



On the application of process management and process mining to Industry 4.0

Flavia Monti¹ · Jerin George Mathew¹ · Francesco Leotta¹ · Agnes Koschmider² · Massimo Mecella¹

Received: 16 January 2023 / Revised: 15 March 2024 / Accepted: 26 March 2024
© The Author(s) 2024

Abstract

The continuous evolution of digital technologies applied to the more traditional world of industrial automation led to Industry 4.0, which envisions production processes subject to continuous monitoring and able to dynamically respond to changes that can affect the production at any stage (resilient factory). The concept of *agility*, which is a core element of Industry 4.0, is defined as the ability to quickly react to breaks and quickly adapt to changes. Accurate approaches should be implemented aiming at managing, optimizing and improving production processes. In this vision paper, we show how process management (BPM) can benefit from the availability of raw data from the industrial internet of things to obtain agile processes by using a top-down approach based on automated synthesis and a bottom-up approach based on mining.

Keywords Industry 4.0 · Process management · BPM · Process mining · Process adaptation

1 Introduction

Industry 4.0 envisions *production processes* subject to continuous monitoring and able to *dynamically respond to changes* that can affect the production at any stage. In particular, Industry 4.0 processes must be *agile*, where agility is defined as a combination of responsiveness and resilience [1]. Responsiveness concerns the ability to adapt to changes in the demand, provide customers with personalized products (mass customization), quickly exploit temporary or permanent advantages and keep their competitive edge, while resilience concerns the ability to react to disruptions along

the supply chain. The resulting processes will be able to successfully adapt to an evolving and uncertain business context in terms of both demand (customization, variability, unpredictability) and supply (new components, uncertainty in the supplies, bottlenecks and risks) taking into account not only the single organization but the entire value chain.

The term *smart manufacturing* is used in some contexts (especially in the US literature and market) as synonymous of Industry 4.0, even though in Industry 4.0 the focus is, in principle, not only on the manufacturing phase, having instead impact on all the phases of the product life-cycle [2, 3]. Anyway, this term highlights the role of smart systems as a fundamental component of the last industrial revolution. A smart system, also known as an intelligent system, refers to a (set of) technology(ies) and/or device(s) that incorporate(s) advanced computing, sensing, and communication capabilities to perform tasks or make decisions autonomously or with minimal human intervention [4]. These systems use data, often collected in real time, to analyze and adapt their behavior in response to changing conditions, ultimately optimizing their performance and enhancing efficiency.

In a *digital factory*, the involved actors falling in different categories, being humans (i.e., final users or participants in the production process), information systems/applications or industrial machines, must be able to communicate and interact at the digital level while operating in the physical world. The industrial internet-of-things (IIoT) represents one

Communicated by Selmin Nurcan and Rainer Schmidt.

✉ Flavia Monti
monti@diag.uniroma1.it

Jerin George Mathew
mathew@diag.uniroma1.it

Francesco Leotta
leotta@diag.uniroma1.it

Agnes Koschmider
agnes.koschmider@uni-bayreuth.de

Massimo Mecella
mecella@diag.uniroma1.it

¹ Sapienza Università di Roma, Rome, Italy

² University of Bayreuth, Bayreuth, Germany

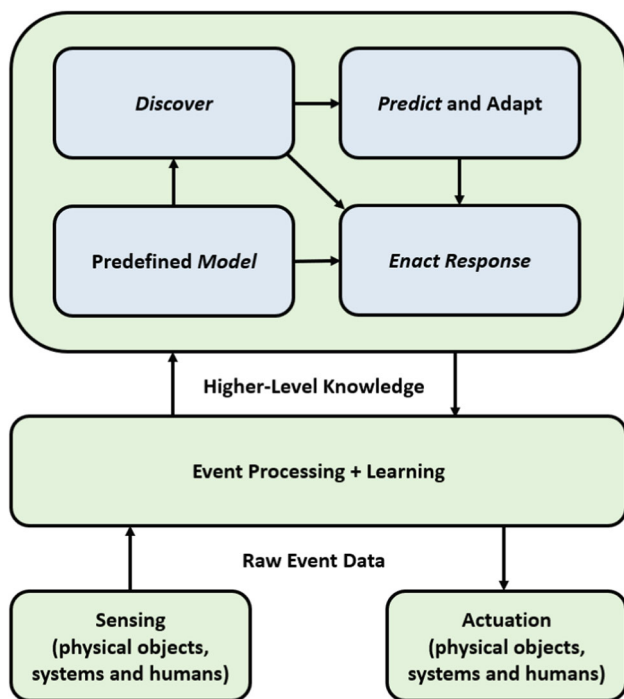


Fig. 1 Interplay of IoT and BPM [8]

of the technological pillars to this end [5]. By covering the domains of machine-to-machine (M2M) and industrial communication technologies, IIoT is the computing concept that enables efficient interaction between the physical world and its digital counterpart [6]. In particular, thanks to IIoT, physical entities involved in the manufacturing process can have faithful representations in the digital world, usually defined as *digital twins* [7].

The employment of raw data coming from IIoT to obtain industrial agile processes is not straightforward. The *IoT-Meets-BPM Manifesto* [8] proposes how business process management (BPM) and the IoT can benefit from each other and be effectively combined (see Fig. 1). On the one hand, the IoT provides a vast amount of data (e.g., from sensors) that can complement already available data (e.g., traces produced by processes) used in BPM. A clear advantage of using IoT data is that it can be automatically retrieved from devices, in contrast to traditional procedures to collect them, which may require having human users manually starting and ending activities. In addition to that, IoT data are less error-prone and thus more accurate than human-generated data. On the other hand, IoT data from a single device (e.g., a converter) may represent simply a routine, i.e., they might lack a broader view of the higher-level context in which they are actually used (e.g., a manufacturing process). In such a scenario, BPM can be leveraged to bridge the gap between raw sensor data and the actual activities in the underlying process.

In this vision paper, in the context of the problems identified by the IoT-Meets-BPM Manifesto, we focus on how to achieve process agility in smart manufacturing. In particular, we propose (i) a top-down approach based on automated synthesis, and (ii) a bottom-up approach based on process mining. To illustrate how the two approaches can be applied, a case study is introduced, abstracted by analogy over the real experiences of the authors in industrial settings.

This paper is based on two works in progress [9, 10], presented during the 23rd business process modeling, development and support (BPMDs) working conference without publication. In particular, authors in [9] discuss the importance of the employment of artificial intelligence for the resilience of business processes in Industry 4.0, whereas authors in [10] discuss the potential of applying process mining to time series produced by IoT powered environments. As mentioned above, these two visions suggested that Industry 4.0 can benefit from a combined top-down and bottom-up approach to BPM, which is the goal of the present work. To define a common framework, the authors of both papers agreed on a common case study (see Sect. 4) on which the proposed approach has been defined and exemplified.

The paper is organized as follows. Section 2 introduces preliminary concepts relevant for the remaining of the manuscript. Section 3 introduces and analyzes related works. Section 4 introduces the case study we refer to in the rest of the paper. Sections 5 and 6 present how to obtain agile industrial processes by using synthesis and process mining, respectively. Finally, a discussion and concluding remarks are presented in Sect. 7.

2 Preliminaries

2.1 BPM and process mining

Business process management (BPM) is a discipline that uses various methods to discover, model, analyze, measure, improve and optimize business processes. A business process coordinates the behavior of people, systems, information and things to produce business outcomes in support of a business strategy [11]. In simpler terms, BPM is a way of managing and improving the processes that are used to create and deliver products and services. It involves identifying the steps of a process, analyzing them to identify areas for improvement, and then making changes to streamline the process and make it more efficient.

Process mining, by merging data mining techniques and processes, enables decision makers to discover process models from data, compare expected and actual behaviors, and enrich models with information retrieved from data [12]. It focuses on the real execution of processes, as reflected by the

event logs collected from the software systems of an organization.

The main techniques of process mining are discovery and conformance checking. Discovery starts from an event log and automatically produces a process model that explains the different behaviors observed in the log, without assuming any prior knowledge on the process. Nowadays, whereas a plethora of process discovery solutions has been developed and successfully employed in several application domains [13], existing techniques are suitable to discover processes that have no data perspective incorporated into them. Conformance checking compares a process model and an event log for the same process, with the aim of understanding the presence and nature of deviations. The majority of conformance checking techniques are based on ad hoc implementations of traditional searching algorithms that work in specific domains [14].

With respect to high level support to processes, process mining can be used in a bottom-up fashion to (i) discover the process actually employed and thus fully exploiting raw data, (ii) identify inefficiencies and problems in the process (e.g., the negative outcome of the process can be predicted by taking into account sensor measurements ignored by traditional information systems), (iii) integrate high level automation procedures with additional conditions mined from raw data, and (iv) suggest a different/better usage of resources.

2.2 AI and the control problem

A control problem is typically a situation in which a system needs to be managed to achieve specific objectives. The main goal in such a problem is the selection of the right actions to be performed in order to maintain the system in an acceptable state. Solutions to such a problem are commonly based on classical feedback control theory. In this context, a well-known approach to support such systems is the feedback control loop called MAPE-K [15] conceived as a sequence of four actions, i.e., monitor, analyze, plan and execute, over a knowledge base.

AI technologies can play an important role in the development of appropriate control systems able to manage complex processes while remaining robust, reactive and adaptive in the presence of both environmental and task changes [16]. Three different approaches can be distinguished: (i) a programming-based approach, where the control is manually generated (commonly by a programmer) and un-anticipated events cannot be handled; (ii) a learning-based approach (e.g., based on reinforcement learning), which does not require a complete knowledge by the experts but instead necessitate the presence of data related to negative conditions (which are hard to collect); (iii) and a model-based approach, where the problem is specified by hand (using, for instance, formal methods) and the control is automatically

derived (e.g., automated planning, synthesis). We propose a model-based approach based on automated planning.

Automated planning, which is one of the oldest areas in AI, is designed to synthesize autonomous behaviors in an automated way from a model [17]. Different types of planning can be distinguished, e.g., classical planning and non-deterministic planning. Intuitively, classical planning deals with deterministic scenarios, while non-deterministic planning deals with non-deterministic and stochastic scenarios.

Formally, a *classical planning problem* [18] (X, I, γ, O) consists of a set of state variables X , a description of the initial state I of the system (i.e., a valuation over X), a goal γ represented as a formula over X , and a list of operations (or actions) O over X that can lead to state transitions. State variables X and actions O constitute the *planning domain* of the problem. Solving a planning problem aims at automatically finding a sequence of actions that, applied to the initial state I , leads to a state s such that $s \models \gamma$ (s satisfies γ). An action $o \in O$ is defined as a tuple $o = \langle \chi, e \rangle$ where χ is the *precondition* and e is the *effect*. Both precondition and effect are conjunctions of literals (positive or negative atomic sentences) over X . χ defines the states in which o can be executed, i.e., o is applicable in a state s if and only if $s \models \chi$; and e defines the result of executing o . For each effect e and state s , the change set $[e_s]$ is a set of state variables whose value is modified upon the action. And, the *successor state* of s with respect to the action o , is the state s' such that $s' \models [e_s]$ and $s'(v) = s(v)$ for all state variables v not mentioned in $[e_s]$.

The state space of the problem may be huge; nevertheless, many algorithms and heuristics have been proposed and integrated over the years into planning systems (i.e., planners) to find a solution efficiently. Planners input the problem model using a standard language PDDL (Planning Domain Definition Language) [19]. The problem model is described in the form of the domain and problem descriptions. Such a division allows for an intuitive separation of the elements related to every specific problem of the given domain (i.e., types, predicates and actions), and the elements that determine the specific problem (i.e., available objects, initial state and goal).

2.3 Information systems in Industry 4.0

Processes of an Industry 4.0 company are usually managed by a set of application systems (also referred to as information systems—ISs), along the set of steps of the product manufacturing, which ranges from the development of the idea to the post-manufacture monitoring. Additionally, to support the business processes at an operational level, business process management systems (BPMSs) can be employed [20].

At design time, computer-aided design (CAD) and computer-aided manufacturing (CAM) software are used,

respectively, to design a product and to determine the machine operations [21]. A product knowledge environment is developed by the product life-cycle management (PLM) that monitors and analyzes the product during its entire life [22]. To support the customers and the suppliers, the customer relationship management (CRM) and the supply chain management (SCM) are employed [23]. Further, the warehouse management system (WMS) controls and administers warehouse operations [24].

Often confused with each other, the enterprise resource planning (ERP) and the manufacturing execution system (MES) represent critical ISs providing fundamental operational support. On the one hand, the ERP supports most of the processes common to any kind of company including sales, marketing, purchasing, production planning, inventory, finance, and human resources [25], thus incorporating CRM, SCM and WMS. On the other hand, a MES is commonly associated with the ERP system and the production lines, storing data from both sources to monitor and optimize the production process. Common MES functionalities include production management, creating the most productive schedule of detailed operations needed to accomplish the master production plan [26]. Also, the MES is interconnected to the other components in the company to support interoperability between them and enable autonomous decisions [27].

The employment and management of processes in factories can range from manually defined procedures (sometimes called recipes) to the employment of resource allocation and scheduling techniques to more advanced approaches based on AI, as proposed in this work. As discussed in Sects. 2.2 and 5, by leveraging AI techniques and embedding them in the ERP and MES, model-based approaches can be deployed to ensure a production plan (i.e., a solution to the control problem of the manufacturing company) [28]. Models are built over the data exchanged between the ISs to suggest the ideal planning. This in turn can be leveraged to define resilient master production plans (in the ERP) and detailed operation schedules (in the MES).

3 Related works

Only very recently, researchers have emphasized the advantages of incorporating BPM into Industry 4.0 companies. For instance, authors in [29] have highlighted that a high level of BPM implementation also impacts a high level of adoption of Industry 4.0 key-enabling technologies. In this context, innovative technologies such as IIoT, enable the collection and analysis of real-time data, empowering businesses to optimize and automate their processes [30].

Authors in [31] also explored the intersection of Industry 4.0 and BPM. This intersection not only opens up new directions for the management and execution of business processes

but also presents new conceptual, technological and methodological challenges for information systems. These systems must become more sensitive to event processing and have to consume a large volume of data permanently. In this regard, authors in [32] demonstrated that BPM enables the development of frameworks for integrating IoT with BPM, enabling bidirectional communication and IoT-aware process execution.

Industry 4.0 has led to the need for new business processes representations that take into account the characteristics of IoT technologies or Cloud Computing applications [8]. Some studies have proposed novel approaches to support IoT-aware processes. Recent surveys present comprehensive lists and comparisons of proposed solutions [33–36], even if they are generic to IoT-aware business process modeling, without any specificity for Industry 4.0 scenarios. Interestingly, whether smart manufacturing poses new challenges in terms of IoT-aware processes has not yet been investigated, to the best of our knowledge.

Again only very recently, researchers have started presenting innovative approaches leveraging process mining as an integral part of the management of dynamic manufacturing processes. Authors in [37] proposed an innovative approach based on joint mining techniques to support the extraction of reliable process models from event data generated in manufacturing systems. Authors in [38] investigated the use of process mining for life cycle assessment in manufacturing aiming at identifying process deviations and interruptions. A comprehensive set of recommendations on how to employ which process mining technique for different goal analyses is discussed in [39]. Presented guidelines can be summarized in: (i) use process discovery as a screening tool and (ii) use conformance checking to detect deviations. Particularly, (iii) use data-aware conformance checking for quality issues, concept drift prediction and predictive maintenance, and (iv) time-aware conformance checking for temporal deviations and organizational problems. Finally, (v) collaborate with domain experts to define process enhancement actions.

A few seminal works proposing AI planning in manufacturing are [40–42], in which the focus is on robotics and single machines, whereas our work attempt to provide a high-level view, by applying planning at the level of business processes considered as an orchestration of intermediate steps (manufacturing services) for achieving a production goal [43, 44].

4 The glass factory case study

We introduce a case study scenario that depicts a glass manufacturing company named *GlassFactory*.¹ A brief description

¹ Adapted from <https://in.saint-gobain-glass.com/glass-manufacturing-process>

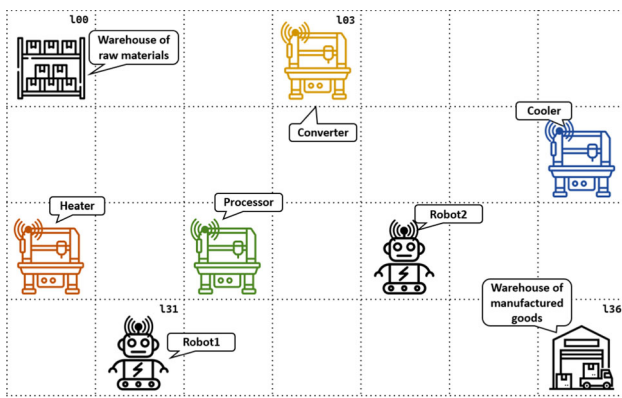


Fig. 2 *GlassFactory* manufacturing actors and layout

of the actors involved is provided below, and their placement in the factory is shown in Fig. 2.

- *Raw materials* are a mixture of different components, comprising silica sand, calcium oxide, soda, magnesium and recycled glass (cullet). They are stored in the warehouse as batches and are passive, meaning that other actors are responsible for them. In the rest of the paper, they are named objects.
- A *warehouse* (of raw materials) stores the batches of raw materials before being used in the manufacturing process.
- A *heater* machinery heats the batched raw materials (one at a time) transforming them into molten glass.
- A *processor* machinery gives a shape to the molten glass (one at a time).
- A *cooler* machinery cools, one at a time, the glass.
- A *converter* machinery performs the tasks of the previous three actors, i.e., heat, process and cool objects, in one pass.
- Two *robotic carriers* (or *robots*) have three capabilities, i.e., loading and unloading an object, and moving around the factory.
- A *warehouse* (of manufactured goods) stores produced glass before their distribution for sale.

Also considered as actors of the *GlassFactory*, the ERP and the MES systems manage the business processes. Specifically, the ERP deals with warehouse-related processes (e.g., inventory, purchasing and planning purposes) and the MES deals with production processes aiming at computing accurate operation planning (coherent with the ERP) and monitoring the production.

The glass production process consists of the ordered sequence of heating, processing and cooling transformations on raw material. Those activities are enabled by the heater, the processor and the cooler, respectively. As an alternative, the converter can be used to perform those same activities all in one. The objects (i.e., raw materials and (semi-)finished

glass) are carried by the robots among the factory machines to undergo the mentioned transformations. The principal production goal of *GlassFactory* is to “manufacture a single glass product and place it in the warehouse of manufactured goods”.

The involved machinery and robots are equipped with sensors and computing capabilities. Usually, the sensors collect real-time working data such as speed, temperature, vibration and voltage. As described in [45], such data can be accessed through a set of application programming interfaces (APIs) which can be implemented as synchronous or asynchronous endpoints following different protocols (e.g., MQTT, Rest APIs, WebSockets). Usually, the accessed sensor data are usually stored in databases in a structured or even unstructured manner.

Commonly, available endpoints also give the chance to retrieve the descriptions of the actors themselves. Such descriptions include functionalities specified in terms of the actions the specific actor is able to perform and status information. Particularly, an action is expressed in terms of prerequisites, post-action conditions and extra details such as execution cost and probability of success. The prerequisites, detail the conditions (typically, Boolean conditions) enabling the execution of an action. The post-action conditions indicate the effects of the execution of an action and they are expressed in terms of which properties change.

Though the just described case study is a fictional one, it is designed to closely mimic real-world Industry 4.0 scenarios. Several real-world examples are considered to devise the *GlassFactory* case study. The real cases range from industrial ones such as die cutters, packaging [46] and spindles [47] manufacturing production to the agri-food sector like wine production [48].

The *GlassFactory* case study captures the essence of real-world manufacturing networks (information systems, machines, products, humans) by simply abstracting their complex interactions to achieve a common goal, i.e., the manufacture of a product. This simplified digital factory emphasizes the pervasive presence of sensors and computing capabilities within each actor. This enables the collection of tons of data for insightful analyses.

In addition, the proposed case study acts as a bridge, transforming complex realities into easily accessible solutions for experimentation. This fictional setting allows us to explore hypothetical scenarios and control for specific variables that may be difficult to isolate in the real world. Although we acknowledge that generalization may be challenging, we strongly believe that the core findings can be applied in similar real-world contexts.

On the other hand, as the case study, though realistic, is fictitious, some limitations may apply. In first place, the factory is modeled in terms of discretized space and time. Even though this is usually considered a fair assumption,

as abstraction is part of the application methodology of both process mining and automated synthesis, the lack of real data may result in an excessively simplified scenario. However, this does not undermine the drawn conclusions as proposed techniques can be applied in more complex situations as well.

5 Synthesis-based agility

In the traditional manufacturing process, the production plan is generated based on new and historical orders. Then the preparation for production is carried out, such as equipment maintenance and material collection. Finally, formal production is executed according to the plan. If some conflicts appear, the plan is modified to adapt to the actual situation. After production, finished products are inspected to ensure whether they meet requirements, and consequently transported into the warehouse or repaired. Information generated during production such as process documents and fault records are kept in files for the next round.

Automated planning techniques can be employed in a context like that to automatically orchestrate the supply chain to satisfy a specific goal. The generated output consists of a synthesis of autonomous behaviors (i.e., plans) from a model that compactly and mathematically describes the problem environment (i.e., the manufacturing company). Its usage enables an automatic recovery, optimization and orchestration of the necessary actions to be executed to achieve the goal.

We propose a resilient automated planning-based approach for the agile orchestration of the involved actors [43]. Particularly, we take advantage of the opportunity to extract the characteristics of the actors (e.g., warehouses and robots) from the available endpoints to define in an automated way the description of the problem environment. The main components of the proposed approach are depicted in Fig. 3.

We devise a module, i.e., *translator*, that combines the actors and the production goal descriptions and converts them into a PDDL problem model composed of the domain and problem files. The *planner* produces the plan based on the problem model and inputs it to the *enactor* module. The latter manages the effective execution of all the necessary actions

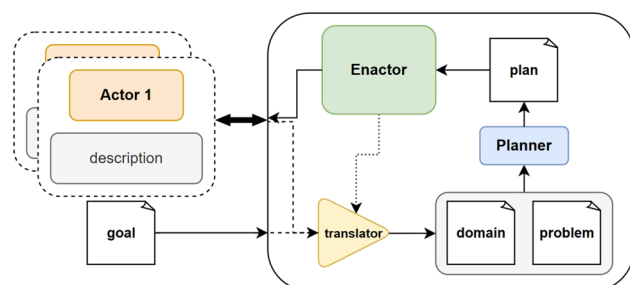


Fig. 3 The architecture of the proposed prototype

```

1 (move rb1 l31 l00)
2 (load rb1 o1 l00)
3 (move rb1 l00 l03)
4 (unload rb1 o1 l03)
5 (convert cnv1 o1 l03)
6 (load rb1 o1 l03)
7 (move rb1 l03 l36)
8 (unload rb1 o1 l36)

```

Fig. 4 Main plan in the *GlassFactory* use case. *rb1* represents *robot 1*, *o1* represents the manufactured object and $l<r><c>$ represents locations with $<r>$ and $<c>$ referring to rows and columns, respectively (see Fig. 2)

specified in the plan by invoking the endpoints of relative actors capable of performing those actions.

5.1 Application to the case study

The production goal of the *GlassFactory* case study corresponds to the manufacture of glass, consisting of the conversion of raw materials into glass successive to the heating, processing and cooling operations. The main steps of the workflow to generate the *GlassFactory* plan for such a production goal are described below and follow the depiction in Fig. 3.

The actors' descriptions, i.e., warehouses, robots, machines and objects, are retrieved and fed, together with the goal description, to the translator which generates the PDDL problem model, i.e., the domain and problem files.

The problem model is then processed by the planner and, upon planning completion, the resulting plan is generated. Figure 4 shows one of the possible plans generated for *GlassFactory*. Note that in this case the *converter* will be used, as it is the cheapest way to achieve the production goal. We can expect that in the case the converter is broken, the other three machines will be used instead.

The sequence of actions listed in the plan is managed by the enactor, that one at a time calls the specific actor able to execute that action. Notably, between the action executions, the enactor checks the state of the actor who executed the previous action to verify if the output is equal to the expected one. Let us assume, for instance, that the fifth operation of the main plan times out. Then the enactor will (i) set the status of the converter to broken and (ii) activate the translator to update the problem model. The planner will then generate a new updated plan (recovery plan) that, starting from the current state of the system can lead to the final goal (involving for instance the heater, processor and cooler).

A tool of the proposed prototype has been developed² and experiments have been conducted by synthesizing different plans based on diverse conditions of the available actors.

² Source code and experiment results of the developed tool available at https://github.com/iaiamomo/agility_synthesis

5.2 Support to agility

The proposed approach enables an agile and responsive solution to the control problem. This option is enabled mainly because of the presence of two main components, i.e., the enactor and actor endpoints.

The presence of endpoints allows the possibility to retrieve information related to the status of the actors involved in the process at any moment. In other words, upon an unexpected situation happens, it is immediately noticed thanks to the presence of the endpoints.

Additionally, the presence of the enactor module, which interfaces with the actors, allows for continuous monitoring of them during the process execution and if something is not coherent with the expected results it triggers the translator to re-compute a new problem model to reach the final target, i.e., the manufacturing goal.

With the integration of classical planning methods and real-time data gathering, smart manufacturing environments are prepped to achieve new levels of efficiency and reactivity ensuring an agile optimization of the production processes.

5.3 Limitations

The proposed approach represents a possible solution to the general problem of orchestration of the supply chain. Despite this, employing the classical planning approach suffers limitations as regards the neglecting of the non-deterministic and probabilistic behavior of the manufacturing scenarios, therefore producing non-optimal results. For instance, machines are subject to wear and can suddenly break down, thus leading to disruption of the supply chain and re-design of the production process.

On the other hand, non-deterministic planning approaches, like the ones proposed by authors in [28, 49], although producing exact solutions, have limitations on performances. Such methods apply algorithms like Markov decision processes (MDPs) which are exponential in the number of states and do not compete with the heuristic algorithms used in classical planning methods. Manufacturing scenarios present a vast number of actors (i.e., machinery, humans and actors) and sometimes tricky production process goals (e.g., supply chain), the number of the states could be massive and non-deterministic approaches result in being inoperable, as of today.

6 Mining-based agility

As stated in the *IoT-meets-BPM Manifesto*, the combination of process mining and data analytics techniques like data mining and machine learning, can give valuable insights into the data [8]. Such a combination allows to identify patterns

within low-level data and to apply predictive analytics to reveal hidden knowledge. Then, based on the analysis results, informed decisions can be made.

Industry 4.0 environment contains a huge number of (intelligent) manufacturing actors typically equipped with sensors that enable the availability of a massive amount of raw data. It is not always straightforward to build a relationship between raw data and business processes. However, by leveraging both process mining and data analytics techniques, the gap between the parts can be filled. We can identify different approaches intending to extract unseen insights leveraging on analytics techniques. For instance, localization systems (e.g., geofencing) output position data and methods like pattern recognition or machine learning could be applied to identify events. Furthermore, hidden Markov models and the use of particle filters make it possible to increase the accuracy by integrating motion models [50]. Moreover, AI methods such as clustering and unsupervised learning show structures that are used by techniques such as event classification to aggregate new patterns [32, 51]. Such approaches lay the foundation for collecting event-related data. However, little effort has been spent on efficient data pre-processing and accurate abstraction of event log data from the raw data. A too fine-grained abstraction leads to the overfitting of the discovered process model, while a too coarse-grained abstraction results in underfitting. To illustrate this, consider the activity of preparing the production related to our use case. Then, the question arises if the start of the activity is initiated when the new order has been checked, when the machine settings have been made, or when the machine starts running. Most of the existing process mining techniques are commonly applied to business events and use data on a higher abstraction. Thus, questions related to the analysis purpose and the start activity need to be answered before starting the analysis. The goal of the analysis significantly impacts the granularity of the abstraction. The *IoT-Meets-BPM Manifesto* [8] discusses several challenges including the one concerning abstracted events which need to be aggregated to an event log in order to apply process mining [52]. The challenge is, on the one hand, the combination of offline and online data and, on the other hand, the integration of current information (e.g., context of the environment) in the analysis. Related to abstraction is the contextualization of activities. For instance, in the smart factory example, wearable sensors of the actors produce similar data for lifting a workpiece from one workstation and putting it down at the next station. Because the pre-processing steps miss contextual information, it are unable to efficiently distinguish between these activities during the abstraction/aggregation step. Thus, the data need to be explicitly contextualized into the realm of the analyzed process.

Additionally, the discovered process should be explained in order to increase the trustworthiness of the approach

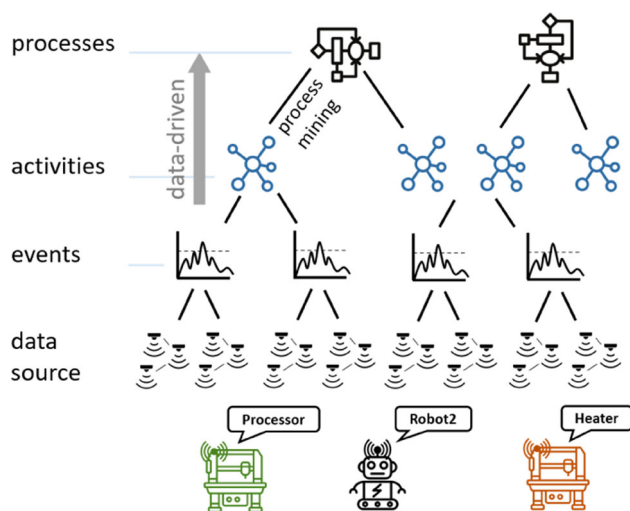


Fig. 5 Visualization of the process analytics pipeline

[53]. The analysis result can be explained in different ways. For example, the combination of process data with domain knowledge (e.g., via labeling or human-in-the-loop approaches [54]) allows transparency and interpretability of the model. Explainability through the annotation of outliers is likewise possible. With this approach, outlier information is propagated through each step of the process from raw data to the analysis results in terms of meta-data annotations [55].

To sum up, the data-driven process mining approach gives valuable insights into the smart factory use case in terms of discovering reasons for, e.g., late packaging and deliveries, identifying bottlenecks, or improving predictions.

6.1 Application to the case study

We now propose a process analytics pipeline that extends the approach proposed in Sect. 5. The pipeline (see Fig. 5) analyzes raw data coming from the manufacturing actors to discover processes and improve the main plan.

The pipeline allows to analyze data and knowledge from different data sources and different levels of abstraction, including scientific models and user (expert) knowledge, through an interactive exploration. First, the raw data must be pre-processed in terms of data cleaning (i.e., removing outliers and extracting representative data). Then, abstraction techniques must be applied to the pre-processed events to enhance the data with the corresponding information. Finally, process mining techniques are used to discover processes by grouping or ordering the aggregated data.

Different types of glass are manufactured by *GlassFactory*, e.g., annealed or tempered, by using different production processes.

The *GlassFactory* main plan found by the orchestrator focuses on the production of one type of glass and consists of the employment of *robot 1* and *converter* machinery. Such

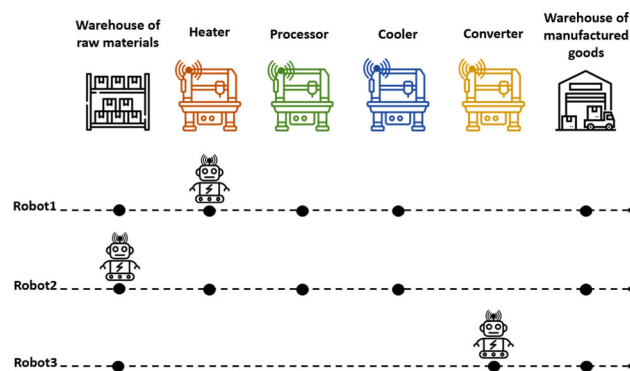


Fig. 6 Enhanced plan with the first two robots employed in the production line composed by the *heater*, *processor* and *cooler*, and the third robot employed in the production line employing the *converter*

a process is executed sequentially until the amount of manufactured raw goods satisfies the main production goal.

Raw sensors data of temperature, voltage and uptime work, allow the extraction of additional information related to machinery/robots: (a) machinery *temperature* reach different values depending on the glass type to be manufactured; (b) machinery status, i.e., waiting, busy or broken, can be accurately inferred considering its *voltage*, and (c) by looking at *uptime work*, production process duration can be upgraded.

The main plan is enhanced in the following way. On the one hand, analytics techniques are applied for predictive maintenance. Currently, if a machinery breaks, the production would be interrupted to repair the machine and the factory would have a great waste of resources (i.e., money and time). Being able to predict machinery (e.g., *converter*) wear out, allows to re-plan before the machinery breaks preventing loss of resources. If *GlassFactory* predicts that the *converter* will break soon, its status would be changed into *broken* and the orchestrator would re-plan the main plan involving the *heater*, *processor* and *cooler*. On the other hand, process mining techniques permit to optimize the plan by considering the uptime work data. The results of such analysis reveal that the production is optimized employing both *robot 1* and *robot 2* which pipeline the operations of the *heater*, *processor* and *cooler* machinery. Finally, it suggests the employment of a third robot so that upon the *converter* is repaired, *GlassFactory* could maximize the production by realizing two production lines, one involving the *converter* which is used sequentially and another one involving the other three machineries used in pipeline. Figure 6 depicts the enhanced production plan.

6.2 Support to agility

For data-driven mining, agility represents a closed-loop approach. This means that the discovered process model is used as essential quality feedback for pre-processing and activity abstractions (see Fig. 5). For this purpose, param-

eter tuning can be used to enhance the process model quality based on, e.g., actor-defined values. The actor in the smart factor pre-selects parameters, and then, the process analytics pipeline is initiated again. It is recommended not to use predefined parameters but rather customize them according to the use case. Mostly, the F1 score is used to evaluate the quality of a process model. To evaluate the hyperparameter tuning, it is recommended to analyze the influence on the F1-score in terms of which parameter values yield average results consistently for the data set.

6.3 Limitations

Even if the produced results are fundamental for the development of a resilient production, it presents some limitations.

The raw sensor data usually have quality problems (in the sense of incompleteness, redundancy or errors). Therefore, the most time-consuming step in the analytics pipeline from processing raw data to discovering knowledge is the processing of the data set [56]. The reason for the time-consuming nature of this activity is usually the quality of the data (i.e., missing or incomplete entries). Some approaches to improving data quality can be found in the literature [57, 58]. However, data preparation is still too time-consuming and appropriate procedures shall be provided.

Also, it is imperative to assess the suitability of the dataset for the intended analysis (i.e., is the data set representative?). An exploratory data analysis might help to get initial insights into whether the gathered data aligns with the analysis goals and is not a “garbage-in-garbage-out” approach [59]. For instance, assume that the wearable of the actors in the smart factory use case are used, but have not been turned off while wearing and thus continuously generate data. This means that certain data cannot be used for analysis or does not exist at all. If we assume that location detection is an analysis purpose in the use case. Then, external data about power failures could be included beside the data from the movement detection sensors making the data set more representative [60].

Commonly, there are often not enough data for the analyses and one way to fix this is to generate synthetic data. It has been shown that synthetic data can not only provide a substitute for real data [61, 62] but even provide new insights into domain-specific research. Synthetic data should be generated that enable different types of analysis. Basically, it must be balanced between the new recording of data mirroring or to evaluate the ML method on synthetic data.

Furthermore, the detection of causal effects from data is highly requested. The challenge, however, is that unstructured data usually do not follow well-defined patterns, so the order in which activities are performed varies depending on the context. This hampers the extraction of causality from data. One solution might be the extraction of probabilistic structures from data.

7 Discussion and concluding remarks

In this paper, we have analyzed the possible applications of BPM and process mining in Industry 4.0. In particular, starting from the general considerations made in the *IoT-Meets-BPM Manifesto* [8], we have proposed an industrial use case showing how process adaptation and process mining can be applied.

Despite the clear potential of the proposed techniques though, real application scenarios are still rare, as even though Industry 4.0 technologies are increasingly frequent, full interconnection of software and hardware systems is very infrequent. Our aim, with the present paper, is to further motivate practitioners and researchers in pursuing such approaches; to this aim, experiments about process adaptivity have been performed, and future work includes the creation of a simulation environment for applying what has been described in this paper.

Limitations of proposed approaches have already been defined in the specific sections above. In this section though, we want to contextualize our work with respect to the more general Industry 4.0 frameworks.

As stated in [1], the role of process management and mining is usually associated with the properly named *Business* and *Functional* logical layers of the Reference Architectural Model for Industry 4.0 (RAMI) [63]. Whereas the business layer specifically focuses on the definition and execution of the business processes, the functional layer implements the *tasks* the business processes refer to. This top-down approach implies that technological changes at the lower hierarchy levels (e.g., *Field Devices*), which are frequent, especially in an Industry 4.0 scenario, are not taken into account and their potential is not exploited.

IoT data can be instead exploited in order to integrate the classical top-down approach (see Sect. 5) and to develop a bottom-up approach leveraging raw data (see Sect. 6). The variety and quantity of IoT data may increase over time, thus producing different perspectives on the processes.

Let us discuss in detail how process mining can be applied to the other logical layers. The *Asset* level represents the source of raw data and, as such, it is the main constraining factor of what can be obtained at the above layers.

Data provided by assets can be analyzed by applying different techniques of data analytics. Data analytics is a fundamental component of the methodology outlined in the *IoT-Meets-BPM Manifesto*. Usually, descriptive, predictive and prescriptive analytics techniques, paired with machine learning and artificial intelligence, are employed to construct models for decision making purposes [64]. Typical goals include finding patterns and correlations of data, e.g., identifying anomalies, forecasting future events, e.g., estimating the remaining useful life (RUL) of machines, and generating recommendations. Such analytics approaches are usually

paired with the concept of digital twin (DT), i.e., a digital representation of a physical asset. DTs are not merely used for simulation purposes in order to continuously improve product design, but they become a way to interact with an asset at any point of its lifecycle [45]. DTs include features such as the ability to interact with the environment (e.g., with other DTs and humans), the ability to collect information about the relative physical entity, the ability to live independently and the ability to modify its behavior using exposed services. In this sense, the term *servitization* is often used.

Advanced data-driven mining approaches require a significant focus on data fusion, i.e., the combination of heterogeneous data sources. This also involves integrating data from distributed sources to create a more complete and accurate picture of the underlying processes. For instance, in the smart factory use case, data from sensors and video cameras are collected to monitor the tasks of the production. By fusing this data, and other structured data sources such as production logs and quality control reports, it is possible to identify patterns that may not be apparent from solely one data source alone [65].

The central role of products is one of the main characteristics of Industry 4.0. Products are followed during their entire life cycle, from the development phase to manufacturing, employment and in some cases to re/de-manufacturing.

Products have always been indirectly part of business processes employed in manufacturing, as the final goal of those processes is the creation of products. In the Industry 4.0 and RAMI vision though, products become active actors that can execute tasks and send notifications. As such, in the future, they will be more and more part of specific processes, in particular with respect to maintenance. In this paper, we mainly discussed adaptivity as seen from the point of view of the manufacturing process. Future works will include the extension of such an approach to other phases of the life-cycle of the product.

A final remark must be made concerning the two implemented approaches. The deployment of such technologies in a real application scenario requires considering integration issues between different involved information systems that have been not considered in our simulations as considered more practical than theoretical.

Acknowledgements The work of Francesco Leotta and Massimo Mecella is partially funded by MICS (Made in Italy—Circular and Sustainable) Extended Partnership (CUP B53C22004130001) and received funding from the European Union Next GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR)—Missione 4 Componente 2, Investimento 1.3—D.D. 1551.1110-2022, PE00000004). Flavia Monti and Jerin George Mathew were partly supported by the Electros spindle 4.0 project (funded by Ministero per lo Sviluppo Economico, Italy, no. F/160038/01-04/X41). The work of Flavia Monti is also supported by the MISE agreement on “Agile&Secure Digital Twins (A&S-DT)”. Jerin George Mathew is financed by the Italian National PhD Program in AI.

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bicocchi, N., Cabri, G., Mandreoli, F., Mecella, M.: Dynamic digital factories for agile supply chains: an architectural approach. *J. Ind. Inf. Integr.* **15**, 111–121 (2019)
2. Thoben, K.D., Wiesner, S., Wuest, T.: “Industrie 4.0” and smart manufacturing—a review of research issues and application examples. *Int. J. Autom. Technol.* **11**(1), 4–16 (2017)
3. Culot, G., Nassimbeni, G., Orzes, G., Sartor, M.: Behind the definition of Industry 4.0: analysis and open questions. *Int. J. Prod. Econ.* **226**, 107617 (2020)
4. Romero, M., Guédria, W., Panetto, H., Barafort, B.: Towards a characterisation of smart systems: a systematic literature review. *Comput. Ind.* **120**, 103224 (2020)
5. Boyes, H., Hallaq, B., Cunningham, J., Watson, T.: The industrial internet of things (IIoT): an analysis framework. *Comput. Ind.* **101**, 1–12 (2018)
6. Sisinni, E., Saifullah, A., Han, S., Jennehag, U., Gidlund, M.: Industrial internet of things: challenges, opportunities, and directions. *IEEE Trans Ind Inf.* **14**(11), 4724–4734 (2018)
7. Grieves, M.: Digital twin: manufacturing excellence through virtual factory replication. *White Paper* **1**, 1–7 (2014)
8. Janiesch, C., Koschmider, A., Mecella, M., Weber, B., Burattin, A., Di Ciccio, C., et al.: The internet of things meets business process management: a manifesto. *IEEE Syst. Man Cybern. Mag.* **6**(4), 34–44 (2020)
9. Leotta, F., Mathew, J.G., Monti, F., Mecella, M.: Towards an information systems-driven maturity model for industry 4.0. In: *Enterprise, Business-Process and Information Systems Modeling: 23rd International Conference, BPMDS 2022 and 27th International Conference, EMMSAD 2022, Held at CAiSE 2022, Leuven, Belgium, June 6–7, 2022, Proceedings.* vol. 450. Springer, p. 351 (2022)
10. Ziolkowski, T., Koschmider, A., Schubert, R., Renz, M.: Process mining for time series data. In: *Enterprise, Business-Process and Information Systems Modeling: 23rd International Conference, BPMDS 2022 and 27th International Conference, EMMSAD 2022, Held at CAiSE 2022, Leuven, Belgium, June 6–7, 2022, Proceedings.* vol. 450. Springer, p. 347 (2022)
11. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: *Fundamentals of Business Process Management*, 2nd edn. Springer, Berlin (2018)
12. Van Der Aalst, W.: *Process Mining: Data Science in Action*, vol. 2. Springer, Berlin (2016)
13. Augusto, A., Conforti, R., Dumas, M., La Rosa, M., Maggi, F.M., Marrella, A., et al.: Automated discovery of process models from

- event logs: review and benchmark. *IEEE Trans. Knowl. Data Eng.* **31**(4), 686–705 (2019)
14. Carmona, J., van Dongen, B.F., Solti, A., Weidlich, M.: *Conformance Checking - Relating Processes and Models*. Springer, Berlin (2018)
 15. Kephart, J.O., Chess, D.M.: The vision of autonomic computing. *Computer* **36**(1), 41–50 (2003)
 16. Myers, K.L., Berry, P.M.: *Workflow management systems: an AI perspective*. AIC-SRI report, pp. 1–34 (1998)
 17. Marrella, A.: Automated planning for business process management. *J. Data Semant.* **8**(2), 79–98 (2019)
 18. Ghallab, M., Nau, D., Traverso, P.: *Automated Planning and Acting*. Cambridge University Press, Cambridge (2016)
 19. Aeronautiques, C., Howe, A., Knoblock, C., McDermott, I.D., Ram, A., Veloso, M., et al.: *Pddl the planning domain definition language*. Technical Report (1998)
 20. van der Aalst, W.M.: Business process management: a personal view. *Bus. Process Manag. J.* **10**(2) (2004)
 21. Connolly, P.E.: CAD software industry trends and directions. *Eng. Des. Graphics J.* **63**(1) (1999)
 22. Mas, F., Arista, R., Oliva, M., Hiebert, B., Gilkerson, I., Ríos, J.: A review of PLM impact on US and EU aerospace industry. *Procedia Eng.* **132**, 1053–1060 (2015)
 23. Manavalan, E., Jayakrishna, K.: A review of Internet of Things (IoT) embedded sustainable supply chain for industry 4.0 requirements. *Comput. Ind. Eng.* **127**, 925–953 (2019)
 24. Barreto, L., Amaral, A., Pereira, T.: Industry 4.0 implications in logistics: an overview. *Procedia Manuf.* **13**, 1245–1252 (2017)
 25. Kletti, J.: *Manufacturing Execution Systems-MES*. Springer, Berlin (2007)
 26. Scholten, B.: *MES guide for executives: why and how to select, implement, and maintain a manufacturing execution system*. Intl. Soc. Autom. (2009)
 27. Jaskó, S., Skrop, A., Holzinger, T., Chován, T., Abonyi, J.: Development of MESs in accordance with Industry 4.0 requirements: a review of standard and ontology-based methodologies and tools. *Comput. Ind.* **123** (2020)
 28. De Giacomo, G., Favorito, M., Leotta, F., Mecella, M., Silo, L.: Digital twins composition in smart manufacturing via Markov decision processes. *Comput. Ind.* **149**, 103916 (2023)
 29. Gažová, A., Papulová, Z., Smolka, D.: Effect of business process management on level of automation and technologies connected to industry 4.0. *Procedia Comput. Sci.* **200**, 1498–1507 (2022)
 30. Steiner, F., et al.: Industry 4.0 and business process management. *Tehnički glasnik.* **13**(4), 349–355 (2019)
 31. Bazan, P., Estevez, E.: Industry 4.0 and business process management: state of the art and new challenges. *Bus. Process. Manag. J.* **28**(1), 62–80 (2022)
 32. Schöning, S., Ackermann, L., Jablonski, S., Ermer, A.: IoT meets BPM: a bidirectional communication architecture for IoT-aware process execution. *Softw. Syst. Model.* **19**, 1443–1459 (2020)
 33. Compagnucci, I., Corradini, F., Fornari, F., Polini, A., Re, B., Tiezzi, F.: A systematic literature review on IoT-aware business process modeling views, requirements and notations. *Softw. Syst. Model.* **22**(3), 969–1004 (2023)
 34. Torres, V., Serral, E., Valderas, P., Pelechano, V., Grefen, P.: Modeling of IoT devices in business processes: a systematic mapping study. In: 2020 IEEE 22nd Conference on Business Informatics (CBI). vol. 1. IEEE, pp. 221–230 (2020)
 35. Fattouch, N., Lahmar, I.B., Boukadi, K.: IoT-aware business process: comprehensive survey, discussion and challenges. In: 2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). IEEE, pp. 100–105 (2020)
 36. Graja, I., Kallel, S., Guermouche, N., Cheikhrouhou, S., Hadj, Kacem A.: A comprehensive survey on modeling of cyber-physical systems. *Concurr. Comput. Pract. Exp.* **32**(15), e4850 (2020)
 37. Friederich, J., Lugaresi, G., Lazarova-Molnar, S., Matta, A.: Process mining for dynamic modeling of smart manufacturing systems: data requirements. *Procedia CIRP.* **107**, 546–551 (2022)
 38. Ortmeier, C., Henningsen, N., Langer, A., Reiswich, A., Karl, A., Herrmann, C.: Framework for the integration of process mining into life cycle assessment. *Procedia CIRP.* **98**, 163–168 (2021)
 39. Rinderle-Ma, S., Stertz, F., Mangler, J., Pauker, F.: Process mining-discovery, conformance, and enhancement of manufacturing processes. In: *Digital Transformation: Core Technologies and Emerging Topics from a Computer Science Perspective*. Springer, pp. 363–383 (2023)
 40. Fernández, S., Aler, R., Borrajo, D.: Machine learning in hybrid hierarchical and partial-order planners for manufacturing domains. *Appl. Artif. Intell.* **19**(8), 783–809 (2005)
 41. Krueger, V., Rovida, F., Grossmann, B., Petrick, R., Crosby, M., Charzoule, A., et al.: Testing the vertical and cyber-physical integration of cognitive robots in manufacturing. *Robot. Comput.-Integr. Manuf.* **57**, 213–229 (2019)
 42. Carreno, Y., Pairet, È., Pétillet, Y.R., Petrick, R.P.A.: Task allocation strategy for heterogeneous robot teams in offshore missions. In: *International Conference on Autonomous Agents and Multiagent Systems, AAMAS*, pp. 222–230 (2020)
 43. Monti, F., Silo, L., Leotta, F., Mecella, M.: Services in smart manufacturing: comparing automated reasoning techniques for composition and orchestration. In: *Symposium and Summer School on Service-Oriented Computing*. Springer, pp. 69–83 (2023)
 44. Marrella, A., Mecella, M., Sardina, S.: Intelligent process adaptation in the SmartPM system. *ACM Trans. Intell. Syst. Technol.* **8**(2), 1–43 (2016)
 45. Catarci, T., Firmani, D., Leotta, F., Mandreoli, F., Mecella, M., Sapiro, F.: A conceptual architecture and model for smart manufacturing relying on service-based digital twins. In: *2019 IEEE International Conference on Web Services (ICWS)*. IEEE, pp. 229–236 (2019)
 46. Calamo, M., De Franceschi, A., De Santis, G., Leotta, F., Mazaroppi, C., Mathew, J.G., et al.: TopKontrol: a monitoring and quality control system for the packaging production. In: *Proceedings of the Research Projects Exhibition Papers Presented at the 35th International Conference on Advanced Information Systems Engineering (CAiSE 2023)*, Zaragoza, Spain, June 12–16 (2023)
 47. Amadori, F., Bardani, M., Bernasconi, E., Cappelletti, F., Catarci, T., Drudi, G., et al.: Electrospindle 4.0: Towards zero defect manufacturing of spindles. In: *Joint Proceedings of RCIS 2022 Workshops and Research Projects Track co-located with the 16th International Conference on Research Challenges in Information Science (RCIS 2022)*, Barcelona, Spain, May 17–20 (2022)
 48. Agostinelli, S., De Luzi, F., Manglaviti, M., Mecella, M., Monti, F., Petriccione, F.M., et al.: BinTraWine-blockchain, tracking and tracing solutions for wine. In: *CAiSE Research Projects Exhibition*; pp. 44–51 (2023)
 49. Ciolek, D., D’Ippolito, N., Pozanco, A., Sardiña, S.: Multi-tier automated planning for adaptive behavior. In: *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 30, pp. 66–74 (2020)
 50. Carrera, J.L.V., Zhao, Z., Braun, T.: Room recognition using discriminative ensemble learning with hidden Markov models for smartphones. *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1–7 (2018)
 51. Koschmider, A., Janssen, D., Mannhardt, F.: Framework for process discovery from sensor data. In: *EMISA*, pp. 32–38 (2020)

52. van Zelst, S.J., Mannhardt, F., de Leoni, M., Koschmider, A.: Event abstraction in process mining: literature review and taxonomy. *Granul. Comput.* **6**(3), 719–736 (2021)
53. Mehdiyev, N., Fettke, P.: Explainable artificial intelligence for process mining: a general overview and application of a novel local explanation approach for predictive process monitoring. In: *Interpretable Artificial Intelligence: A Perspective of Granular Computing*. Studies in Computational Intelligence. Springer, pp. 1–28 (2021)
54. Dixit, P.M., Buijs, J.C., van der Aalst, W.M.: Prodigy: Human-in-the-loop process discovery. In: *2018 12th International Conference on Research Challenges in Information Science (RCIS)*. IEEE; pp. 1–12 (2018)
55. Ziolkowski, T., Koschmider, A., Kröger, P., Devey, C.: Outlier quantification for multibeam data. *Informatik Spektrum.* **45**(4), 218–222 (2022)
56. Ter Hofstede, A.H., Koschmider, A., Marrella, A., Andrews, R., Fischer, D.A., Sadeghianasl, S., et al.: Process-data quality: the true frontier of process mining. *ACM J. Data Inf. Qual.* **15**(3), 1–21 (2023)
57. Dixit, P.M., Suriadi, S., Andrews, R., Wynn, M.T., ter Hofstede, A.H., Buijs, J.C., et al.: Detection and interactive repair of event ordering imperfection in process logs. In: *International Conference on Advanced Information Systems Engineering*. Springer; pp. 274–290 (2018)
58. Koschmider, A., Kaczmarek, K., Krause, M., Zelst, S.J.: Demystifying noise and outliers in event logs: review and future directions. In: *International Conference on Business Process Management*. Springer, pp. 123–135 (2021)
59. Koschmider, A., Aleknytytė-Resch, M., Fonger, F., Imenkamp, C., Lepsien, A., Apaydin, K., et al.: Process mining for unstructured data: challenges and research directions
60. Fonger, F., Aleknytytė-Resch, M., Koschmider, A.: Mapping time-series data on process patterns to generate synthetic data. In: *International Conference on Advanced Information Systems Engineering*. Springer; p. 50–61 (2023)
61. Patki, N., Wedge, R., Veeramachaneni, K.: The synthetic data vault. In: *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE; pp. 399–410 (2016)
62. Zisgen, Y., Janssen, D., Koschmider, A.: Generating synthetic sensor event logs for process mining. In: *International Conference on Advanced Information Systems Engineering*. Springer; p. 130–137 (2022)
63. Hankel, M., Rexroth, B.: The reference architectural model industrie 4.0 (rami 4.0). *ZVEI.* **2**(2), 4–9 (2015)
64. Duan, L., Da Xu, L.: Data analytics in industry 4.0: a survey. *Information Systems Frontiers.* pp. 1–17 (2021)
65. Murty, M.N., Devi, V.S.: *Introduction to Pattern Recognition and Machine Learning*, vol. 5. World Scientific, Singapore (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Flavia Monti is a Ph.D. student in Engineering in Computer Science at the Department of Computer, Control and Management Engineering Antonio Ruberti at Sapienza University of Rome. She got the bachelor's degree in Computer and System Engineering in 2018 at Sapienza University of Rome, where she also got the master's degree in Engineering in Computer Science in 2021. Her research interests mainly concern Industry 4.0 and Smart Manufacturing. Specifically, the integra-

tion of new technologies and approaches like computer vision, machine learning and artificial intelligence with the purpose of improving production quality, reducing costs, increasing machinery uptime, analyze failures and defects towards zero defect manufacturing.



Jerin George Mathew is a Ph.D. student of the National Doctorate in AI at Sapienza University of Rome. He pursued his Bachelor's degree in Computer Engineering at Roma Tre University in 2017 and then earned a Master's degree in Computer Engineering in 2020 at the same university. His research interests are in data management related topics, including entity resolution, knowledge graphs and assessment of fairness of rankings, as well as the application of NLP techniques in these

fields.



Francesco Leotta got his Ph.D. in Engineering in Computer Science at Sapienza in 2014, where he currently covers the position of Assistant Professor. Since the beginning of his research activity, he addressed several challenges related to how users interact with a smart space and how the environment senses the users and reactively perform actions to meet user requirements. In this research context, he developed an approach, called habit mining, where techniques typical of business process

management (BPM) and process mining can be adapted to user habit modelling and discovery. Beside this main research activity, his research interests cover advanced user interfaces, service-oriented architectures (SOA), matchmaking applied to entrepreneurship, and e-Government. He regularly serves as a reviewer for international conferences and journals in the fields of smart spaces and information systems engineering.



Agnes Koschmider from 2019 to 2022 was professor of business informatics at the Computer Science Institute of the University of Kiel. Before that she was a professor (associate professor) at the Poznań University of Economics and Business, visiting professor at the University of Cologne and Postdoc at the Institute for Applied Computer Science and Formal Description Methods at KIT. She completed her Ph.D. in 2007 and her habilitation in Applied Informatics in 2015 at KIT. Between

the Ph.D. and habilitation, several scholarships followed with research stays at the TU Eindhoven, the Politechnical University of Valencia, the Alpen-Adria-Universität Klagenfurt and IBM Research (Almaden).



Massimo Mecella Ph.D. in Engineering in Computer Science, is a full professor at Sapienza, where he is conducting research in the fields of software architectures, information systems, distributed middleware and service oriented computing, mobile and pervasive computing, process management, data and process mining, big data analytics, advanced interfaces and human-computer interaction, focusing on smart applications, environments and communities. He is author of more than 200

papers (h-index 40, cf. <https://scholar.google.com/citations?user=x844E6sAAAAJ>). He has been/is currently involved in several European and Italian research projects, and has been the technical manager of the projects WORKPAD and SM4All, coordinated by Sapienza, which already a few years ago were investigating ubiquitous processes. He has a large experience in organizing research events. He was the General Chair of CAiSE 2019 in Rome and of BPM 2021 in Rome, and he will act as General Chair of ICSOC 2023 in Rome. Currently, he is the director of the the B.Sc. in Engineering in Computer and Control Sciences, the M.Sc. in Engineering in Computer Science, and the MSc in Artificial Intelligence and Robotics offered by Sapienza.