Technical Report

Industry 4.0 Asset-Based Risk Mitigation for Production Operation

Dietmar Winkler[1]
Petr Novak[2]
Jiri Vyskocil[2]
Kristof Meixner[1]
Stefan Biffl[1]

[1] CDL-SQI, TU Wien, Austria
[first.last]@tuwien.ac.at

[2] Czech Institute of Informatics, Robotics, and Cybernetics, Czech Technical University, Czech Republic
[first.last]@cvut.cz

Technical Report No. CDL-SQI 2021-02
Issued: March 2021

Citation: D. Winkler, P. Novak, J. Vyskocil, K. Meixner, S. Biffl “Industry 4.0 Asset-Based Mitigation for Production Operation”, Technical Report CDL-SQI 2021-02, TU Wien, Vienna, Austria, March 2021, submitted to Robotics and Automation Letter (RA-L) and CASE 2021 (under review)

Industry 4.0 Asset-based Risk Mitigation for Production Operation

Dietmar Winkler[1], Petr Novák[2], Jiří Vyskočil[2], Kristof Meixner[1], Stefan Biffl[1]

Abstract— **During engineering and operation of flexible robot-based production systems meeting the Industry 4.0 (I40) paradigm, users require guidance to analyze and resolve issues that may disturb the production process. Challenging issues stem from causes in several heterogeneous engineering disciplines. Unfortunately, current risk mitigation guidelines frequently focus isolated components and not on the risk of the entire system. This fragmented guidance is hard to apply for users who are not aware of existing dependencies between components of different types. Therefore, risks in the engineering and operation of I40 components are hard to identify and mitigate. In this paper, we propose the *Industry 4.0 Asset-based Risk Mitigation* (I4ARM) approach, providing knowledge for efficient root cause analysis to non-expert users based on (a) a minimal model for knowledge representation as an I40 asset network with cause-effect annotations and (b) the I4ARM method for model building and risk mitigation with structured guidance. We build on the I40 asset network concept, cause-effect analysis, and decision trees to enable efficient and effective risk mitigation with structured guidance. I4ARM facilitates for engineers (a) defining an *Industry 4.0 asset network* and relationships, (b) identifying risks, and (c) supporting risk mitigation. We conceptually evaluate I4ARM for a real-world I40 use case. The results showed that the I40 Asset Network with Cause-Effect Relationships and Decision Trees is usable and useful both for experienced and novice users to efficiently and systematically mitigate risks in I40 environments.**

I. Introduction

Industry 4.0 (I40) and *Cyber-Physical Production Systems (CPPSs)* aim at addressing business demands for flexible production in terms of volume and product variants [17], [31]. Modern robot-based production systems, such as the *Testbed for Industry 4.0* hosted at CTU[3], are flexible for adaptation to automate changing production processes with production system components and assets [15]. An I40 asset can be a physical or a logical asset, such as a production process, described by an Asset Administration Shell [1] that can collect and provide integrated knowledge during engineering and operation [5].

*The financial support by the Christian Doppler Research Association, the Austrian Federal Ministry for Digital & Economic Affairs and the National Foundation for Research,Technology and Development is gratefully acknowledged. This result was also funded by Ministry of Education, Youth and Sport of the Czech Republic within the project Cluster 4.0, reg. number CZ.02.1.01/0.0/0.0/16_026/0008432.

[1]Dietmar Winkler, Kristof Meixner, and Stefan Biffl are with the Christian Doppler Laboratory for Security and Quality Improvement in the Production System Lifecycle, Institute for Information Systems Engineering, TU Wien, Vienna, Austria (email: dietmar.winkler@tuwien.ac.at; kristof.meixner@tuwien.ac.at; stefan.biffl@tuwien.ac.at).

[2]Petr Novák and Jiří Vyskočil are with the Department of Intelligent Systems for Industry and Smart Distribution Networks, Czech Institute of Informatics, Robotics, and Cybernetics, Czech Technical University in Prague, Prague 160 00, Czech Republic (e-mail: petr.novak@cvut.cz; jiri.vyskocil@cvut.cz).

[3]Industry 4.0 Testbed: ciirc.cvut.cz/teams-labs/testbed/

However, in an I40 asset network, the interaction of production systems components [17] may have risky effects, like imprecise product placement by a robot, leading to issues in the production or even to a process failure that requires intervention by human operators. Therefore, domain experts have to analyze the risks and issues of the component-based system during engineering and operation.

The concept of I40 assets enables full digital monitoring while performing all actions on these components [15], [31]. In addition, the digital twin/shadow [9] of these I40 assets can hold data and models to facilitate analysis on component level during engineering and operation. I40 assets (data and models), forming the entire I40 production system, can be seen as an I40 asset network that facilitates knowledge and technology transfer from engineering to practice [15]. This network can support engineers in error handling during performing specific operations on concrete I40 components. For example, if a shuttle stops moving, the entire sequence of consecutive operations, which depend on this shuttle, needs to be canceled and an operator should be notified to address this issue. However, it is not clear how to leverage the knowledge in an I40 asset network for guiding issue analysis.

Risk analysis and mitigation [13] and vendor-specific guidelines currently focus on the risk of a specific component, providing only fragmented guidance for users. As these approaches often do not focus on the entire system, they are hard to use for novice users, who are not aware of dependencies between components in the I40 asset network [5]. In this context, we identified two main challenges.

Challenge 1. *Single-discipline engineering plans make it hard to understand and mitigate multi-disciplinary risks*. An example risk is *imprecise position of product*, which may be caused by the robot program (IT), temperature (environment), or voltage (electrics), designed by different engineers and domains. There is currently no successful approach to represent multi-disciplinary risk mitigation guidance for robot-based production systems, which build on I40 assets.

Challenge 2. *No clear method to analyze issues, learn from previous problems, and mitigate risks.* Performing a *Root Cause Analysis (RCA)* [22] is difficult with a limited amount of data and without a multi-disciplinary cause-effect model: (a) RCA in operation is difficult for an inexperienced user without guidance and (b) providing relevant guidance takes time to gain experience with a novel/changed CPPS [31], even for an expert.

Main goal of this paper is to improve the efficiency of multi-disciplinary risk mitigation in the engineering and operation of an automated robot-based production system (a) by integrating engineering data of the CPPS and for a scope

of risky effects to define possible and most likely *Cause-Effect* (C-E) dependencies and root causes based on I40 asset concepts [5] (cf. Figure 2); and (b) by structuring risk analysis questions and risk mitigation guidelines in a *Decision Tree* [23], based on the C-E information for selected, prioritized, and currently identified risks (cf. Figure 3).

This paper introduces the *I40 Asset-based Risk Mitigation* (I4ARM) approach, which is motivated by lessons learned from working with domain experts and unskilled users of the *Testbed for Industry 4.0* at CTU. The I4ARM builds on the concepts of the *I40 Asset Network* [4] and *Cause-Effect (C-E)* relationships [22] that together form a *Cause-Effect Knowledge Graph (CEKG)* and on *Decision Trees (DTs)* [28] to support engineers in (a) defining I40 assets and relationships for risk analysis, (b) identifying risk causes during engineering and operation, and (c) supporting efficient risk mitigation by guiding users in identifying and mitigating risky effects.

We evaluate the proposed approach to operational risk mitigation on a simplified real-world example originating from the *I40 Testbed* by comparing variants with and without using decision trees and *CEKGs*.

Main contributions of this paper include (a) the knowledge representation capabilities of a *CEKG* that allow representing I40 assets and relationships effectively and efficiently for risk analysis and mitigation [4] and (b) efficient derivation of a *Decision Tree*, based on a *CEKG*, for guiding risk mitigation for a robot-based production system with I40 components.

The remainder of this paper is structured as follows. Section II summarizes related work on Industry 4.0, risk management, and decision trees. Section III motivates the research questions and approach. Section IV introduces an illustrative use case for evaluation. Section V introduces the *Industry 4.0 Asset based Risk Mitigation (I4ARM)* approach and provides an example application. Section VI evaluates and discusses the I4ARM approach with the I40 Testbed. Finally, Section VII concludes and identifies future work.

II. Related Work

This section summarizes related work on CPPS and Industry 4.0, Risk Management, Cause-Effect Analysis, and Decision Trees for Risk Analysis Guidance.

A. Cyber-Physical Production Systems and Industry 4.0

The CPPS [31] in the *Industry 4.0 Testbed* (see Section IV) is a representative example of an industrial production system that incorporates a set of *I40 components*, like robots [30]. An I40 component typically includes physical and logical assets [14], like products, production processes, and resources with dependencies and technical data [1], [26]. Dependencies between I40 components within a CPPS will require increasing awareness of *multi-aspect risks* that come from *(i)* hidden complexity within this component and *(ii)* distributed domain knowledge on the interaction with I40 components [20]. CPPS engineering, following an I40 asset-oriented view, is based on multi-model digital shadows [31] in several domains like functional, mechanical, and control engineering, semantically linked by *common concepts* [3].

In this work, we build on engineering data to derive the basic I40 asset network [5]. Further, we build on the *I40 Asset Network* concept to represent the information in assets, similar to I40 components, to connect risk effects to causes in CPPS design.

B. Risk Management and Cause-Effect Analysis

Model-based risk assessment in CPPS focuses on automating process steps, e.g., as defined by the *Failure Mode and Effect Analysis* (FMEA) [29], for single-discipline models. This includes, for instance, risks from incorrect computation and timing in signal processing components [16], [21]. The analysis of risks and the identification of causes, effects, and counter measures is typically based on a modular *Cause-Effect Analysis* [10], [16], such as the *Ishikawa* process approach [12]. Such traditional approaches typically focus on investigating isolated effects, with limitations regarding multi-disciplinary views that are common in CPPS environments. However, we see the need for supporting the *Cause-Effect Analysis* in an *I40 Asset network* to focus on *I40 components* with their dependencies. Loucopoulos *et al.* [14] emphasizes addressing risky assets required for a transition towards CPPS engineering. Liu *et al.* [13] categorize applications and shortcomings of works on FMEA approaches, concluding that the examined approaches lack capabilities to address interdependencies between failure modes.

In this paper we address this issue by focusing on the *Cause-Effect Knowledge Graph (CEKG)* [4] that is based on the *Cause-Effect Analysis* supporting the systematic analysis of root causes based on observed effects [22] following the Ishikawa approach [12] but extends this approach by taking into consideration possible and likely risks that go beyond an isolated risk assessment approach. Hence, we build on the logical conjunction of risk conditions [21] and on linking *components* with *ports* [10] to formulate multi-aspect queries for a *CEKG* represented in an *I40 Asset Network* in CPPS [5].

C. Decision Trees for Risk Analysis Guidance

In CPPS, risk management during production operation requires an approach (a) to prioritize possible root causes of risky effects and (b) to provide guidance in systematically exploring causes of risky issues in the real system. For this context, the *Decision Tree (DT)* is a suitable and well-proven paradigm in *Artificial Intelligence (AI)* and *Machine Learning (ML)* [18]. *DT*s have been successfully used for the classification and regression in data mining [36]. In the CPPS context, *DT*s can be used for automatically generating guidance for identifying root causes based on *CEKG* and based on collected data coming from the entire *I40 Asset Network*, i.e., the *I40 Testbed* environment. Contrary to other AI concepts (such as artificial neural networks or inductive logic programming), algorithms for constructing *DT*s are typically deterministic, allowing to asymptotically estimate[4]

[4]The time complexity of *C4.5* is super-linear with $O(|T|log_2|T|)$ but there are even faster near-linear implementations [7].

how long the computation/construction of the *DT* will take. Therefore, we consider the *DT* approach suitable for an industrial production environment.

There are numerous algorithms for constructing *DTs* [2]. Quinlan's pioneering algorithm *ID3* (Iterative Dichotomiser 3) constructs a DT [23] on the principles of Occam's razor and minimization of information entropy in each decision node. Quinlan improved the *ID3* into the *C4.5* algorithm [24], providing the new features: discrete and continuous attributes, missing values, assignment of differing costs, and pruning trees. The last famous improvement is Quinlan's *C5.0* algorithm [25] with the advantages: several orders of magnitude faster (because of parallelism support), memory efficient, smaller *DT*s, boosting (more accuracy), ability to weight different attributes, and winnowing (reducing noise). Most improvements in *C4.5* and *C5.0* can be used directly in our work and their new features like discrete and continuous attributes and missing values match the characteristics of data collected from the *I40 Testbed*.

There also exist more complex random forest algorithms [6], based on *DT* ideas with overall better results (in classification and regression domain). However, in comparison to a standard decision tree, where users can understand the tree in a white-box way, the understanding of random forests is quite difficult, because of the much higher number of parameters and significantly more complex effort for interpretation.

In this paper, we build on the *C4.5 DT* algorithm and on information derived from the *CEKG* to select relevant variables for the design of the *DT* as foundation for deriving guidance for non-expert users.

III. Research Questions and Approach

This paper aims at improving the efficiency of multi-disciplinary risk mitigation in the engineering and operation of an automated robot-based production system, such as the *I40 Testbed*. We employ *Design Science* research [8], [35], to investigate how to improve shortcomings in the context of CPPS engineering processes. As preliminary research work, we conducted a domain analysis to elicit use cases on risk mitigation in robot-based production system engineering and operation, i.e., workshops and interviews with relevant domain expert roles in the context of the *I40 Testbed*, leading to the illustrative use case *Industry 4.0 Production Line* (cf. Section IV). We investigate the following research questions to (a) build up a *CEKG* and (b) use this graph to derive a *DT* for guiding users in issue analysis and risk mitigation:

RQ1. *What knowledge model can represent knowledge on and dependencies between I40 components and risky effects and their causes?* The commission and operation of robot-based production systems requires various discipline-specific views including related artifacts and knowledge, such as mechanical, electrical, and automation plans and experience. However, the links between these isolated views are often implicit and not expressed explicitly. The first research question focuses on a minimum set of required knowledge and experience for information and knowledge sharing for risk analysis and mitigation. We build on knowledge requirements for root cause analysis during commissioning and operation including a *I40 Asset Network* [5] that holds relevant information and knowledge usable for a cause-effect analysis with annotations as foundation for a decision tree [4]. To make the *CEKG* efficiently usable for non-expert users in a situation that requires taking decisions fast, we investigate the following RQ.

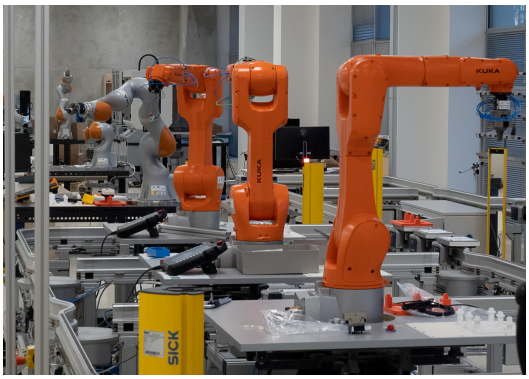

Fig. 1: The Industry 4.0 Testbed at CTU in Prague – CIIRC.

RQ2. *How can domain experts derive a Decision Tree, building on a root cause analysis for efficient risk mitigation guidance?* Although the *CEKG* can help experts in efficiently identifying root causes based on observed effects, *DTs* [18] can be derived from Cause-Effect representation to help non-experts in guiding them through the root cause analysis process. Therefore, the second research question focuses on a method to identify such a *DT* in CPPS contexts as foundation for deriving risk mitigation guidelines. This outcome improves risk mitigation efficiency by representing effect-specific user guideline decisions for operational risk analysis and risk mitigation guidelines on system level going beyond the current focus of risk mitigation on components.

We evaluate *Industry 4.0 Asset-based Risk Mitigation (I4ARM)* process approach in a real-world use case in the *I40 Testbed* by comparing the capabilities of the I4ARM approach to traditional risk mitigation design approaches in the engineering of automated robot-based production systems (cf. Section VI). We discuss the usefulness and usability of the I4ARM method results with domain experts from academia and industry in various contexts, i.e., in the manufacturing, automotive, and chemical industry domain.

IV. Use Case Industry 4.0 Production Line

This section introduces the illustrative use case *Industry 4.0 Production Line* (cf. *I40 Testbed*, Figure 1) in the context of evaluation of risk mitigation. The *I40 Testbed*[5] bridges the gap between scientific state-of-the-art and industrial practice in various domains, including advanced process control and planning [33], automated precise robot calibration [19], and the design of (collaborative) robotic working cells. The *I40 Testbed* focuses on developing and transferring new

[5]Industry 4.0 Testbed: www.ciirc.cvut.cz/teams-labs/testbed/

outcomes to industrial partners. Since significant solution knowledge is only implicit, understanding and transferring deep insight to system operation poses a challenging issue (cf. challenges C1 and C2 in Section I). We have experienced that especially recognition of possible roots of stopping or failing the production process is critical. Therefore, this paper focuses on mitigating such operational risks with the newly developed hardware and software. Table I summarizes requirements and capabilities for risk mitigation collected from domain analysis.

R1. Representation of I40 Asset network with multi-disciplinary dependencies between assets.
R2. Representation of cause-effect pathway between selected risky effects and root cause candidates in an I40 Asset network.
R3. Cause-effect driven data collection for analysis.
R4. Representation of effect-specific user guideline decisions for operational risk analysis and risk mitigation guidelines on system level.
R5. Cost-benefit driven guidelines for effective and efficient risk mitigation based on effect-specific decisions.

TABLE I: Identified Risk Mitigation Requirements.

Figure 1 presents the *I40 Testbed*, consisting of three industrial robots, *KUKA*[6] *Agilus* and one cooperative robot *KUKA iiwa*. The robots are interconnected with a transportation system *montrac*[7]. *Montrac* is a mono-rail transportation system consisting of tracks, *shuttles*, and *positioning units*, which assure exact stopping and positioning of the shuttles in specific locations, such as work cells close by robots. Shuttles are relatively autonomous components that move on tracks according to the given production plan [32]. Shuttles are equipped with electrical motors, re-programmable control units memorizing target station of the current movement, and an infra-red sensors to detect the free space in the front, available for movement. Supply of DC voltage is provided by tracks. The production line is generic with focus on the final assembly of products, including a set of basic production operations: *(i) Pick* a component from given coordinates by a *robot*; *(ii) Place* a component to given coordinates by the *robot*; *(iii) Move* a semi-product on a *shuttle*.

Although the *Testbed* is orchestrated by industrial machinery of high reliability, the production process can fail due to various reasons. For example, the robot movement to specified coordinates may not be precise enough (e.g., due to wrong calibration), or a component may drop from a robot gripper. The following section builds on this use case to demonstrate the proposed I4ARM method on a practical example from the Testbed domain.

V. Industry 4.0 Asset based Risk Mitigation

This section introduces the *I4ARM* knowledge graph and method (cf. Sections V-A and V-B), the *Cause-Effect Knowledge Graph* (cf. Figure 2) and elaborates how to derive a Decision Tree (cf. Section V-C) to structure risk mitigation guidelines on a system level, based on the *I40 Testbed* (cf. Section IV).

[6]KUKA: www.kuka.com

[7]montratec: www.montratec.de/en/

A. I4ARM Knowledge Graph for Cause-Effect Annotation

This section describes how to build the *Cause-Effect Knowledge Graph (CEKG)* and illustrates an example based on the *I40 Testbed*. The underlying meta-model [5] consists of (a) *Risk analysis scope*, (b) *I40 Assets* connected with *links* and *hypotheses*, and (c) *failure modes* related to *effects*, *risks*, and *causes* related to the risk assessment approach (see [5] for details). Domain experts coming from different disciplines identify related aspects in the *I40 Asset Network* by analyzing product, process, resources (as building blocks of the model), and extend the basic model with causes, effects, and dependencies. Figure 2 presents an example *CEKG* that consists of observable *effects*, related *production processes* and system *resources*, influenced by a set of *root causes* (see legend of Figure 2). Note that this *CEKG* (a) provides the foundation for deriving a *DT* for guiding novice users in the risk assessment approach or (b) provides added value to expert users for risk assessment.

Industry 4.0 Asset network representation. The production system consists of a set of I40 components. Such components are frequently considered as resources, but they can consist of sub-resources as well. This detailed view on I40 components can be provided by domain experts, practitioners, or by machine vendors for their specific components. This view is well aligned with the I40 concept of virtualization/data transparency [31] and the I40 Asset Administration Shell [1], which encapsulates knowledge on components as software.

Cause-effect annotation to the I40 Asset network. The representation of causes and effects related to I40 components is a foundation for the cause-effect analysis and the *CEKG*. Both, causes and effects can be organized within hierarchies and assigned to resources and sub-resources. Figure 2 shows an example *CEKG* for the *I40 Testbed*. There is a set of four effects represented as boxes in orange color. In addition, the I40 component robot concerns two further effects, I40 component *montrac* brings in three effects. Root causes are represented as boxes in violet color.

Signals that represent logical or physical variables, such as *environment temperature*, play important roles in industrial system operation and integration. In Figure 2, the *CEKG* includes signals (boxes in white color) as data sources for cause conditions and for questions in the *DT*. Since these signal variables are important for system operation, we expect these signals to be readable in the system and its components and to be logged for further analysis.

B. I4ARM Method with Cause-Effect Diagrams

To address RQ1, the *Industry 4.0 Asset-based Risk Mitigation* (I40ARM) method consists of three steps to capture knowledge on causes and effects for enabling risk mitigation guidance in a multi-disciplinary CPPS environment: (1) Design the *CEKG*; (2) Design the *Decision Tree and Guidelines*; and (3) *Validate Decision Tree and User Guidelines*.

Step 1. *Design Cause-Effect Knowledge Graph*. For a selected risk assessment scope, domain experts design (a) an *I40 Asset Network* based on I40 assets and links from engineering plans and dependencies coming from various

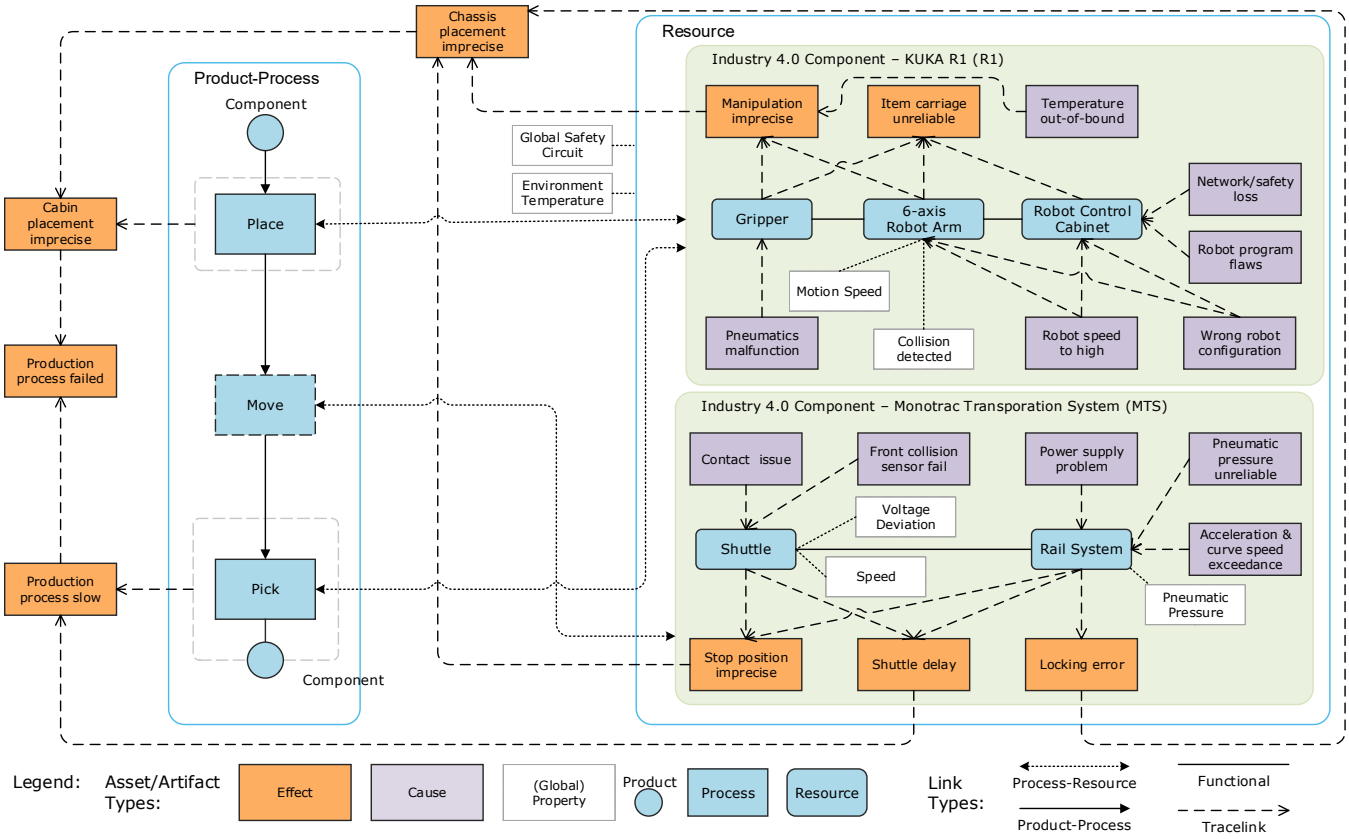

Fig. 2: Cause-Effect graph for the exemplary production process in I40 Testbed, including products (blue circles), production processes (blue boxes), and resources (blue rounded boxes); I40 components (green areas); and variables (white boxes).

domain experts and (b) for *risky effects*, e.g., imprecise product placement, domain experts can annotate elements in the I40 asset network by highlighting the Cause-Effect pathway from risky effects to root causes and counter measures [5]. Creating the I40 asset network relies on knowledge on the product (i.e., the product to be built), the process (i.e., individual steps how to build a product), and production resources (i.e., required components to construct the products). Furthermore, dependency relationships (between assets and views) need to be added to represent cause-effect paths from I40 components via process assets to effects, such as product quality or process performance. For each effect, an engineer or operator of the system obtains a set of plausible causes that can be roots of the specific effect and marks the cause-effect path from root cause to effect.

Step 2. *Design Decision Tree for Structured Guidance on selected Effects.* This process step focuses on the design of the *DT* based on the *CEKG* (cf. Figure 2), selected variables, and data sources (cf. Table II) related to selected risky effects. Following the most relevant Cause-Effect pathways (prioritization), the result of this step is a sequence of analysis questions. For each node (question) and leaf (root cause) in the *DT*, guidance is available to support non-experts in checking the system state and providing issue mitigation options. Hence, experts and non-experts can follow these suggestions to identify root causes for risky effects.

Step 3. *Apply and Validate the DT for User Guideline in the Target Scope.* During engineering and operation, non-expert users can follow these guidelines and execute resulting actions based on traceable and repeatable results and guidance. Expert users can (a) validate the results provided by the *DT* and (b) iteratively improve the *CEKG*, the *DT*, and derived guidelines.

C. I4ARM Decision Tree Design based on the CEKG

During I40 production system operation, it is crucial to quickly identify an I40 component that has the most significant impact on a concrete negative effect. Such a quick identification requires a guideline of good and targeted questions, leading to the most probable root cause.

Therefore, in I4ARM, we apply the C4.5 algorithm to make a *DT* of analysis questions. The I4ARM method relies on selected system variables (specified in the *CEKG*) whose values are used as inputs for the decision making algorithm together with specification of root causes indicated during production as outputs. *DT*s can be easily created for each individual effect in *CEKG* (i.e., set of variables and root causes) separately so they can be individually tailored for any specific purpose.

In I40, slightly more pieces of information compared to conventional systems are available and required: (i) *Expert knowledge* represented as knowledge graph (including I40

components and sub-graphs); and (ii) *Production and maintenance logs* of all I40 components (including duration, positions coming from camera/positioning systems, failure rates, response time, latency, power consumption, or environmental parameters, such as temperature or humidity). This is a huge amount of data, a so-called *Big Data* problem [11].

For successful and robust application in I4.0 environments, the following properties of *DT* creation have to be fulfilled: (1) *Input for the DT making algorithm* (both training and evaluating) should be (i) discrete values and enumerations (such as state running, idle, stopped, and failure), (ii) continuous values (physical variables such as speed, voltage, duration, position), and (iii) missing values (some value could not be measured due to the current failure, or due to ramp-up phase of the system). (2) We consider an *assumption that just one component has the most significant impact on the specific (negative) effect*. Leaves in *DTs* can have typically just one output that is in our case the most significant root cause. (3) We focus on a *single tree for each individual effect* rather than on a forest, if we consider that the tree is evaluated (and the system is maintained by a human technician rather than an automated error handling system). The single *DT* is a white box in comparison to forests or neural networks that pose grey or black boxes. From a quality perspective, white box approach is preferable to a black box approach.

Input Variables							Output Root Cause
Global System		ransportation System montra			Robot R1		
Enironment Temparature	Safety Circuit Open	Shuttle Speed	Pneumatic Pressure	Voltage std. Deviation in last 15 minutes	Collision Detected	Motion Speed	
21.6°C	yes	medium	nominal	2%	no	70%	Robot speed to high
19.6°C	yes	?	?	?	no	72%	Robot speed to high
45.6°C	no	low	nominal	1%	no	55%	Temperature out of bound
25.3°C	no	high	not nominal	3%	no	90%	Pneumatics malfunction
...	...	...	...	...	...	...	...
35.6°C	no	high	nominal	10%	no	40%	Front collision sensor fail

TABLE II: Example input data from the *I40 Testbed* for the use case *Place-Move-Pick* ("?" represents a missing value).

Training Data. The I40 Testbed can provide a range of promising training data for this particular *DT* task: (1) We assume that the *Cause-Effect Knowledge Graph* structure was prepared/designed by domain experts from numerous component descriptions (e.g., sub-graphs for robots, conveyor belts) Some of those description parts can be delivered by machine/component vendors, e.g., in the frame of I40 Asset Administration Shell [1]. From this we get the domain for *DT* output (i.e., *DT* leaves; cf. Figure 3, rendered with *GraphViz*[8] based on input data from the I40 Testbed, cf. Table II). (2) We *combine* (i) sensoric and other data from the specific production system, (ii) statistical and aggregated data related to the production system (performance per minute, failure rates per day, etc.), and (iii) available component vendor data (such as KUKA, Siemens, and B&R) across production lines, countries, and companies. These combined data can be serialized into feature vectors including columns representing individual variables with values. The variables/values whose values are not available in the respective context, are substituted by so-called missing values, playing a role of placeholders in the vector.

[8]GraphViz: graphviz.org/

VI. Evaluation, Discussion and Limitations

In this section, we conceptually evaluate the the I4ARM approach in context of the *I40 Testbed* by comparing four different variants, i.e., risk mitigation with/without a *CEKG* and with/without a derived *DT*. Therefore, we derived the following variants for evaluation: (1) *NoCEKG.NoDT.* In case of neither using *CEKG* nor *DT*, guidelines are based on experience without cause-effect network and without a decision tree. This requires long-time experience with system and typically this approach is not systematic. (2) *DT.NoCEKG.* In case of guidelines based on a *DT*, but not on a cause-effect network, it is difficult to select the right data variables. If we mix all relevant causes for example from different components, there is a high chance of false positives and noise affecting data can play a significant negative role. (3) *CEKG.NoDT.* Considering utilizing guidelines based on a *CEKG*, but not on a *DT* leads to missing statistical power of data. (4) *I4ARM.* Our approach includes a *CEKG* as a systematic cause effect network and - based on this network a derived *DT*.

Comparing the capabilities of the I4ARM approach to traditional risk mitigation design approaches in the engineering of automated robot-based production systems, we discussed with domain experts the usefulness and usability of the I4ARM method results. We discussed the quality of the *DT* and the resulting guidelines. We discussed the applicability of the knowledge model to novice and experienced users and further applications building on the *I40 Asset Network* with cause-effect annotations and the *DT*/forest.

Requirements \ Risk Mitigation Guidance Approaches	NoCEKG.NoDT	DT.NoCEKG	CEKG.NoDT	I4ARM
R1. Representation of I40 Asset network with multi-disciplinary dependencies between assets.	--	--	++	++
R2. Representation of cause-effect pathway between selected risky effects and root cause candidates in an I40 Asset network.	--	--	++	++
R3. Cause-effect driven data collection for analysis.	N/A	o	N/A	++
R4. Representation of effect-specific user guideline decisions for operational risk analysis and risk mitigation guidelines on system level.	--	+	o	++
R5. Cost-benefit driven guidelines for effective and efficient risk mitigation based on effect-specific decisions.	-	o	o	+
B1. Effectiveness of risk mitigation.	-	+	o	++
B2. Efficiency of risk mitigation.	--	o	o	++
E. Setup effort	++	-	+	+

TABLE III: Evaluation of Risk Mitigation Guidance Options.

Table III compares I4ARM to these alternative approaches regarding the requirements (i.e., R1 ...R5) introduced in Section IV, effectiveness (B1) and efficiency (B2) for risk mitigation; and set up effort for data collection and graph construction (E). The ratings in Table III have been discussed with experts from academia and industry and follow a 5-point *Likert* scale ($++$, $+$, o, $-$, $--$), where $++/--$ indicate very high/low capabilities. While there is considerable initial setup effort for creating *CEKG* and *DT*, there are limitations for risk mitigation for approaches

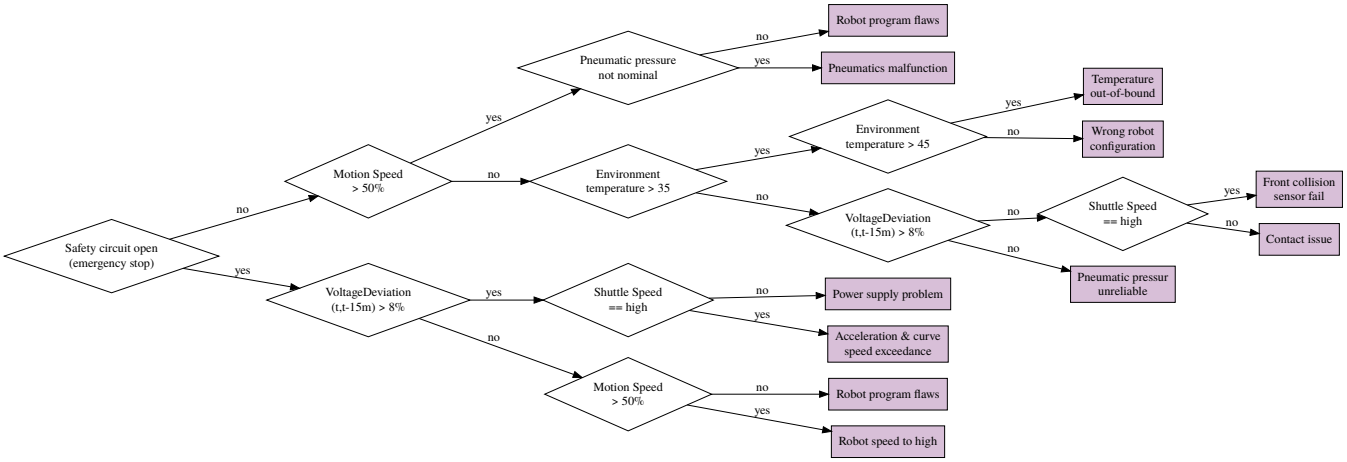

Fig. 3: Decision tree to identify the root cause for the effect *Imprecise Cabin Placement*, based on I40 asset properties in the *Cause-Effect Knowledge Graph (CEKG)* in Figure 2.

without *CEKG* and/or *DT* and strong benefits for the I4ARM approach.

Discussion. In this subsection, we discuss the results according to the research questions (cf. Section III).

RQ1. What knowledge model can represent knowledge on and dependencies between I40 components and risky effects and their causes? Section V-A introduced the *I4ARM Cause-Effect Knowledge Graph (CEKG)* for cause-effect annotation; and Section V-B described the I4ARM method that applies the *CEKG* approach. In the evaluation with domain experts from academia and industry, experts found the *CEKG* useful for supporting risk analysis and risk mitigation in CPPS environments. While the *CEKG* was partly hard to understand for non-expert users, experts found the approach useful for improving and extending their risk analysis process based on the FMEA. However, the initial effort for analyzing and initially constructing the *CEKG* pays off benefits that arise in the operation phase (due to reduced downtime of the production system because of a faster identification of root causes in case of observed negative effects).

RQ2. How can domain experts derive a Decision Tree building on a root cause analysis for efficient risk mitigation guidance? Section V-C introduced the I4ARM decision tree design based on the *CEKG* and derived data of system logs. *DT*s can help non-expert users in efficiently identifying root causes based on observed effects. Therefore, the *DT* approach that is based on data, derived from system logs or vendor data can help to improve risk analysis and risk mitigation in CPPS environments.

Limitations. In this paper we introduced I4ARM for risk analysis and mitigation. However, the following limitations require further investigation: *Evaluation.* The comparative study focused on the use case *Industry 4.0 Production Line* in a state of the art I40 Testbed. This may introduce bias due to the specific selection of production issues challenges and alternative risk mitigation approaches considered, as well as the roles or individual preferences of the domain experts. To overcome these limitations, we plan case studies in a wider variety of application contexts.

Limitations of the I4ARM knowledge graph. In this work, the knowledge graph requires *CEKG* experts that support domain experts in constructing the *CEKG*. We plan tool support to improve the I4ARM method, to support domain experts in constructing I4ARM more efficiently and effectively.

Limitations of the Decision Tree. In this paper, we applied a sample set of test data for manually constructing the *DT* to demonstrate the concept of I4ARM. In future research work we plan to implement the *DT* approach with test data from a real-world use case.

VII. Conclusion and Future Work

The goal of this paper was to improve the efficiency of multi-disciplinary risk mitigation in the engineering and operation of an automated robot-based production system. We introduced the I4ARM method that is based on a *Cause-Effect Knowledge graph (CEKG)* based on an I40 Asset Network in CPPS environments. Traditional cause-effect analysis approaches, such as the Ishikawa approach [22] typically focus on one individual effect as foundation for identifying root causes. In this paper we go beyond the state of the art by modelling a set of effects in a *CEKG* approach within an *I40 Asset Network*. For providing guidelines for root cause analysis, we followed the Decision Tree approach to support experts and non-experts in identifying and mitigating risks in a CPPS. In a traditional approach [18] a huge amount of data is required for deriving the DT. However, in the I4ARM approach, we focus on a subset of data that address risky effects with a minimum set of data.

Based on a conceptual evaluation with domain experts from academia and industry (including expert and non-expert users), expected benefits have been confirmed: (1) *Non-Experts* can use the *DT* approach to efficiently identify root causes based on guidelines, derived from *DT* and *CEKG*. (2) *Advanced Users* can improve over the

guidelines coming from the decision tree by going back to the cause-effect information in the I40 Asset Network to systematically look risk sources and adapt the guidelines. (3) Even with a reasonably moderate amount of data, a meaningful decision tree can be designed due to the information coming from the cause-effect information in the I40 Asset Network. (4) In an organization, there is the need for a dedicated role that takes care on the *CEKG* to improve the decision tree (and the associated guidelines) as more data becomes available from CPPS operation.

Future Work. We plan to investigate the standardized representation of data relevant for *CEKG* modeling, with the use of *AutomationML*[9] data format or ISA-95 [27] information modeling and categorization, in the spirit of [34]. We will investigate the I4ARM approach in a large-scale real industrial environment at a partner in industrial product packing. For improving the construction of *DT*, we will consider integrating vendor guidelines to improve the validity of derived *DT* options.

REFERENCES

[1] Plattform Industrie 4.0 and ZVEI. Part 1 - The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC01 Review). Standard, German BMWI, Nov. 2020. https://bit.ly/37A002I.

[2] Matthew Anyanwu and S. Shiva. Comparative analysis of serial decision tree classification algorithms. *International Journal of Computer Science and Security*, 3(3), 09 2009.

[3] Stefan Biffl, Arndt Lüder, and Detlef Gerhard, editors. *Multi-Disciplinary Engineering for Cyber-Physical Production Systems, Data Models and Software Solutions for Handling Complex Engineering Projects*. Springer, 2017.

[4] Stefan Biffl, Arndt Lüder, Elmar Kiesling, Kristof Meixner, Felix Rinker, Christian Engelbrecht, Matthias Eckhart, and Dietmar Winkler. Multi-Aspect Risk Exploration in Models for Positioning and Joining Simulation (Case Study) Part II. Technical Report CDL-SQI-2020-07, CDL-SQI, Institute for ISE, TU Wien, November 2020. https://url.tuwien.at/ujyge.

[5] Stefan Biffl, Arndt Lüder, Kristof Meixner, Felix Rinker, Matthias Eckhart, and Dietmar Winkler. Multi-View-Model Risk Assessment in Cyber-Physical Production Systems Engineering (in press). In *8th Int. Conf. on Model-Driven Eng. and Softw. Dev., MODELSWARD 2021, online, Feb. 8-10, 2021*. SciTePress, 2021.

[6] Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, 2001.

[7] Anis Cherfi, Kaouther Nouira, and Ahmed Ferchichi. Very fast C4.5 decision tree algorithm. *Appl. Artif. Intell.*, 32(2):119–137, 2018.

[8] Emelie Engström, Margaret-Anne Storey, Per Runeson, Martin Höst, and Maria Teresa Baldassarre. How software engineering research aligns with design science: a review. *Empirical Software Eng.*, 25(4):2630–2660, 2020.

[9] Sebastian Haag and Reiner Anderl. Digital twin–proof of concept. *Manufacturing Letters*, 15:64–66, 2018.

[10] Bernhard Kaiser, Peter Liggesmeyer, and Oliver Mäckel. A new component concept for fault trees. In *Proceedings of the 8th Australian workshop on Safety critical systems and software-Volume 33*, pages 37–46. Citeseer, 2003.

[11] Jay Lee, Edzel Lapira, Behrad Bagheri, and Hung-an Kao. Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing letters*, 1(1):38–41, 2013.

[12] Luca Liliana. A new model of ishikawa diagram for quality assessment. In *IOP Conference Series: Materials Science and Engineering*, volume 161, page 012099. IOP Publishing, 2016.

[13] Hu-Chen Liu, Long Liu, and Nan Liu. Risk evaluation approaches in failure mode and effects analysis: A literature review. *Expert systems with applications*, 40(2):828–838, 2013.

[14] P. Loucopoulos, E. Kavakli, and N. Chechina. Requirements engineering for cyber physical production systems. In *Int. Conf. on Adv. Inf. Sys. Eng.*, pages 276–291. Springer, 2019.

[15] A. Lüder, M. Schleipen, N. Schmidt, J. Pfrommer, and R. Henßen. One step towards an industry 4.0 component. In *2017 13th IEEE Conference on Automation Science and Engineering (CASE)*, pages 1268–1273, 2017.

[16] Oliver Mäckel and Martin Rothfelder. Challenges and Solutions for Fault Tree Analysis Arising from Automatic Fault Tree Generation. In *World Multiconf. on Systemics, Cybernetics and Inf., ISAS-SCIs 2001, USA, Proc., Vol. I: Inf. Sys. Dev.*, pages 583–588. IIIS, 2001.

[17] László Monostori, Botond Kádár, Thomas Bauernhansl, Shinsuke Kondoh, S Kumara, Gunther Reinhart, Olaf Sauer, Gunther Schuh, Wilfried Sihn, and Kenichi Ueda. Cyber-physical systems in manufacturing. *CIRP Annals*, 65(2):621 – 641, 2016.

[18] Arundhati Navada, Aamir Nizam Ansari, Siddharth Patil, and Balwant A Sonkamble. Overview of use of decision tree algorithms in machine learning. In *2011 IEEE control and system graduate research colloquium*, pages 37–42. IEEE, 2011.

[19] P. Novák, Š. Stoszek, and J. Vyskočil. Calibrating industrial robots with absolute position tracking system. In *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, volume 1, pages 1187–1190, 2020.

[20] F. Ocker, M. Seitz, M. Oligschläger, M. Zou, and B. Vogel-Heuser. Increasing Awareness for Potential Technical Debt in the Engineering of Production Systems. In *2019 IEEE 17th Int. Conf. Ind. Inf. (INDIN)*, volume 1, pages 478–484, 2019.

[21] Yiannis Papadopoulos and John A McDermid. Hierarchically performed hazard origin and propagation studies. In *International Conference on Computer Safety, Reliability, and Security*, pages 139–152. Springer, 1999.

[22] Judea Pear and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.

[23] J. Ross Quinlan. Induction of decision trees. *Mach. Learn.*, 1(1):81–106, 1986.

[24] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

[25] J. Ross Quinlan. Improved use of continuous attributes in C4.5. *J. Artif. Intell. Res.*, 4:77–90, 1996.

[26] Miriam Schleipen, Arndt Lüder, Olaf Sauer, Holger Flatt, and Jürgen Jasperneite. Requirements and concept for plug-and-work. *at-Automatisierungstechnik*, 63(10):801–820, 2015.

[27] Bianca Scholten. *The road to integration: A guide to applying the ISA-95 standard in manufacturing*. Isa, 2007.

[28] Praveen Ranjan Srivastava, Parshad Patel, and Siddharth Chatrola. Cause effect graph to decision table generation. *ACM SIGSOFT Software Engineering Notes*, 34(2):1–4, 2009.

[29] Diomidis H Stamatis. *Failure mode and effect analysis: FMEA from theory to execution*. Quality Press, 2003.

[30] Birgit Vogel-Heuser, Thomas Bauernhansl, and Michael Ten Hompel. Handbuch industrie 4.0 bd. 4. *Allgemeine Grundlagen*, 2, 2017.

[31] Birgit Vogel-Heuser, Thomas Bauernhansl, and Michael Ten Hompel. Handbuch Industrie 4.0 Bd. 4. *Allgemeine Grundlagen*, 2, 2020.

[32] B. Wally, J. Vyskočil, P. Novák, C. Huemer, R. Šindelar, P. Kadera, A. Mazak, and M. Wimmer. Flexible production systems: Automated generation of operations plans based on isa-95 and pddl. *IEEE Robotics and Automation Letters*, 4(4):4062–4069, 2019.

[33] B. Wally, J. Vyskočil, P. Novák, C. Huemer, R. Šindelář, P. Kadera, A. Mazak-Huemer, and M. Wimmer. Leveraging iterative plan refinement for reactive smart manufacturing systems. *IEEE Transactions on Automation Science and Engineering*, 18(1):230–243, 2021.

[34] Bernhard Wally, Christian Huemer, Alexandra Mazak, and Manuel Wimmer. Automationml, ISA-95 and others: Rendezvous in the OPC UA universe. In *14th IEEE International Conference on Automation Science and Engineering, CASE 2018, Munich, Germany, August 20-24, 2018*, pages 1381–1387. IEEE, 2018.

[35] Roel J. Wieringa. *Design science methodology for information systems and software engineering*. Springer, 2014.

[36] Xindong Wu, Vipin Kumar, J. Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus F. M. Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael S. Steinbach, David J. Hand, and Dan Steinberg. Top 10 algorithms in data mining. *Knowl. Inf. Syst.*, 14(1):1–37, 2008.

[9] AutomationML: www.automationML.org