehensive, our approach provides a solid starting point for the description of process chains in engineering sciences according to the results of the expert consultation and thus represents a reliable basis for future improvements and extensions.

Data validity includes threats to data extraction and data analysis. One of the most central threats to data validity is the *sample size limitation* of 10 TFPCs due to the smaller number of included publications. As explained above, we limited the number of publications to obtain a more homogenous dataset which mitigates the threat of *heterogeneity of the publications*. Heterogeneous data is not safe to synthesize and is therefore prone to involve a high degree of subjectivity (Ampatzoglou et al., 2020). However, this decision means that the results are prone to bias and not safe to generalize. The selection is inevitably accompanied by a *publication bias* as we only consider publications from the CRC 1153, which was the central motivation for our use case due to the need for an overview of their own publications. We mitigated this threat to validity by applying the ORKG template to three randomly selected publications on AMPCs reporting a total of 20 process chains. This part of the validation demonstrates the applicability of our solution design to publications outside the CRC 1153. For the mitigation of *data extraction bias* and *researcher bias*, we had a clear strategy by involving the first three authors and an expert consultation (Ampatzoglou et al., 2020). Regarding the data extraction, the first three authors were involved. In particular, the first author is a knowledge engineer and thus a technical expert in the semantic modeling and organization of knowledge. The other two authors are engineers and thus domain experts with professional expertise in engineering sciences. For each publication, they applied the following procedure: 1) The knowledge engineer and one domain expert extracted the knowledge from the publications, 2) The remaining domain expert reviewed the described knowledge by comparing it with the publication, and 3) All three experts directly discussed and resolved any identified inconsistencies or ambiguities. For the data extraction, the first three authors developed an ORKG template in three iterations over a total period of two weeks, which was finally reviewed and confirmed by the other authors. The development of the ORKG template was based on a clear idea about the topics for data extraction by the co-authors from the CRC 1153. This clear idea ensured the correctness of the *choice of variables to be extracted*. The structured organization of these topics and their associated variables to be extracted as an ORKG template mitigated the threats of *data extraction inaccuracies* and *classification scheme bias*. In addition, an expert consultation was conducted in which 21 experts assessed the usefulness of the data extraction topics, their organization in the ORKG template, and the relevance of the data analysis results. The results of this consultation show a clear positive agreement of the experts regarding the usefulness of the data extraction topics (cf. Figure 15), their organization (cf. Figure 16), and the relevance of the data analysis results (cf. Figure 17) with a potential for future improvements and expansion.

Research validity includes threats to the entire research process and design. First, our research approach has a *research method bias* resulting from the use of only one method for data collection, namely a review of literature. Despite its weaknesses, this method enables for systematic reflection as a proven means for building an initial, sound knowledge base (Glinz and Fricker, 2015; Karras et al., 2020). Furthermore, our use case is a first attempt to semantically describe TFPCs in the ORKG. However, we conducted a validation consisting of an expert consultation and the reuse of the ORKG template to publications outside the CRC 1153 to increase the research validity of our results and finding. For these reasons, we assume that our research approach is a suitable to illustrate how engineers can utilize the ORKG in innovative ways to organize scientific knowledge for communication and (re-)use beyond publications.

We described our research approach in detail and made all data and materials⁴⁴ (Karras, 2024) openly accessible to ensure their *reliability*, *repeatability*, and *reproducibility*. Regarding *generalizability*, we must emphasize that this article presents initial work for describing TFPCs semantically in the ORKG. The focus of this work is on the development and provision of an initial ORKG template for the semantic description of TFPCs based on an appropriate use case and in close cooperation with domain experts. The data analysis is therefore not intended to provide general insights into the research on TFPCs, but rather to demonstrate how engineers can utilize the ORKG infrastructure in innovative ways to sustainably organize FAIR scientific knowledge for communication and (re-)use and thus advance the field of engineering sciences.

7 DISCUSSION

Along our research approach, we discuss our results and findings below, focusing on three central aspects: The semantic description for data extraction, the ORKG comparison for communicating scientific knowledge, and the answering of the competency questions for (re-)using scientific knowledge, both of which are part of the data analysis. Finally, we present our plan for future work with short-, mid-, and long-term actions.

7.1 SEMANTIC DESCRIPTION

Data extraction was the first challenge, as the organization of scientific knowledge required the development of a structure for the semantic description of the scientific knowledge, for which we used an ORKG template. The efforts to establish and adhere to a common structure and terminology are crucial to ensure semantic consistency and interoperability, which are necessary to make the knowledge accessible and reusable.

Despite a clear idea about the topics for data extraction (cf. section 3.1.2), it was challenging to develop an ORKG template that is as generic as possible and thus reusable, as engineering sciences with all its disciplines has domain-specific terminologies. There is a lack of established metadata standards and ontologies focusing on engineering sciences (Schmitt et al., 2020) that we could have used when developing the ORKG template. The initial version of a domain-specific ontology for Tailored Forming technology by Sheveleva et al. (2020) was a good starting point for developing the ORKG template, but this ontology was never officially published. After we extracted all knowledge using the reviewed and confirmed ORKG template and published the ORKG comparison, we became aware that NFDI4Ing had published the first version of its Metadata4Ing ontology. This ontology is used for the semantic description of research data focusing on engineering sciences and its disciplines (Arndt et al., 2023). The special interest group *Metadata & Ontologies* of NFDI4Ing⁴⁵ released this ontology at the time of the data extraction. In a first subsequent comparison of the Metadata4Ing ontology with the ORKG template, we found fundamental and promising similarities in the semantic description, especially regarding processes, their steps, their sequences, and measurements as central elements. These aspects are one of the most distinctive features of the CRC 1153 (cf. section 2.3) and are therefore highly important. The usefulness of these and other topics is also confirmed by the expert evaluation (cf. Figure 15). However, it is not yet possible to predict whether the Metadata4Ing ontology will become established in the engineering sciences in the future. For this reason, we are not currently harmonizing the ORKG template with this ontology, as a complete revision and reclassification of all 1918 statements with their 3662 resources (see Table 3) would require a lot of effort, which we consider to be disproportionate given the unclear establishment of the ontology. If the Metadata4Ing ontology becomes established in the long term, we plan to harmonize the ORKG template with this ontology in the future to increase the accuracy of the conceptual entities and their relationships. Nevertheless, we interpret the agreement in the semantic description with its focus on processes, their steps, their sequence, and measurements as a positive indicator that we have developed the ORKG template in the right direction. This type of description enables better traceability and reproduction of all steps in the engineering (research) process which are necessary to guarantee the trustworthiness of published results, prevent redundancies, and ensure social acceptance (Schmitt et al., 2020).

⁴⁴ <https://github.com/okarras/TailoredForming-Analysis>.

⁴⁵ <https://nfdi4ing.de/special-interest-groups-sig/metadata-ontologies/>.

The semantic description is the basis to sustainably organize of scientific knowledge for communication and (re-)use. However, general provision is insufficient, as engineers must be able to access the knowledge in an understandable way to (re-)use it. For this reason, we divided the data analysis into two parts, one focusing on the communication of knowledge by means of an ORKG comparison and the other one on the (re-)use of knowledge by answering competence questions. Below, we discuss these two aspects.

7.2 COMMUNICATION OF KNOWLEDGE

For the communication of knowledge, we needed a central access point that provides a solid and easily understandable overview of the interrelated knowledge. This central access point was important for the CRC 1153 as its employees lack such an overview (cf. section 2.3).

ORKG comparisons offer precisely this type of access point to a specific topic as an innovative way for communication of knowledge. Given the specific topic of our use case from the CRC 1153, using an ORKG comparison as a central access point to the described knowledge is a suitable solution. Its visualization in the form of an interactive table can be used easily by experts and laypersons to explore the contents and their interrelationships. For example, the CRC 1153 employees and every ORKG user can access and tailor the ORKG comparison to their individual needs due to its interactive nature. These tailored tables can, in turn, form the basis for new ORKG comparisons. However, the CRC 1153 employees and ORKG users can also expand the existing ORKG comparison with additional publications and knowledge, resulting in a revised version. These possibilities of (re-)use allow for continuous expansion and evolutionary tracking, which makes research more sustainable overall. With the feature of enriching ORKG comparisons with supplementary materials, such as our Jupyter notebook and its generated visualizations, we compiled all results on a single web page to explore them. This feature has rarely been used in the ORKG so far (cf. Table 3), but our use case clearly demonstrates how useful it is to provide a central overview of all materials. In this way, we organized and linked the described knowledge with auxiliary information, provenance data, and appropriate visualizations, including the source code for reproducibility, to facilitate the understanding and interpretation of the knowledge in its specific context of the CRC 1153, but also for hitherto unknown purposes in the future.

We have to emphasize that the solution of using an ORKG comparison was, in particular, suitable as we had the use case with its specific topic on *“Tailored Forming Process Chains for the Manufacturing of Hybrid Components with Bearing Raceways Using Different Material Combinations.”* For a central overview of an overarching theme such as Tailored Forming, which covers several specific topics, the ORKG offers the possibility to establish a so-called ORKG observatory. An ORKG observatory is an open group that anyone can join to contribute to its overarching theme by maintaining, (re-)using, updating, and expanding its contents, such as ORKG comparisons and related publications. Each ORKG observatory has its own web page as central access point, where its description, focused research problems, members with their organization, and all related contents in the ORKG are compiled. We have refrained from setting up such an ORKG observatory for the time being, as we currently have little content and the overarching theme is not yet clearly defined among the authors of this article. Nevertheless, we keep this possibility in mind for future work as it enables the organization, linking, and communication of FAIR scientific knowledge across individual publications, different organizations, and entire projects. For a collaborative, open-science project like the CRC 1153, an ORKG observatory offers significant benefits, enhancing the project’s efficiency and transparency by addressing the challenges of communication and (re-)use of scientific knowledge.

7.3 (RE-)USE OF KNOWLEDGE

As mentioned above, the communication of semantically described knowledge with a central access point lays the foundation for its (re-)use, which is a cornerstone in advancing research and innovation, particularly in engineering sciences. For example, engineers (re-)use knowledge and data from mechanical processes to develop new or revised processes for manufacturing components that meet modern society’s challenges. As a part of this article, it was important for us to show how the described knowledge can be meaningfully (re-)used to obtain new overviews and insights.

The semantic description of the extracted knowledge provides a comprehensive and interconnected representation that fosters a deeper understanding of the engineering concepts

and their relationships. This understanding, especially of the structure, is important to retrieve the knowledge from ORKG in a targeted manner for its (re-)use. Thus, the machine-actionable representation enables data-driven analysis methods including the use of artificial intelligence, machine learning, natural language processing, or large language models. As a first step into this direction, we used the established method of answering competency questions for analyzing and evaluating KGs using a manual analysis script. Based on the described knowledge, we provided answers and visualizations to the competency questions asked by two domain experts. In this way, we (re-)used the described knowledge to provide relevant information that meets their information needs. Our results show that we answered all competency questions by providing relevant information for the respective information needs. In expert consultation, the nine experts from the CRC 1153 indicated a high level of agreement, that the competency questions and answers with visualizations are relevant for TFPCs in engineering sciences (cf. [Figure 17](#)). However, the more specific a competency question is, the more difficult it is to provide an answer as the required detailed information might be missing in the publications. This challenge applies not only to our manual analysis but also to more sophisticated approaches such as scientific question answering systems that use artificial intelligence and large language models which require detailed contextual information to provide correct answer ([Auer et al., 2023a](#)).

Overall, our analysis shows how the described knowledge can be (re-)used for data-driven analysis methods. However, we had to refrain from using more sophisticated methods, i.e., artificial intelligence, machine learning, natural language processing or large language models as our collected amount of data is too small. Nevertheless, the organization of scientific knowledge using the ORKG paves the way for these kinds of data-driven analysis methods once sufficient data is available. In engineering sciences, this type of analysis can enable engineers to gain valuable insights, optimize processes, and drive innovation through the exploration and (re-)use of scientific knowledge.

Given these insights, the use of the ORKG to organize scientific knowledge for communication and (re-)use differs from traditional approaches as the knowledge is not encapsulated in a published PDF file. Instead, the described knowledge is organized in an openly accessible knowledge graph, which, to put it simply, is nothing more than a graph-based database. Essentially, the ORKG provides a ready-to-use and sustainably governed infrastructure that adheres to established best practices, such as the FAIR principles ([Wilkinson et al., 2016](#)) and versioning ([Stocker et al., 2023](#)). Even if the semantic description of the knowledge is challenging, especially in our case with the lack of established metadata standards and ontologies focusing on engineering sciences ([Schmitt et al., 2020](#)), the ORKG offers the necessary support with its ORKG template feature, flexibility, and openness. Overall, the ORKG offers a comprehensive infrastructure with its services supports researchers in the sustainable organization of FAIR scientific knowledge for communication and (re-)use, that can be understood, processed, and used by humans and machines, so that scientific knowledge is openly accessible to everyone in the long term. As an answer to our research question, we can summarize:

Answer to the research question:

The ORKG is a ready-to-use infrastructure with services designed to support researchers from all research fields in sustainably organizing FAIR scientific knowledge on any topic in an openly accessible and long-term manner. Our use case confirms this intention for engineering sciences and the topic of TFPCs. The direct use of the ORKG by engineers is feasible, so the ORKG is a promising infrastructure that offers innovative ways for communication and (re-)use of FAIR scientific knowledge to advance the field of engineering sciences.

7.4 FUTURE WORK

For future work, we have a plan with short-, mid-, and long-term actions, which we present below.

As short-term action, we started to describe more publications using the ORKG template (cf. section 5.2). In this way, we verified the applicability of the ORKG template and at the same time increase the amount of data.

As mid-term actions, we extend the use of the ORKG template to publications with a broader scope on the overarching themes of Tailored Forming and process chains. With this approach, we can investigate how generic and reusable the ORKG template is to revise and extend it for improvement and refinement. Given the expansion of the amount of data regarding the overarching themes, we establish a corresponding ORKG observatory. This observatory serves as central access point to the scientific knowledge for all employees of the CRC 1153, but also any interested member of the engineering sciences community.

As long-term actions, we work on solutions regarding scalability and integration. While the ORKG provides the basis for transferring the presented approach to other areas in which data management is of crucial importance, the semantic description of knowledge represents the central challenge. Engineers and researchers in general need better support to address this challenge. A promising direction is the integration of semantic description into the research lifecycle so that knowledge is captured semantically at an early stage and, in the best case, as a by-product of the actual work, based on the idea of “FAIR-by-Design” (Karras et al., 2023b). Recently, a number of such approaches have been developed that enable the semantic description of knowledge in different stages of the research life cycle, for example during measurements (Mozgova et al., 2022), data analysis (Stocker et al., 2024), or the writing of a publication (Bless et al., 2022; Bless et al., 2023; Martin and Henrich, 2022). In the best case, the generated semantic descriptions can even be automatically stored in RKGs. In this regard, we also continuously monitor the developments surrounding the Metadata4Ing ontology concerning the currently open question of its establishment. In the case of its establishment or another similar ontology, we harmonize the ORKG template with the respective ontology to increase the accuracy of the conceptual entities and their relationships. Based on these actions, we expect a significant increase in the amount of data required to investigate the use of more sophisticated data analysis methods, including the use of artificial intelligence, machine learning, natural language processing, and large language models to advance the field of engineering sciences.

With this plan, we work towards establishing the ORKG as a promising infrastructure for the sustainable organization of FAIR scientific knowledge in engineering sciences.

8 CONCLUSION

Engineering sciences are essential for addressing contemporary technical, environmental, and economic challenges. Despite the increasing importance of FAIR scientific knowledge and data in this research field, including the efforts of NFDI4Ing, this topic remains understudied. Given the data-intensive and interdisciplinary nature of engineering sciences, engineers need infrastructures with services that support them in organizing FAIR scientific knowledge and data for communication and (re-)use.

The ORKG is one of these infrastructures. Based on its several success stories in different research fields and central role in NFDI4Ing, we examined how engineers can utilize the ORKG in innovative ways for communication and (re-)use of FAIR scientific knowledge in engineering sciences. We have focused on a concrete use case to provide specific examples that others can easily understand in detail. On the one hand, we published the ORKG comparison to show how scientific knowledge can be communicated beyond publications. This type of presentation makes knowledge more accessible not only for humans, but also for machines, so that it can be (re-)used to gain new overviews and insights. We demonstrated this (re-)use by answering eight competency questions asked by two domain experts. The validation of our results shows a clear agreement from experts in engineering sciences regarding the usefulness of the data extraction topics, their organization, and the relevance of the competency questions and answers for engineering sciences. We also reuse of the ORKG template on publications outside the CRC 1153 to demonstrate its applicability.

Given our use case and the validation with their results and insights, we conclude that the ORKG is a promising infrastructure for engineering sciences. Although it is ready-to-use and sustainably governed, thus providing a solid foundation, further actions are required as future work to better understand the needs of engineers for improving ORKG to support them. Nevertheless, the use of ORKG is already feasible for engineers, providing them with innovative ways for the communication and (re-)use of FAIR scientific knowledge, which in turn advances the field of engineering sciences.

The data of this article is openly accessible in the Open Research Knowledge Graph (<https://orkg.org/>). We provide all supplementary materials on GitHub (<https://github.com/okarras/TailoredForming-Analysis>) and the interactive Jupyter notebooks on mybinder.org (<https://mybinder.org/v2/gh/okarras/TailoredForming-Analysis/HEAD?labpath=%2Ftailored%20forming%20analysis.ipynb> and <https://mybinder.org/v2/gh/okarras/TailoredForming-Analysis/HEAD?labpath=%2Fcomparison%20analysis.ipynb>).

FUNDING INFORMATION

The authors thank the Federal Government, the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the NFDI4Ing, NFDI4DataScience, and NFDI4Energy consortia. This work was funded by the German Research Foundation (DFG) for the projects NFDI4Ing: 442146713, NFDI4DataScience: 460234259, NFDI4Energy: 501865131, and CRC 1153: 252662854, and by the TIB - Leibniz Information Centre for Science and Technology. The publication of this article was funded by the Open Access Fund of Technische Informationsbibliothek (TIB).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

We use the CRediT author statement⁴⁶ to recognize individual author contributions.

Conceptualization: OK, LB, SA

Methodology: OK, LB, PM

Software: OK

Validation: OK, LB, PM

Formal analysis: OK

Investigation: OK, LB, PM

Resources: OK, LB, PM

Data curation: OK, LB, PM

Writing – Original draft: OK, LB, PM

Writing – Review & editing: OK, LB, PM, JH

Visualization: OK, PM

Supervision: MS, LO, BB, SA

Project administration: OK, MS, LO, BB, SA

Funding acquisition: MS, LO, BB, SA

AUTHOR AFFILIATIONS

Oliver Karras  orcid.org/0000-0001-5336-6899

TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany

Laura Budde  orcid.org/0000-0002-0168-3363

Laser Zentrum Hannover e.V., Hannover, Germany

Paulina Merkel  orcid.org/0000-0001-5594-8622

IPH – Institut für Integrierte Produktion Hannover gGmbH, Hannover, Germany

Jörg Hermsdorf  orcid.org/0000-0003-1820-1516

Laser Zentrum Hannover e.V., Hannover, Germany

⁴⁶ <https://credit.niso.org/>.

Malte Stonis  orcid.org/0000-0001-5957-3469
IPH – Institut für Integrierte Produktion Hannover gGmbH, Hannover, Germany

Ludger Overmeyer  orcid.org/0000-0002-7734-8895
Laser Zentrum Hannover e.V., Hannover, Germany

Bernd-Arno Behrens  orcid.org/0000-0001-6845-9975
IPH – Institut für Integrierte Produktion Hannover gGmbH, Hannover, Germany;
Institute of Forming Technology and Machines, Hannover, Germany

Sören Auer  orcid.org/0000-0002-0698-2864
TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany;
L3S Research Center, Leibniz University Hannover, Hannover, Germany

REFERENCES

- Ammar, W. et al.** (2018) ‘Construction of the Literature graph in semantic scholar’, in S. Bangalore, J. Chu-Carroll, and Y. Li (eds) *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)*. New Orleans – Louisiana: Association for Computational Linguistics, pp. 84–91. Available at: <https://doi.org/10.18653/v1/N18-3011>
- Ampatzoglou, A. et al.** (2020) ‘Guidelines for Managing Threats to Validity of Secondary Studies in Software Engineering’, in M. Felderer and G.H. Travassos (eds.) *Contemporary Empirical Methods in Software Engineering*. Cham: Springer International Publishing, pp. 415–441. Available at: https://doi.org/10.1007/978-3-030-32489-6_15
- Anteghini, M. et al.** (2020) ‘Representing Semantified Biological Assays in the Open Research Knowledge Graph’, in E. Ishita, N.L.S. Pang, and L. Zhou (eds.) *Digital Libraries at Times of Massive Societal Transition*. Cham: Springer International Publishing, pp. 89–98. Available at: https://doi.org/10.1007/978-3-030-64452-9_8
- Arndt, S. et al.** (2023) ‘Metadata4Ing: An ontology for describing the generation of research data within a scientific activity’. *Zenodo*. Available at: <https://doi.org/10.5281/zenodo.8382665>
- Aryani, A. and Wang, J.** (2017) ‘Research Graph: Building a Distributed Graph of Scholarly Works using Research Data Switchboard’, in *Open Repositories Conference*. Available at: <https://doi.org/10.4225/03/58c696655af8a>
- Auer, S. et al.** (2007) ‘DBpedia: A Nucleus for a Web of Open Data’, in K. Aberer et al. (eds.) *The Semantic Web*. Berlin, Heidelberg: Springer, pp. 722–735. Available at: https://doi.org/10.1007/978-3-540-76298-0_52
- Auer, S. et al.** (2018) ‘Towards a Knowledge Graph for Science’, in *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics (WIMS '18)*. New York, NY, USA: Association for Computing Machinery, pp. 1–6. Available at: <https://doi.org/10.1145/3227609.3227689>
- Auer, S. et al.** (2020) ‘Improving Access to Scientific Literature with Knowledge Graphs’, *Bibliothek Forschung und Praxis*, 44(3), pp. 516–529. Available at: <https://doi.org/10.1515/bfp-2020-2042>
- Auer, S. et al.** (2023a) ‘The sciq scientific question answering benchmark for scholarly knowledge’, *Scientific Reports*, 13(1), p. 7240. Available at: <https://doi.org/10.1038/s41598-023-33607-z>
- Auer, S. et al.** (2023b) ‘Organizing Scholarly Knowledge in the Open Research Knowledge Graph: An Open-Science Platform for FAIR Scholarly Knowledge’, *Proceedings of the Conference on Research Data Infrastructure*, 1. Available at: <https://doi.org/10.52825/cordi.v1i.272>
- Basili, V.R., Caldiera, C. and Rombach, H.D.** (1994) ‘Goal Question Metric Paradigm’, *Encyclopedia of Software Engineering*, 1, pp. 528–532. Available at: <https://doi.org/10.1002/0471028959.sof142>
- Behrens, B.A. et al.** (2019) ‘Manufacturing and Evaluation of Multi-Material Axial-Bearing Washers by Tailored Forming’, *Metals*, 9(2). Available at: <https://doi.org/10.3390/met9020232>
- Bless, C., Baimuratov, I. and Karras, O.** (2022) ‘SciKGTex – Scientific Knowledge Graph Tex’. Available at: <https://github.com/Christof93/SciKGTex>
- Bless, C., Baimuratov, I. and Karras, O.** (2023) ‘SciKGTex - A LATEX Package to Semantically Annotate Contributions in Scientific Publications’, in *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pp. 155–164. Available at: <https://doi.org/10.1109/JCDL57899.2023.00030>
- Brack, A. et al.** (2022) ‘Analysing the Requirements for an Open Research Knowledge Graph: Use Cases, Quality Requirements, and Construction Strategies’, *International Journal on Digital Libraries*, 23(1), pp. 33–55. Available at: <https://doi.org/10.1007/s00799-021-00306-x>
- Brockmüller, T. et al.** (2020) ‘Computer-Aided Engineering Environment for Designing Tailored Forming Components’, *Metals*, 10(12). Available at: <https://doi.org/10.3390/met10121589>
- Budde, L. et al.** (2022) ‘Investigation of the Material Combination 20MnCr5 and X45CrSi9-3 in the Tailored Forming of Shafts with Bearing Seats’, *Production Engineering*, 16(5), pp. 661–671. Available at: <https://doi.org/10.1007/s11740-022-01119-w>
- Burton, A. et al.** (2017) ‘The Scholix Framework for Interoperability in Data-Literature Information Exchange’, *D-Lib Magazine*, 23(1/2). Available at: <https://doi.org/10.1045/january2017-burton>

- Coors, T. et al. (2020) 'Investigations on Tailored Forming of AISI 52100 as Rolling Bearing Raceway', *Metals*, 10(10), p. 1363. Available at: <https://doi.org/10.3390/met10101363>
- COVID-19 Air Quality Data Collection (2021) 'COVID-19 Air Quality Data Collection'. Available at: <https://covid-aqs.fz-juelich.de> (Accessed: 21 January 2024).
- Darari, F. et al. (2018) 'Completeness Management for RDF Data Sources', *ACM Trans. Web*, 12(3). Available at: <https://doi.org/10.1145/3196248>
- Dessi, D. et al. (2022) 'CS-KG: A Large-Scale Knowledge Graph of Research Entities and Claims in Computer Science', in *The Semantic Web – ISWC 2022: 21st International Semantic Web Conference, Virtual Event, October 23–27, 2022, Proceedings*. Berlin, Heidelberg: Springer-Verlag, pp. 678–696. Available at: https://doi.org/10.1007/978-3-031-19433-7_39
- Domingo-Fernández, D. et al. (2020) 'COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology', *Bioinformatics*, 37(9), pp. 1332–1334. Available at: <https://doi.org/10.1093/bioinformatics/btaa834>
- D'Souza, J. et al. (2024) 'Quality Assessment of Research Comparisons in the Open Research Knowledge Graph: A Case Study', *JLIS.it*, 15(1), pp. 126–143. Available at: <https://doi.org/10.36253/jlis.it-547>
- Färber, M. (2019) 'The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data', in *The Semantic Web – ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18*. Springer, pp. 113–129. Available at: https://doi.org/10.1007/978-3-030-30796-7_8
- Gkatzelis, G.I. et al. (2021) 'The Global Impacts of COVID-19 Lockdowns on Urban Air Pollution: A Critical Review and Recommendations', *Elementa: Science of the Anthropocene*, 9(1), p. 00176. Available at: <https://doi.org/10.1525/elementa.2021.00176>
- Glinz, M. and Fricker, S.A. (2015) 'On Shared Understanding in Software Engineering: An Essay', *Computer Science – Research and Development*, 30, pp. 363–376. Available at: <https://doi.org/10.1007/s00450-014-0256-x>
- Goedicke, M. et al. (2024) 'National Research Data Infrastructure for and with Computer Science (NFDIxCs)'. Zenodo. Available at: <https://doi.org/10.5281/zenodo.10557968>
- Grüninger, M. and Fox, M.S. (1995) 'The Role of Competency Questions in Enterprise Engineering', in A. Røstadås (ed.) *Benchmarking — Theory and Practice*. Boston, MA: Springer US, pp. 22–31. Available at: https://doi.org/10.1007/978-0-387-34847-6_3
- Hammond, T., Pasin, M. and Theodoridis, E. (2017) 'Data Integration and Disintegration: Managing Springer Nature SciGraph with SHACL and OWL', in *International Workshop on the Semantic Web*. Available at: <https://ceur-ws.org/Vol-1963/paper493.pdf>
- Hogan, A. et al. (2021) *Knowledge Graphs*. Springer (Synthesis Lectures on Data, Semantics, and Knowledge). Available at: <https://doi.org/10.2200/S01125ED1V01Y202109DSK022>
- Hussein, H. et al. (2022) 'KGMM – A Maturity Model for Scholarly Knowledge Graphs Based on Intertwined Human-Machine Collaboration', in Y.-H. Tseng, M. Katsurai, and H.N. Nguyen (eds.) *From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries*. Cham: Springer International Publishing, pp. 253–269. Available at: https://doi.org/10.1007/978-3-031-21756-2_21
- Hussein, H. et al. (2023) 'Increasing Reproducibility in Science by Interlinking Semantic Artifact Descriptions in a Knowledge Graph', in D.H. Goh, S.-J. Chen, and S. Tuarob (eds.) *Leveraging Generative Intelligence in Digital Libraries: Towards Human-Machine Collaboration*. Singapore: Springer Nature Singapore, pp. 220–229. Available at: https://doi.org/10.1007/978-981-99-8088-8_19
- Jaradeh, M.Y. et al. (2019) 'Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge', in *Proceedings of the 10th International Conference on Knowledge Capture*. New York, NY, USA: Association for Computing Machinery (K-CAP '19), pp. 243–246. Available at: <https://doi.org/10.1145/3360901.3364435>
- Jaradeh, M.Y., Stocker, M. and Auer, S. (2020) 'Question Answering on Scholarly Knowledge Graphs', in M. Hall et al. (eds.) *Digital Libraries for Open Knowledge*. Cham: Springer International Publishing, pp. 19–32. Available at: https://doi.org/10.1007/978-3-030-54956-5_2
- Jeschke, J.M. et al. (2020) Hi-Knowledge.org. Available at: <https://hi-knowledge.org/> (Accessed: 21 January 2024).
- Kapogiannis, G. and Sherratt, F. (2018) 'Impact of Integrated Collaborative Technologies to Form a Collaborative Culture in Construction Projects', *Built Environment Project and Asset Management*, 8(1), pp. 24–38. Available at: <https://doi.org/10.1108/BEPAM-07-2017-0043>
- Karras, O. (2024) 'Analysis of Tailored Forming Process Chains in Mechanical Process Engineering'. Zenodo. Available at: <https://doi.org/10.5281/zenodo.13752753>
- Karras, O. et al. (2021) 'Researcher or Crowd Member? Why not both! The Open Research Knowledge Graph for Applying and Communicating CrowdRE Research', in *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*. Los Alamitos, CA, USA: IEEE Computer Society, pp. 320–327. Available at: <https://doi.org/10.1109/REW53955.2021.00056>
- Karras, O. et al. (2023b) "'FAIR-by-Design" Artifacts: Enriching Publications and Software with FAIR Scientific Information at the Time of Creation', in *Proceedings of the NFDI4Ing Conference 2023*

– Innovation in Research Data Management: Bridging the Gaps Between Disciplines and Opening New Perspectives for Research in Engineering Science. *NFDI4Ing*, pp. 34–35. Available at: <https://doi.org/10.5281/zenodo.10036957>

- Karras, O.** et al. (2023c) ‘Divide and Conquer the EmpiRE: A Community-Maintainable Knowledge Graph of Empirical Research in Requirements Engineering’, in *2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. Los Alamitos, CA, USA: IEEE Computer Society, pp. 1–12. Available at: <https://doi.org/10.1109/ESEM56168.2023.10304795>
- Karras, O.** et al. (2024) ‘Organizing Scientific Knowledge From Energy System Research Using the Open Research Knowledge Graph’, *arXiv preprint arXiv:2401.13365* [Preprint]. Available at: <https://doi.org/10.48550/arXiv.2401.13365>
- Karras, O., Budde, L. and Merkel, P.** (2023a) ‘Tailored Forming Process Chain for the Manufacturing of Hybrid Components with Bearing Raceways Using Different Material Combinations’. *Open Research Knowledge Graph*. Available at: <https://doi.org/10.48366/R187049>
- Karras, O. and Groen, E.C.** (2021) ‘Overview of Approaches that Classify User Feedback as Feature Request’. *Open Research Knowledge Graph*. Available at: <https://doi.org/10.48366/R112387>
- Karras, O. and Khan, J.A.** (2021) ‘Overview of Crowd Intelligence in Requirements Engineering’, *Open Research Knowledge Graph*. Available at: <https://doi.org/10.48366/R114155>
- Karras, O., Schneider, K. and Fricker, S.A.** (2020) ‘Representing Software Project Vision by Means of Video: A Quality Model for Vision Videos’, *Journal of Systems and Software*, 162, p. 110479. Available at: <https://doi.org/10.1016/j.jss.2019.110479>
- Keet, C.M.** (2013) ‘Open World Assumption’, in W. Dubitzky et al. (eds.) *Encyclopedia of Systems Biology*. New York, NY: Springer New York, pp. 1567–1567. Available at: https://doi.org/10.1007/978-1-4419-9863-7_734
- Kejriwal, M.** (2022) ‘Knowledge Graphs: A Practical Review of the Research Landscape’, *Information*, 13(4). Available at: <https://doi.org/10.3390/info13040161>
- Khan, J.A.** et al. (2019) ‘Crowd Intelligence in Requirements Engineering: Current Status and Future Directions’, in E. Knauss and M. Goedicke (eds.) *Requirements Engineering: Foundation for Software Quality*. Cham: Springer International Publishing, pp. 245–261. Available at: https://doi.org/10.1007/978-3-030-15538-4_18
- Kitchenham, B.A. and Pfleeger, S.L.** (2008) ‘Personal Opinion Surveys’, in F. Shull, J. Singer, and D.I.K. Sjøberg (eds.) *Guide to Advanced Empirical Software Engineering*. London: Springer London, pp. 63–92. Available at: https://doi.org/10.1007/978-1-84800-044-5_3
- Knoll, C.** (2022) ‘Examining the ORKG towards Representation of Control Theoretic Knowledge – Preliminary Experiences and Conclusions’, in *Companion Proceedings of the Web Conference 2022*. New York, NY, USA: Association for Computing Machinery (WWW ’22), pp. 810–817. Available at: <https://doi.org/10.1145/3487553.3524661>
- Knublauch, H. and Kontokostas, D.** (2017) Shapes Constraint Language (SHACL). W3C. Available at: <https://www.w3.org/TR/2017/REC-shacl-20170720/>
- Kruse, J.** et al. (2019) ‘Cross-wedge rolling of PTA-welded hybrid steel billets with rolling bearing steel and hard material coatings’, *AIP Conference Proceedings*, 2113(1), p. 040019. Available at: <https://doi.org/10.1063/1.5112553>
- Kullmann, F.** et al. (2021) ‘Comparison of Studies on Germany’s Energy Supply in 2050’. *Open Research Knowledge Graph*. Available at: <https://doi.org/10.48366/r153801>
- Manghi, P.** et al. (2019) ‘The OpenAIRE Research Graph Data Model’. *Zenodo*. Available at: <https://doi.org/10.5281/zenodo.2643199>
- Martin, L. and Henrich, A.** (2022) ‘RDFtex: Knowledge Exchange Between LaTeX-Based Research Publications and Scientific Knowledge Graphs’, in G. Silvello et al. (eds.) *Linking Theory and Practice of Digital Libraries*. Cham: Springer International Publishing, pp. 26–38. Available at: https://doi.org/10.1007/978-3-031-16802-4_3
- Mozgova, I.** et al. (2022) ‘Knowledge Annotation within Research Data Management System for Oxygen-Free Production Technologies’, *Proceedings of the Design Society*, 2, pp. 525–532. Available at: <https://doi.org/10.1017/pds.2022.54>
- Nieße, A.** et al. (2022) ‘NFDI4Energy – National Research Data Infrastructure for the Interdisciplinary Energy System Research’. *Zenodo*. Available at: <https://doi.org/10.5281/zenodo.6772013>
- Oelen, A.** et al. (2020) ‘COVID-19 Reproductive Number Estimates’. *Open Research Knowledge Graph*. Available at: <https://doi.org/10.48366/R44930>
- Pape, F.** et al. (2018) ‘Tribological Study on Tailored-Formed Axial Bearing Washers’, *Tribology Online*, 13(6), pp. 320–326. Available at: <https://doi.org/10.2474/trol.13.320>
- Papers With Code** (2020) *Papers With Code*. Available at: <https://paperswithcode.com/about> (Accessed: 21 January 2024).
- Papke, T.** et al. (2018) Bulk Metal Forming of Additively Manufactured Elements’, in F. Vollertsen et al. (eds.) *MATEC Web of Conferences*, 190, p. 03002. Available at: <https://doi.org/10.1051/mateconf/201819003002>

- Papke, T. et al.** (2023) 'Alternating Additive Manufacturing and Forming—An Innovative Manufacturing Approach', *Journal of Manufacturing and Materials Processing*, 7(3), p. 90. Available at: <https://doi.org/10.3390/jmmp7030090>
- Penev, L. et al.** (2019) 'OpenBiodiv: A Knowledge Graph for Literature-Extracted Linked Open Data in Biodiversity Science', *Publications*, 7(2), p. 38. Available at: <https://doi.org/10.3390/publications7020038>
- Prakash, A. and Sandfeld, S.** (2018) 'Chances and Challenges in Fusing Data Science with Materials Science', *Practical Metallography*, 55(8), pp. 493–514. Available at: <https://doi.org/10.3139/147.110539>
- Priem, J., Piwowar, H. and Orr, R.** (2022) 'OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts', *arXiv preprint* [Preprint]. Available at: <https://doi.org/10.48550/arXiv.2205.01833>
- Robson, C. and McCartan, K.** (2016) *Real World Research*. John Wiley & Sons.
- Rossenova, L. et al.** (2023) 'KGI4NFDI: Knowledge graph infrastructure for the German national research data infrastructure'. *Zenodo*. Available at: <https://doi.org/10.5281/zenodo.8337432>
- Runeson, P., Engström, E. and Storey, M.-A.** (2020) 'The Design Science Paradigm as a Frame for Empirical Software Engineering', in M. Felderer and G.H. Travassos (eds.) *Contemporary Empirical Methods in Software Engineering*. Cham: Springer International Publishing, pp. 127–147. Available at: https://doi.org/10.1007/978-3-030-32489-6_5
- Runnwerth, M., Stocker, M. and Auer, S.** (2020) 'Operational Research Literature as a Use Case for the Open Research Knowledge Graph', in A.M. Bigatti et al. (eds.) *Mathematical Software—ICMS 2020*. Cham: Springer International Publishing, pp. 327–334. Available at: https://doi.org/10.1007/978-3-030-52200-1_32
- Santos, R., Groen, E.C. and Villela, K.** (2019) 'An Overview of User Feedback Classification Approaches', in *REFSQ Workshops*, pp. 357–369. Available at: https://ceur-ws.org/Vol-2376/NLP4RE19_paper11.pdf
- Schimmler, S.** (2023) 'NFDI4DataScience: National Research Data Infrastructure for Data Science and AI'. Available at: <https://www.dfg.de/resource/blob/174308/e173d5d18c9e740d2b1e830280ad3b30/nfdi4datascience-abstract-data.pdf>
- Schindler, D. et al.** (2021) 'SoMeSci - A 5 Star Open Data Gold Standard Knowledge Graph of Software Mentions in Scientific Articles', in *30th ACM International Conference on Information & Knowledge Management*. New York, NY, USA: Association for Computing Machinery (CIKM '21), pp. 4574–4583. Available at: <https://doi.org/10.1145/3459637.3482017>
- Schindler, D. et al.** (2022) 'The Role of Software in Science: A Knowledge Graph-Based Analysis of Software Mentions in PubMed Central', *PeerJ Computer Science*, 8, p. e835. Available at: <https://doi.org/10.7717/peerj-cs.835>
- Schindler, D., Zapilko, B. and Krüger, F.** (2020) 'Investigating Software Usage in the Social Sciences: A Knowledge Graph Approach', in A. Harth et al. (eds.) *The Semantic Web*. Cham: Springer International Publishing, pp. 271–286. Available at: https://doi.org/10.1007/978-3-030-49461-2_16
- Schirrwagen, J. et al.** (2013) 'Data Curation in the OpenAIRE Scholarly Communication Infrastructure', *Information Standards Quarterly*, 25(3), pp. 13–19. Available at: <https://doi.org/10.3789/isqv25no3.2013.03>
- Schmitt, R.H. et al.** (2020) 'NFDI4Ing - The National Research Data Infrastructure for Engineering Sciences'. *Zenodo*. Available at: <https://doi.org/10.5281/zenodo.4015201>
- Schröcker, K. et al.** (2022) 'Mechanical properties of the LMD-processed material Ferro55 in as-built and heat-treated conditions', *Procedia CIRP*, 111, pp. 228–232. Available at: <https://doi.org/10.1016/j.procir.2022.08.055>
- Seaborne, A. and Harris, S.** (2013) 'SPARQL 1.1 Query Language', *W3C*. Available at: <https://www.w3.org/TR/2013/REC-sparql11-query-20130321/>
- Sheveleva, T. et al.** (2020) 'Development of a Domain-Specific Ontology to Support Research Data Management for the Tailored Forming Technology', *Procedia Manufacturing*, 52, pp. 107–112. Available at: <https://doi.org/10.1016/j.promfg.2020.11.020>
- Spadaro, G. et al.** (2022) 'The Cooperation Databank: Machine-Readable Science Accelerates Research Synthesis', *Perspectives on Psychological Science*, 17(5), pp. 1472–1489. Available at: <https://doi.org/10.1177/17456916211053319>
- Stocker, M. et al.** (2022) 'SKG4EOSC - Scholarly Knowledge Graphs for EOSC: Establishing a backbone of knowledge graphs for FAIR Scholarly Information in EOSC', *Research Ideas and Outcomes*, 8, p. e83789. Available at: <https://doi.org/10.3897/rio.8.e83789>
- Stocker, M. et al.** (2023) 'FAIR Scientific Information with the Open Research Knowledge Graph', *FAIR Connect*, 1(1), pp. 19–21. Available at: <https://doi.org/10.3233/FC-221513>
- Stocker, M. et al.** (2024) 'Rethinking the Production and Publication of Machine-Reusable Expressions of Research Findings'. *arXiv preprint*. Available at: <https://doi.org/10.48550/ARXIV.2405.13129>
- Tongco, M.D.C.** (2007) 'Purposive Sampling as a Tool for Informant Selection', *Ethnobotany Research and Applications*, 5, pp. 147–158. <https://ethnobotanyjournal.org/index.php/era/article/view/126>

- Uhe, J. and Behrens, B.A.** (2019) 'Manufacturing of Hybrid Solid Components by Tailored Forming', in J.P. Wulfsberg, W. Hintze, and B.-A. Behrens (eds.) *Production at the leading edge of technology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 199–208. Available at: https://doi.org/10.1007/978-3-662-60417-5_20
- van de Sompel, H. and Lagoze, C.** (2009) 'All aboard: toward a machine-friendly scholarly communication system', in T. Hey, S. Tansley, and K.M. Tolle (eds.) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, pp. 193–199. Available at: http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part4_sompel_lagoze.pdf
- Vrandečić, D. and Krötzsch, M.** (2014) 'Wikidata: a free collaborative knowledgebase', *Commun. ACM*, 57(10), pp. 78–85. Available at: <https://doi.org/10.1145/2629489>
- Wiens, V., Stocker, M. and Auer, S.** (2020) 'Towards Customizable Chart Visualizations of Tabular Data Using Knowledge Graphs', in E. Ishita, N.L.S. Pang, and L. Zhou (eds.) *Digital Libraries at Times of Massive Societal Transition*. Cham: Springer International Publishing, pp. 71–80. Available at: https://doi.org/10.1007/978-3-030-64452-9_6
- Wilkinson, M.D. et al.** (2016) 'The FAIR Guiding Principles for Scientific Data Management and Stewardship', *Scientific data*, 3(1), pp. 1–9. Available at: <https://doi.org/10.1038/sdata.2016.18>
- Zou, X.** (2020) 'A Survey on Application of Knowledge Graph', *Journal of Physics: Conference Series*, 1487(1), p. 012016. Available at: <https://doi.org/10.1088/1742-6596/1487/1/012016>

TO CITE THIS ARTICLE:

Karras, O., Budde, L., Merkel, P., Hermsdorf, J., Stonis, M., Overmeyer, L., Behrens, B.-A. and Auer, S. (2024) Organizing Scientific Knowledge from Engineering Sciences Using the Open Research Knowledge Graph: The Tailored Forming Process Chain Use Case. *Data Science Journal*, 23: 52, pp. 1–36. DOI: <https://doi.org/10.5334/dsj-2024-052>

Submitted: 01 March 2024

Accepted: 04 October 2024

Published: 06 November 2024

COPYRIGHT:

© 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.