

Knowledge-Graph Based Approach for Automated Selection of Spare Parts Suitable for Additive Manufacturing: A Railway Use-Case

Theresa Madreiter^{a,b*}, Philipp Besinger^a, Sebastian Archila^a,
Linus Kohl^{a,b}, Fazel Ansari^{a,b}

^aFraunhofer Austria Research GmbH, Vienna, Austria

^bTU Wien, Research Group of Production and Maintenance Management, Vienna, Austria

* Corresponding author. Tel.: +43676 88861621; E-mail address: theresa.madreiter@fraunhofer.at

Abstract

Spare part inventory management (SPIM) in the railway sector highly demands reliability and transparency for decentralized inventory control. Optimal SPIM should ensure the availability of needed spare parts for a service request, considering the frequency of use and criticality criterion for effective maintenance. Additive manufacturing (AM) technologies enable cost-effective production of small batch sizes often required for spare parts. However, critical component-specific information is often unstructured within engineering drawings (ED), making digital processing, and linking to existing data from enterprise resource planning (ERP) and maintenance management systems difficult. To ensure effective maintenance logistics, this paper introduces a knowledge graph (KG) that can facilitate i) interlinking multiple sources through data integration and ii) establishing a semantic data hub, thus serving as a backbone for automated assessment of component's suitability for AM. The proposed KG-based approach merges relevant (existing) ontologies, multi-structured data from ED, ERP system information, and external data sources. The approach is developed and evaluated in real-world use-cases in cooperation with the Austrian railway and public transit industry.

Keywords: Spare part inventory management; knowledge graph; additive manufacturing; railway industry

1. Introduction

Due to the long service life of trains, rail operators face the challenge of needing to ensure the availability of spare parts for their rolling stock. Critical components must be available on demand and may have a long lead time, requiring a safety stock of components at all times [1]. This leads to spare parts being the largest cost factor for maintenance costs, especially due to replacement costs, high logistic costs and supply chain inefficiencies [2]. A study on urban railway operation reveals that spare part costs are the highest influential costs with approximately 13.8% compared to life cycle costs (11%), preventive maintenance costs (6.4%) and other costs (4.6%) including workforce, corrective maintenance and cost of ownership [2]. Individual rolling stock spare parts (often critical components) are not mass-produced making them suitable for additive manufacturing (AM) [3]. Further, the costs and time of producing spare parts conventionally can dramatically increase if the original manufacturer is no longer available or if the parts need to be produced in small batches.

Optimized spare part inventory management (SPIM) is crucial for the railway industry. AM improves SPIM's efficiency by enabling decentralized print-on-demand of spare parts [4]. Literature suggests

that approximately 5-10% of manufacturing spare parts [5] and 10-15% of rail inventory [6] could be additively manufactured. However, whether spare parts are applicable for AM depends on technical specifications, the supply chain, maintenance considerations, and financial aspects [7]. Manual selection and consideration of AM suitability for spare parts is often not feasible, due to the large number of potentially suitable spare parts and diversity of AM technologies. Instead, spare part data from various heterogeneous data sources (e.g., engineering drawings (ED), enterprise resource planning (ERP) systems, domain knowledge, etc.) can deepen insights and facilitate informed decision-making. Yet, automated analysis is challenging because component-specific information is often only present in an unstructured format (e.g. non-digitalized in EDs). Further, multiple data sources are not typically semantically connected and holistically analyzable. Knowledge Graphs (KG) reinforce data integration from heterogeneous data sources and an automated selection of spare parts suitable for AM. The underlying concept of KG is based on knowledge representation methods, especially ontologies, which enables the creation of a comprehensive knowledge base in a scalable form through interoperating different ontologies. Interlinked information can be structured easily by

describing entities and their respective relationships in a machine-readable format [8]. It is worth noting that literature further emphasizes on KG capabilities not only for information search and retrieval but also for knowledge discovery and reasoning over large amounts of heterogeneous data [8].

Considering the above discussion, this paper presents a methodology for a KG-based approach to automate spare parts selection suitable for AM in the railway industry. The rest of the paper is structured as follows. Section 2 describes the state-of-the-art, focusing on the necessary measures for evaluating AM-suitable spare parts, the challenges of dealing with heterogeneous data sources, and the industrial applications of KG. Based on the identified research gap, Section 3 presents a methodology for automated spare parts selection in the railway industry, while Section 4 demonstrates its application in a specific use case. Finally, Section 5 discusses the proposed approach and future research directions.

2. State-of-the-art

2.1. Additive manufacturing in the spare parts supply chain

AM is a manufacturing process that creates objects from 3D models through the layering of materials [9]. It is applicable to a wide variety of materials, including metals, ceramics, and polymers, and can be used with different AM systems such as powder bed, powder feed, or wire feed, using energy sources such as electron beam, laser, or arc, and include the identification of suitable substitution materials [10]. This allows the creation of near-net-shaped components and complex designs and provides the possibility of customization [11]. The need to avoid downtime, lack of control over the supply chain, and difficulty in predicting inventory needs are all factors that can enable AM in the spare part supply chain [12]. For the identification of spare parts suitable for AM, several approaches exist. Cardeal et al. propose a multi-criteria classification system that allows an efficient pre-selection of relevant spare parts based on historical orders, demand frequency, price, criticality criteria (e.g., cost of losing sales), and lead time. They also mention that there often is a lack of information, especially for pre-processed, comprehensive information sources, regarding the technical specifications of a spare part as well as historical demand and lead time data [12]. Heinen et al. suggest that factors such as low demand, high fixed costs, and minimum order requirements associated with conventional manufacturing methods serve as catalysts for adopting AM [13]. They also point out that AM can significantly reduce supply chain costs,

lower minimum order quantities, and a streamlined inventory management process [13]. Despite advantages, they also acknowledge the complexities involved in assessing the suitability of AM, primarily due to the limited comprehensive data availability [13]. This issue is highlighted by Chaudhuri et al, who used structured interviews to generate and collect data. In particular, they emphasized that the lack of data often hinders the accurate screening and evaluation of parts [7].

2.2. Challenges of multi-structured data

Heterogeneous data is an essential driver for many aspects of manufacturing (i.e., informed decision-making, quality optimization, or product design) [14]. However, the usage of heterogeneous data sources also comes with challenges. On top of all, data is typically acquired from multiple data sources and is multi-structured (i.e., structured and unstructured). Therefore, it is often noisy, incomplete, erroneous, and inconsistent, which makes data pre-processing essential [8]. Further, due to the multiple sources (i.e., sensors, ERP-systems, the internet, etc.), data can be acquired through diverse options, which raises demand for comprehensive data storage systems [15]. Based on an extensive literature survey, Li et al. suggest that dealing with multi-structured data from various data sources requires pre-processing steps according to the data structure. At the same time, also a comprehensive platform should be provided, where available data can be synchronized and semantically connected for decision-making [16]. To address this issue Ansari et al. propose a multiple-level process that combines natural language processing (NLP) and ontologies to deal with multi-structured data and the respective storing for a competence-based maintenance planning approach [8]. NLP is also used by Mao et al. to tackle the challenge of multi-structured data while developing a process safety KG for the chemical industry that orchestrates process data [17].

2.3. Knowledge graphs (KG)

The concept of KG represents the acquisition and integration of information into an ontology and the application of a reasoner to infer new knowledge exploitatively or exploratively. Through reasoning, new knowledge can be generated from the stored information [18]. While the term KG was originally introduced as an umbrella term for graph-based knowledge base technologies [19], it has become increasingly popular in various domains, including manufacturing, where the term Industrial Knowledge Graph (IKG) was introduced [20]. Zhou et al. use a KG to provide accurate information for

decision support in resource allocation and emphasize the need for unified information modeling to extract semantic information from massive information data [21]. Zheng et al. establish a KG that incorporates manufacturing information from multiple data sources and links them systematically and semantically to enable quick retrieval of information [22]. While Ko et al. even propose a machine learning and KG design rule construction for AM [23].

2.4. Research gap

Focusing on SPIM optimization in the railway industry, literature shows that a lot of publications have already established technical and economic criteria for the suitability of spare parts for AM. However, the information required for such evaluations is often widely distributed across multiple heterogeneous data sources. To use available information comprehensively, it is essential i) to employ data pre-processing to handle multi-structured data sources, and ii) KG to establish semantic data storage (hub) solutions. The contribution of this paper is to introduce a KG methodology, to facilitate well-informed decision-making in the assessment of AM applicability to spare parts.

3. Methodology for the implementation of a semantic data hub

The proposed methodology for developing a semantic data hub to determine the applicability of AM for spare parts utilizes multi-sourced and multi-structured data. It includes three stages: i) data acquisition, ii) data pre-processing, and iii) ontology modeling and KG creation, as displayed in Fig. 1.

3.1. Data acquisition

The data acquisition stage aims to collect relevant data for the evaluation of whether a spare part is applicable for AM. Relevant technical parameters include the size and weight of the parts, as well as the material of the spare part. This is because only a limited number of materials can be substituted by AM materials, and this substitution further depends on the specific AM system, as described in the state-of-the-art [7]. From an economic point of view, financial aspects, lead time, demand patterns as well as the availability of suppliers should be taken into account too [7]. Therefore, data from multiple sources and of multiple structures need to be considered. In this paper (cf. Section 4 – Use Case), enterprise resource planning (ERP) data, engineering drawings (EDs) and already existing ontologies (e.g. EMMO [24], MMOY [25], Material

selection ontology [26], etc.) are in focus because those data sources typically include most of the relevant information for a comprehensive evaluation and selection. The ERP system provides comprehensive information on spare parts consumption, potential suppliers, associated price details, expected delivery times, and current stock levels. However, it often lacks details on the actual spare part, such as material information, weight, or size. This information is an integral part of the AM suitability assessment. Therefore, also (scanned) EDs are considered to obtain more detailed information. Notably, non-digitalized EDs are incorporated using state-of-the-art optical character recognition (OCR) tools.

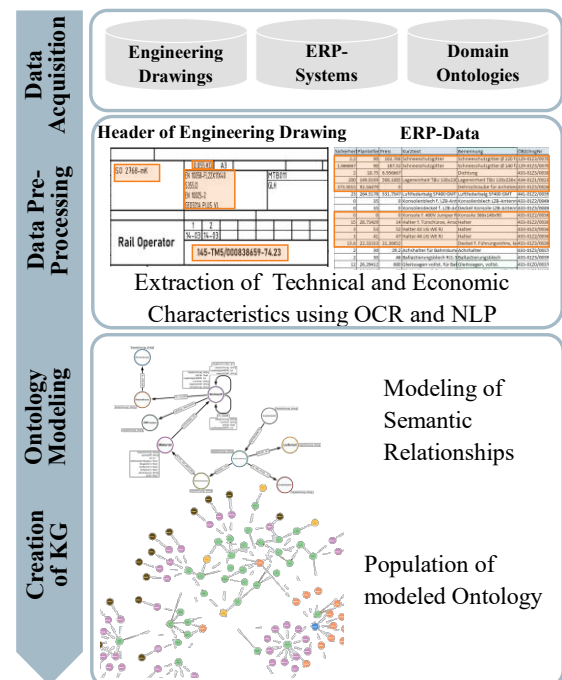


Fig. 1: KG implementation methodology

3.2. Data pre-processing

To transform available data into a format for AM-suitability evaluation, data pre-processing should be applied. The data pre-processing methods depend on the source and structure of the data. Since ERP-data is already in a structured format, only basic filter criteria must be applied. To extract information from the EDs header (i.e., identification number, material, weight, DIN standards, etc.) and transform it into a more structured format, OCR combined with NLP is used. OCR (e.g. through mmocr [27]) enables the extraction of all textual elements within the header of the ED. This process involves detecting lines and cells within the headers to isolate text from each table cell individually. Next, NLP (e.g., via Spacy [28]) allows accurate interpretation of these

extracted strings. To do this, the text from each cell is tokenized into words and phrases for detailed processing. Classification algorithms in the form of named entity recognition are used to categorize critical information such as materials, weights, and identification numbers. This structured data is then matched to ERP data using the ED identification for seamless integration. In the final stage, data errors are mitigated by cross-verifying outputs with existing material databases and DIN norms, supplemented by manual verification.

3.3. Ontology modelling and population of the KG

The modeled ontology represents the foundation for the KG. For the data modeling, each node represents an entity relevant to the AM-suitability evaluation process. Then, relationships between the defined entities are established. It also considers constraints based on technical domain knowledge, such as which AM systems require specific AM materials. Relevant properties are assigned to nodes and relationships. This enables inferences between different entities within the KG.

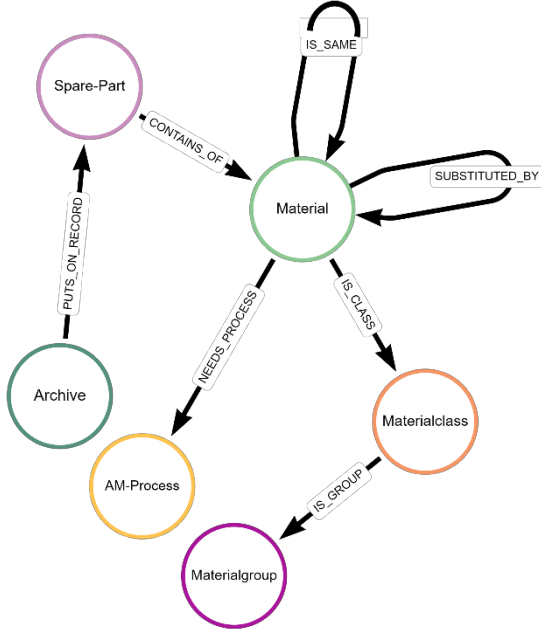


Fig. 2: Excerpt of the modelled ontology (meta-level)

Fig. 2 shows how to, e.g., infer a possible AM process from a spare part. Then, in the data acquisition phase, identified relevant ontologies can be integrated into the data model. It should be considered that the collected data adheres to constraints and rules defined within the ontology. The created ontology is then populated with filtered data from the ERP-system and the pre-processed data from the EDs.

4. Use case from the railway industry

The effectiveness of the previously proposed methodology is validated with two Austrian railway operators facing the challenges of limited spare parts availability and high costs due to dependence on conventional manufacturing. This approach addresses the relevant issue of inefficient identification of required parts resulting from fragmented and heterogeneous data sources. For both operators, spare part data for the past 60 years is available, including ERP-data, as well as corresponding EDs. The ERP-data contains information on spare part consumption, potential suppliers, associated price details, expected delivery times and current stock levels, the identification number of the spare part as well as the respective ID of the corresponding engineering drawing. For newer parts, the ERP data even provides details on sizing, weight, and material. However, due to data collection issues, this information is often neither accurate nor complete. Some parts are entered into the systems without full knowledge, resulting in entries that are both incomplete and occasionally incorrect. This data is especially missing for older spare parts. The missing information is extracted from EDs linked to ERP data by a unique identification number. Essential details that can be extracted are, e.g., sizing, material, and weight of the spare part. During the data pre-processing phase of ERP-data, spare parts that do not have a unique identification number are removed. In addition, parts that are unsuitable for AM, based on their description (e.g. glass or pictograms) and likewise standardized parts are excluded as well. For the pre-processing phase of the EDs, relevant information is extracted, by using the OCR-open source tool kit mmocr [27]. To minimize errors, the OCR network used for the string extraction has been fine-tuned with domain and language-specific characters and expressions to reduce error rates.

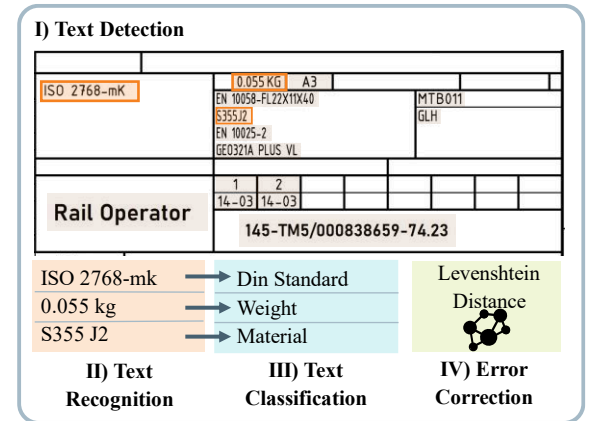


Fig. 3: Data pre-processing of engineering drawing

Then a multi-level approach (see Fig. 3) has been applied. In the first step, text detection is performed on the ED. In the second step, the text recognition is used to extract strings. It's important to note that these extracted strings are not yet mapped to their corresponding metadata, such as material or weight, so a Support Vector Machine (SVM) algorithm is used to classify the tokenized OCR output based on its structural patterns. This classification is feasible because attributes such as weight, material, and identification number typically follow a distinct format in the processed EDs. However, OCR results may contain recognition errors. Therefore, in the final step, the Levenshtein distance, a measure of the similarity between two strings [29], matches the extracted output to materials from a material list or to the most similar ED identification number, eliminating possible errors caused by the OCR algorithm.

The ontology model has been then built based on domain expert knowledge considering the assessment criteria for identifying the suitability of spare parts for additive manufacturing, including materials, alternative material substitutions, and associated AM processes. This is achieved by using OWL-design rules. In addition, the domain expertise required to assess the suitability of a part for AM accurately has been considered in the modeling, covering semantic relationships.

The KG was then created by populating the modeled ontology with the pre-processed and merged data from the ERP-system and the EDs, using the open-source KG creation tool neo4j [29]. The established KG enables reasoning on whether a spare part is eligible for AM by already being able to present details on possible substitute materials, suitable AM processes, as well as further economic details on pricing and delivery times of the conventional manufacturing method. A simple reasoning rule would be “If a spare part is composed of material M1 that isn't suitable for additive manufacturing but can be substituted by material M2, and M2 requires an available AM_Process P1, then the spare part, when made of M2, becomes feasible for AM”. Fig. 4 depicts an excerpt of the KG, showing the mentioned reasoning rule.

This approach provides sufficient details to make an initial decision on whether the spare part is suitable for AM. However, the implemented methodology has limitations. The primary challenge to the method's effectiveness is the accuracy and completeness of the ERP data, where inaccuracies and missing entries can lead to inaccurate knowledge graph outputs. In addition, relying on OCR to extract data from engineering drawings introduces further risk of error, particularly with complex or poor-quality images, despite error

correction measures such as SVM classification and Levenshtein distance checks.



Fig. 4: Populated KG

5. Conclusion and outlook

The proposed KG methodology provides a reproducible and scalable approach to process and integrate multiple data sources into a semantic data hub and to discover spare parts suitable for AM automatically. As a result, SPIM's efficiency is improved, enabling the use of AM for certain spare parts on demand. Prior to the application, for each spare part, the corresponding ERP data entry had to be checked. Further, the associated ED had to be connected and then be evaluated manually. The KG approach demonstrates the potential benefits of employing a semantic data hub in industrial applications. This method is anticipated to significantly decrease the time required to identify spare parts suitable for AM, with time reductions of 50-70% of identification achieved by automating this process. Furthermore, costs are reduced if parts suitable for AM can be identified because traditional manufacturing methods are often more costly, especially when only spare parts are manufactured in small quantities, which is often the case for spare parts. Hueng et al. state a cost reduction between 15-

35% [30], using AM. The proposed approach can be transferred to other industries, mainly when machinery is used for a long time. Future research will involve the advancement of the proposed concept by including more detailed information, i.e., characteristics like water and temperature resistance, strength-to-weight ratio and stiffness-to-weight ratio, or building speed and layer thickness, as mentioned by Chaudhuri et al. that currently still have to be evaluated manually [7]. This knowledge is only present from industry experts. Further, an automated recommender system could be established that automatically recommends spare parts suitable for AM based on the information present in the semantic data hub. This approach can be implemented in a maintenance planning system in an advanced version, enabling on-demand part production and predictive maintenance strategies.

Acknowledgments

The authors would like to thank and acknowledge Austrian Research Funding Agency (FFG) for supporting this research through AM4Rail project.

References

- [1] Mohamad Idris MF, Saad NH, Yahaya MI, Wan Mohamed WM, Shuib A, Mohamed Amin AN. Factor and Influential Variables Affecting Maintenance Cost of Urban Rail Rolling Stock. *IJIE* 2023;15(1).
- [2] Mohamad Idris MF, Saad NH, Yahaya MI, Shuib A, Wan Mohamed WM, Mohamed Amin AN. Cost of Rolling Stock Maintenance in Urban Railway Operation: Literature Review and Direction. *JST* 2022;30(2):1045–71.
- [3] Toth AD, Padayachee J, Mahlatji T, Vilakazi S. Report on case studies of additive manufacturing in the South African railway industry. *Scientific African* 2022;16:e01219.
- [4] Sgarbossa F, Peron M, Lolli F, Balugani E. Conventional or additive manufacturing for spare parts management: An extensive comparison for Poisson demand. *International Journal of Production Economics* 2021;233:107993.
- [5] Binkhuysen AB, Hermesen N, Schipper M. Additive Manufacturing for Spare Parts: Reducing Risks and Costs - Increasing Service and Functionality. *Netherlands*; 2020.
- [6] Pannett L. Supercharged: How 3D printing will drive your supply chain. Hoboken, New Jersey: John Wiley & Sons, Inc; 2019.
- [7] Chaudhuri A, Gerlich HA, Jayaram J, Ghadge A, Shack J, Brix BH, Hoffbeck LH, Ulriksen N. Selecting spare parts suitable for additive manufacturing: a design science approach. *Production Planning & Control* 2021;32(8):670–87.
- [8] Ansari F, Kohl L, Giner J, Meier H. Text mining for AI enhanced failure detection and availability optimization in production systems. *CIRP Annals* 2021;70(1):373–6.
- [9] Standard terminology for additive manufacturing technologies; 2012.
- [10] Frazier WE. Metal Additive Manufacturing: A Review. *J. of Materi Eng and Perform* 2014;23(6):1917–28.
- [11] Kumar SA, Prasad R. Basic principles of additive manufacturing: different additive manufacturing technologies. In: *Additive Manufacturing*. Elsevier; 2021. p. 17–35.
- [12] Cardeal G, Leite M, Ribeiro I. Decision-support model to select spare parts suitable for additive manufacturing. *Computers in Industry* 2023;144:103798.
- [13] Heinen JJ, Hoberg K. Assessing the potential of additive manufacturing for the provision of spare parts. *J of Ops Management* 2019;65(8):810–26.
- [14] Wang J, Xu C, Zhang J, Zhong R. Big data analytics for intelligent manufacturing systems: A review. *Journal of Manufacturing Systems* 2022;62:738–52.
- [15] Pulikottil T, Estrada-Jimenez LA, Abadía JJP, Carrera-Rivera A, Torayev A, Rehman HU, Mo F, Nikghadam-Hojjati S, Barata J. Big Data Life Cycle in Shop-Floor-Trends and Challenges. *IEEE Access* 2023;11:30008–26.
- [16] Li C, Chen Y, Shang Y. A review of industrial big data for decision making in intelligent manufacturing. *Engineering Science and Technology, an International Journal* 2022;29:101021.
- [17] Mao S, Zhao Y, Chen J, Wang B, Tang Y. Development of process safety knowledge graph: A Case study on delayed coking process. *Computers & Chemical Engineering* 2020;143:107094.
- [18] Ehrlinger L, Wolfram W. Towards a Definition of Knowledge Graphs. *International Conference on Semantic Systems* 2016.
- [19] Färber M, Basil E, Menne C, Rettinger A, Bartscherer F. Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web Journal* 2016.
- [20] He L, Jiang P. Manufacturing Knowledge Graph: A Connectivism to Answer Production Problems Query With Knowledge Reuse. *IEEE Access* 2019;7:101231–44.
- [21] Zhou B, Bao J, Li J, Lu Y, Liu T, Zhang Q. A novel knowledge graph-based optimization approach for resource allocation in discrete manufacturing workshops. *Robotics and Computer-Integrated Manufacturing* 2021;71:102160.
- [22] Zheng P, Xia L, Li C, Li X, Liu B. Towards Self-X cognitive manufacturing network: An industrial knowledge graph-based multi-agent reinforcement learning approach. *Journal of Manufacturing Systems* 2021;61:16–26.
- [23] Ko H, Witherell P, Lu Y, Kim S, Rosen DW. Machine learning and knowledge graph based design rule construction for additive manufacturing. *Additive Manufacturing* 2021;37:101620.
- [24] EMMC & EMMO Governance Committee. EMMO: Elementary Multiperspective Material Ontology; Available from: <https://github.com/emmo-repo/EMMO>.
- [25] Zhang X, Pan D, Zhao C, Li K. MMOY: Towards deriving a metallic materials ontology from Yago. *Advanced Engineering Informatics* 2016;30(4):687–702.
- [26] Zhang Y, Luo X, Zhao Y, Zhang H. An ontology-based knowledge framework for engineering material selection. *Advanced Engineering Informatics* 2015;29(4):985–1000.
- [27] MMOCR: Welcome to MMOCR's documentation! - MMOCR 1.0.1 documentation; Available from: <https://mmocr.readthedocs.io/en/dev-1.x/> (29 October 2023).
- [28] spaCy · Industrial-strength Natural Language Processing in Python; Available from: <https://spacy.io/> (29 April 2024).
- [29] Graph Database & Analytics. Neo4j Graph Database & Analytics – The Leader in Graph Databases; Available from: <https://neo4j.com/> (29 October 2023).
- [30] Huang R, Riddle ME, Graziano D, Das S, Nimbalkar S, Cresko J, Masanet E. Environmental and Economic Implications of Distributed Additive Manufacturing: The Case of Injection Mold Tooling. *J of Industrial Ecology* 2017;21(S1).