



Multi-perspective Data Modelling in Cyber Physical Production Networks: Data, Services and Actors

Ada Bagozi¹ · Devis Bianchini¹ · Anisa Rula¹

Received: 23 March 2022 / Revised: 29 June 2022 / Accepted: 9 August 2022 / Published online: 20 August 2022
© The Author(s) 2022

Abstract

In recent years, Cyber Physical Production Systems and Digital Threads opened the vision on the importance of data modelling and management to lead the smart factory towards a full-fledged vertical and horizontal integration. Vertical integration refers to the full connection of smart factory levels from the work centers on the shop floor up to the business layer. Horizontal integration is realised when a single smart factory participates in multiple interleaved supply chains with different roles (e.g., main producer, supplier), sharing data and services and forming a Cyber Physical Production Network. In such an interconnected world, data and services become fundamental elements in the cyberspace to implement advanced data-driven applications such as production scheduling, energy consumption optimisation, anomaly detection, predictive maintenance, change management in Product Lifecycle Management, process monitoring and so forth. In this paper, we propose a methodology that guides the design of a portfolio of data-oriented services in a Cyber Physical Production Network. The methodology starts from the goals of the actors in the network, as well as their requirements on data and functions. Therefore, a data model is designed to represent the information shared across actors according to three interleaved perspectives, namely, product, process and industrial assets. Finally, multi-perspective data-oriented services for collecting, monitoring, dispatching and displaying data are built on top of the data model, according to the three perspectives. The methodology also includes a set of access policies for the actors in order to enable controlled access to data and services. The methodology is tested on a real case study for the production of valves in deep and ultra-deep water applications. Experimental validation in the real case study demonstrates the benefits of providing a methodological support for the design of multi-perspective data-oriented services in Cyber Physical Production Networks, both in terms of usability of the data navigation through the services and in terms of service performances in presence of Big Data.

Keywords Cyber physical production network · Data models · Service-oriented architecture · Digital thread · Industry 4.0

1 Introduction

The ever-growing application of digital technologies in modern smart factories has enabled the integration of product design, manufacturing processes and general collaboration across factories over the supply chain with the exploitation of data management techniques and service-oriented

architectures [1]. Recent approaches toward Industry 4.0 digital revolution shifted from the design of Digital Twins of machines or parts in isolation, as virtual representations of physical assets in a cyber-physical system (CPS), to the design of cyber physical production systems (CPPS). CPPS are hybrid networked cyber and engineered physical systems that record data (e.g., using sensors), analyze it using data-oriented services (e.g., over cloud computing infrastructures), influence physical processes and interact with human actors using multi-channel interfaces [2]. In industrial production environments, CPPS enable workers to supervise the operations of industrial machines or work centers, according to the Human-in-the-loop paradigm [3]. Moreover, research on digital threads (DTs) is gaining momentum, where DTs are conceived as the cyber side representation of a product, to enable the holistic view and traceability along its entire

✉ Devis Bianchini
devis.bianchini@unibs.it

Ada Bagozi
ada.bagozi@unibs.it

Anisa Rula
anisa.rula@unibs.it

¹ Department of Computer Science, University of Brescia, Via
Branze 38, Brescia, Italy

lifecycle [4]. Digital Thread, in its modern formulation, includes any data, functions, models and protocols related to the product and to the context where the product is designed, produced, used and maintained. As such, it goes beyond the understanding of the thread as mostly a collection of data, being the new digital counterpart of the established product lifecycle management (PLM) [5].

Cyber-Physical Production Systems and Digital Threads open the vision on the importance of data modelling and management in the smart factory over multiple perspectives, ranging over the product lifecycle [6] (e.g., product design, manufacturing, product quality control), the production process phases (e.g., process monitoring), work centers and shop floor. This vision is also coherent with the RAMI 4.0 Reference Architectural Model for Industry 4.0 [7], where the dimension of the lifecycle value stream (IEC 62890), spanning over the development, production and maintenance phases, is interleaved with the IEC 62264/IEC 61512 levels (from product to the connected world level) and with the smart factory layers (starting from industrial assets up to the business layer). Driving the digital transformation of the smart factory over these multiple dimensions leads to a full-fledged *vertical integration*, to coordinate work centers in a single smart factory, connecting the shop floor level with the business level, and a *horizontal integration*, where a smart factory participates to multiple interleaved supply chains, forming a so-called Internet of Production [8] or Cyber-Physical Production Network [9]. Although vertical integration has been extensively investigated in recent years, also thanks to the research on Digital Twins [10], horizontal integration still deserves further attention.

In this scenario, data collection, organization, analysis and exploration for implementing horizontal integration of the smart factory call for new methods and techniques for: (1) data modelling under multiple perspectives, namely, the product, the production process, the industrial assets and resources; (2) service-oriented design and implementation of functions spanning across all the layers of the smart factory and all the phases of the product lifecycle management; (3) design of proper access policies for all the involved actors, who have different permissions to access data and services through multiple channels by exploiting the modularity of service-oriented computing.

We propose an information model that contains all the data gathered at the shop floor level (for process monitoring, advanced anomaly detection and predictive maintenance applications on the involved work centers) as long as the product is moved forward in the production line. The model relies on the three perspectives of the product, the production process and the industrial assets used in the production. For each perspective, the physical world is connected with the cyber world, where both calculated indicators and collected sensor data are properly organised

to enable data analysis and exploration according to the three perspectives in an interleaved way. A portfolio of services has been also introduced to be invoked by actors in the production network, such as the main producer and the suppliers, to access data and monitor the production. Our approach is devoted to the production of costly and complex products, where the conceptualisation of a *smart product* as the integration of data according to the product, process and industrial assets perspectives aims at ensuring high product quality levels, long-lasting operations, less frequent and efficient maintenance activities and performance scalability over time.

We can summarize the contributions of this paper as follows:

- We provide a detailed multi-perspective data model by expanding the product perspective with different kinds of bill of materials over the production lifecycle and by detailing the process and industrial asset perspectives to highlight the involvement of different actors in the process phases and resource usage;
- We propose a methodology for the design of a multi-perspective data-oriented service portfolio; the methodology starts from the goals of the actors in the network, as well as their requirements on data and functions, and produces the data model and the portfolio of services, distinguishing among different kinds of services (e.g., collect, monitor, dispatch and display services) over the product, process and industrial assets perspectives;
- We designed data and service access policies on top of the production network, to let actors access data and use services according to their respective permissions.

The result of the above contributions, which are specifically targeted at the implementation of horizontal integration of the smart factory, is a *cyber physical model (CPM)*. A CPM is composed of the multi-perspective data model, a portfolio of data-oriented services and a multi-perspective dashboard designed for data visualisation purposes. The methodology is tested on the real case study, i.e., the production of valves to be used in deep and ultra-deep water applications. In particular, we focus here on a specific use case, namely, production scheduling, to effectively demonstrate the benefits of providing a methodological support for the design of multi-perspective data-oriented digital services in Cyber Physical Production Networks.

This article is an extension of the work in [11]. With respect to the previous work, where we sketched the main concepts of the multi-perspective data model and we implemented an ad-hoc set of collect, dispatch and display services to demonstrate the feasibility of the approach, the novel contributions here are the following:

- We provide a detailed version of the data model, as the result of the application of a set of refinement primitives to the model introduced in the previous work; through the model refinement, we introduced different kinds of Bill of Materials, corresponding to distinct phases in the product lifecycle, and we modelled the production process in a finer-grained way, distinguishing between phases executed internally to the main producer and operations outsourced through the collaboration with suppliers;
- We provide a methodology to guide the actors to design the multi-perspective data-oriented services in a systematic way, going beyond the ad-hoc implementation described in the previous work; furthermore, we added a fourth category of services (*monitor*) to serve additional requirements and address Big Data issues;
- We provide a data access model that is completely new in this work, properly evolving the dashboard GUI to customise the visualised features through access permission policies;
- The usability experiments have been revised to include the new features; in particular, questions to users have been detailed by making explicit the perspectives that have been involved in each question, in order to better identify usability drawbacks; finally, the performance of the newly introduced monitor services for anomaly detection has been tested.

The paper is organised as follows: Sect. 2 introduces the real case study and provides an overview of the methodology; Sects. 3–6 describe the methodological steps in more details; in Sect. 7 implementation issues are addressed; usability and performance experiments are described in Sect. 8; cutting edge features of the approach compared to the state of the art are presented in Sect. 9; finally, Sect. 10 closes the paper and sketches future research directions.

2 Case Study and Methodology Overview

The methodology for designing the portfolio of multi-perspective data-oriented digital services in a Cyber-Physical Production Network has been developed and applied in an industrial regional project in Northern Italy. Specifically, the production of valves to be used in deep and ultra-deep water applications has been considered. Nevertheless, the methodological steps have been conceived for a generic production network in the manufacturing sector.

Valves of deep and ultra-deep water applications are placed in prohibitive environments and, once installed, are difficult to remove and maintain over time and require high-quality levels. Figure 1 provides an overview of the considered valve production case study. The production of the final product (i.e., the valve), its installation on-field and maintenance are time-consuming and costly tasks and the product is delivered on-demand in low volumes, very often designed to serve specific needs of clients. The case study

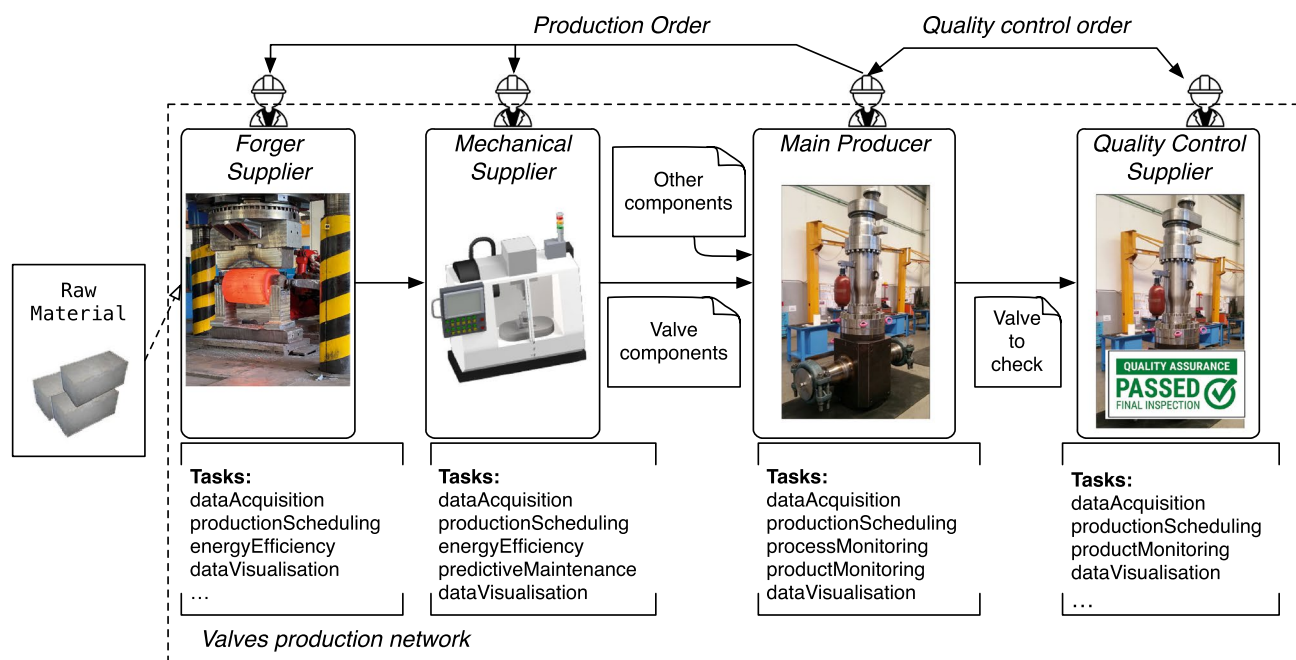


Fig. 1 Production steps of deep and ultra-deep water valves: from the top to the bottom in the figure, involved actors, valves production steps and sample tasks are shown

targets different categories of actors involved in this production network or similar ones in the manufacturing sector: the main producer (e.g., the valves producer); the raw materials suppliers (e.g., the forger); the suppliers of mechanical processing tasks (e.g., the supplier who is in charge of machining raw materials provided by the forger to be assembled in the valves); the suppliers of specific tools used in the production stages (e.g., to perform quality controls on the valves).

In the production network, collaboration among actors is the main aspect in order to deliver on time high-quality products. The main producer and the suppliers may perform different tasks, requiring data owned by other actors and services delivered across actors' boundaries. The tasks for the case study of deep and ultra-deep water valves production are partially sketched at the bottom of Fig. 1 for each actor. For instance, the supplier of mechanical processing tasks transforms raw material into valve sub-parts by using multi-spindle machines, that are specifically conceived for flexible production. This supplier is interested in monitoring the performance of the spindles by implementing predictive maintenance and anomaly detection techniques. On the other hand, the same kind of supplier and the raw materials supplier (i.e., the forger) are interested in adopting energy efficient strategies on their assets, given the high cost of energy in this sector. Furthermore, the main producer (valve producer) is interested in the optimisation of the production scheduling, that involves all actors. Design strategies of the services in the portfolio should take into account several challenges, such as modularity, platform independence and reuse, that are very common in service-oriented applications but have not been properly addressed in existing approaches providing advanced Industry 4.0 solutions. Moreover, existing solutions are focused on specific needs in the production network, such as energy efficiency [12], anomaly detection [13] and predictive maintenance [14] on work centers, production scheduling [15] and process monitoring [16], without considering a holistic view that includes multiple aspects together. In this context, the following challenges can be identified.

Strong correlations among observed phenomena Observed phenomena in the production network are often *strongly interleaved* with each other. For example, a peak in energy consumption does not always mean a low-efficiency issue. Indeed, such a peak may be due to intense production scheduling. Services in the portfolio might be used to make these correlations explicit, bringing to advanced functionalities. For instance, by correlating the output of the production scheduler with the energy efficiency optimisation services, the above situations can be properly distinguished and managed. On the other hand, if the production scheduling is not recognised as the cause of the energy peak, the peak may be due to disruptions of the machines used throughout the production, properly detected through an anomaly detection

service. Similarly, problems raised during product quality controls may be correlated with anomalies raised on the machines of one of the participants in the production process. To implement all these services, acquisition, processing and visualisation of data under the three perspectives of product, process and industrial assets are required for all the actors of the production network.

Modular design and reuse of functionalities Service design may benefit from an organisation of services within a *reusable service catalogue*. The design of services in the catalogue can be driven by the three perspectives in the data model (product, process and industrial assets) and by the data flow in the production network, namely data collection services to gather data from the shop floor, dispatch services to share data over the production network, data monitoring and display services for different actors taking into account their access permissions on data and functions. Therefore, a service design approach should be conceived according to the multi-perspective data model and to the information flow throughout the production network.

Data-intensive functionalities Service design according to the different perspectives over the production data aims at serving several advanced applications, namely product traceability and monitoring (over the product perspective), production process monitoring (over the process perspective) and predictive maintenance and anomaly detection (over the industrial assets perspective). The abovementioned applications also rely on data characterised by high volumes and collection and processing speed never seen before, as well as by heterogeneity and variety in data formats (Big Data). Therefore, the methodology for the design of data-intensive services must also take into account volume, velocity and variety of data collected from the Cyber Physical Production Network.

2.1 Methodology at a Glance

Figure 2 provides an overview of the methodology we propose for designing multi-perspective data-oriented services for the CPM. The methodology comprises four steps defined in the following.

- (1) *Requirements analysis* provides a clear understanding of the general context in which the CPM will be adopted, through the identification of actors and their objectives, data and functional requirements; actors' objectives are formalised using Use Cases Diagrams, data and functional requirements are specified through a *data model template* and through sequence diagrams, respectively. This step will be detailed in Sect. 3.
- (2) The *multi-perspective data design* step produces a detailed information model for the CPM, containing all the relevant entities, attributes and relationships

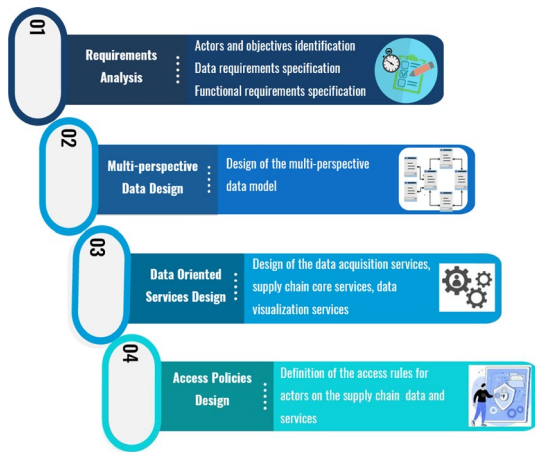


Fig. 2 Steps of the methodology for designing multi-perspective data-oriented services of the CPM

according to the Entity-Relationship formalism. The model is obtained by refining the data model template from the previous step by applying a set of *refinement primitives*, as detailed in Sect. 4.

- (3) The *data-oriented service design* step produces the catalogue of services implemented on top of the multi-perspective data model. Services in the catalogue are organised depending on the perspective(s) they are focused on and with respect to their role in the data flow of the Cyber Physical Production Network, that is, collect services, monitor services, dispatch services and display services. This step will be detailed in Sect. 5.
- (4) The goal of the last step is to define the policies to manage the access permissions on data and services for the actors of the CPM. The access policies are based on a conceptualisation of the production network at different levels of abstraction, distinguishing between the actors involved and their respective roles (e.g., main producer, raw materials suppliers). This step will be detailed in Sect. 6.

3 Requirements Analysis

In this section, we describe the first step of the methodology that regards the *Requirements analysis*, aimed at identifying the actors in the Cyber Physical Production Network, their objectives, their data and functional requirements, as detailed in the following.

Identification of actors and their objectives In this sub-step, the actors who are involved in the production network and their objectives are identified and formalised by adopting the well-known Use Cases notation. The methodology starts from a set of high-level business goals that have been considered relevant for the production of costly and complex

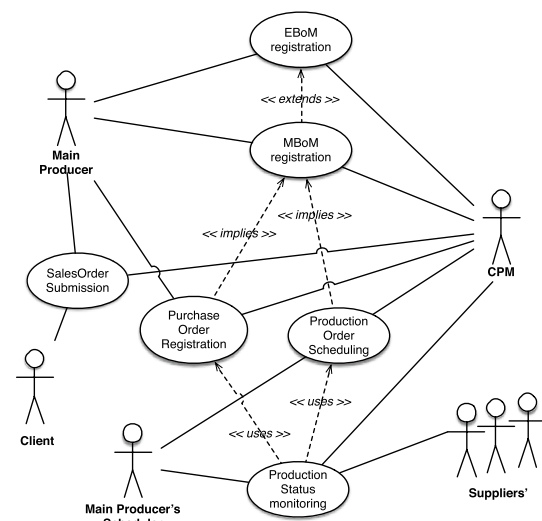


Fig. 3 Use Case Diagram for the production scheduling business goal

products, to ensure high product quality levels, long-lasting operations, less frequent and efficient maintenance activities, production scalability and sustainability over time. Such business goals have been identified through a feasibility study in partnership with the project. The business goals are production scheduling, process monitoring, product quality monitoring, and energy consumption optimisation.

For each business goal, we distinguish between the preparation of information, according to the product, process and industrial assets perspectives, and the daily execution of the actions to meet the objective. For example, considering the production scheduling business goal (see the Use Case Diagram shown in Fig. 3), information about the *Bill of Material (BoM)* is firstly registered, to prepare the schedule. Different kinds of BoM are taken into account throughout the entire product lifecycle, such as engineering BoM (EBoM), manufacturing BoM (MBoM) or service BoM (SBoM). Depending on the target phase in the production lifecycle, a specific BoM is considered, that brings its own information. For example, the MBoM concerns the production phase and also specifies who makes each part code and from whom the main producer will buy some items. For ensuring product change management, the MBoM also refers to an EBoM, that must be previously registered. The modelling of different kinds of Bill of Materials is a cutting edge feature of this work with respect to [11]. After EBoM and MBoM registration, the production scheduling is executed, thus requiring the previous submission of a sales order by the client (*sales order submission* use case) and the interaction between the production network actors to schedule the production. Production scheduling is performed both by issuing orders toward some suppliers for parts of the BoM that must be produced externally (*purchase order registration* use case) and

by planning internal production orders for the other parts (*production order scheduling* use case). This use case also includes the registration of industrial assets and resources required for production. The distinction between purchase orders and production orders is another difference in the data model with respect to [11]. Finally, the production status is constantly monitored (*production status monitoring* use case).

Data requirements analysis In this step, the data requirements necessary to achieve the business goals are specified. Here, the main concepts and relationships between them are used to produce a *data model template*, useful to guide the next methodological steps, according to the three perspectives of the product, the production process and the assets.

Figure 4 shows the considered multi-perspective data model template, already presented in [11]. This template will be valid in all the production networks that are similar to the real case study adopted here and can be conceived as a standardisation effort in the identification of data elements. Nevertheless, slight changes can be applied to the template depending on the specific case study in which it is applied. In the following we describe each perspective separately and how they relate to each other.

- **Product** Each product is composed of a set of parts which are identified by a part code and can be composed of other sub-parts. This relationship between product parts is represented through a recursive hierarchy making the navigation structure of a product flexible. The hierarchy represents the *Bill of Material (BoM)* which will be further specialised into different kinds of BoM in the data model design step.
- **Process** The process represents the various processing phases that must be executed to obtain the final product: each processing phase includes different sub-phases. A recursive hierarchy is used to model this relationship as well. The relationship between the product and the process can be very complex, depending on the organisation of the production network (see the distinction between purchase order and production order made above). For

example, some parts of the final product could be bought from suppliers instead of being produced internally or externally. These details will be refined in the data model design step.

- **Work centers and resources** The process is executed using work centers and resources. A work center comprises one or more machines. The hierarchical relationship in this case is between the work center and the component machines, that in a recursive way may be composed of other parts (for example, an oil pump, electrical engines, spindles, and so forth). The hierarchical organisation of assets reflects the IEC62264/IEC61512 standards of the RAMI 4.0 specification [7]. Resources can be of different kinds (e.g., operators, tools, software) and will be refined in the data model design step.
- **Parameters** Different kinds of parameters are used to monitor the behavior of the production network according to the three perspectives. On each product part in the BoM some *product parameters* are measured, for instance, to be used in quality controls. Values of these parameters must stay within acceptable ranges. On each process phase, proper *process parameters* are measured as well, concerning the phase duration which must be compliant with the end timestamp of each phase, as established by the production schedule. *Work center parameters* are gathered to monitor the working conditions of each work center at different levels. Parameters are monitored through proper thresholds, established by domain experts who possess the knowledge about the production process. Parameter bounds are used to establish if a critical condition has occurred on the monitored work center or one of its components. Parameters will be also refined in the data model design step.

Functional requirements analysis Implementing business goals corresponds to the design of several services in order to access, share and visualise the data between all the actors involved in the production network. During functional requirements analysis, each use case previously identified is refined to specify the dependencies with other use cases, defining the CRUD operations (Create-Read-Update-Delete) to be executed on entities and relationships of the data model template and using sequence diagram notation to model the interactions between the actors involved in the use case. To give a (non exhaustive) example of functional requirements analysis, let's consider the *Production Order Scheduling* use case in Fig. 3. This use case requires the execution of the *MBoM registration* use case (an MBoM must be previously registered in the data model of the CPM) and the execution of the *Sales Order submission* use case (a sales order must be received by the main producer from the client and properly registered in the data model of the CPM). Moreover, to implement the

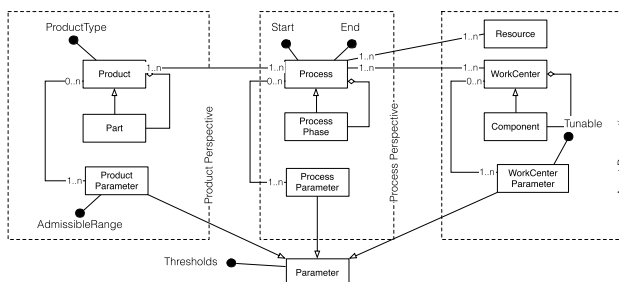


Fig. 4 Multi-perspective conceptual data model template for the CPM data requirements analysis

use case, process phases must be created and connected to the product to be produced and to the work centers and resources required for the production, respectively, that is:

```
create(Process)
connect(Process, Product)
connect(Process, Resource)
connect(Process, WorkCenter)
```

The above coarse-grained CRUD primitives are rather intuitive in their syntax and will be further refined in the multi-perspective data-oriented service design step, after data model refinement. Finally, the sequence diagram for the use case is designed, as shown in Fig. 5. In this case, the main producer forwards the Sales Order (SO), the MBoM, the production cycle (recipe) and resources to the scheduler for production scheduling. The scheduled production is communicated to the CPM. Some missing information in this schedule, concerning the delivery date of some product parts, must be asked to the parts suppliers by interacting with their schedulers, that will reply with the delivery date, returned back to the main producer's scheduler. This process is repeated until the final plan is obtained. According to this sequence diagram, and from the point of view of the CPM, the implementation of the *Production Order Scheduling* use case requires a *collect* service to receive the production order from the main producer's scheduler and a set of *dispatch* services to notify the production order to suppliers' schedulers and to share the received suppliers' delivery date with the main producer's scheduler. A *display* service can be implemented for visualising the production schedule. No *monitor* services are required for this Use Case (but, for instance, they will be required for the *Production Status Monitoring*, not detailed here).

4 Data-Oriented Data Design

During the data design step, the three perspectives in the data model template, namely product, process and asset perspectives, are refined with the application of *refinement primitives*, to create the final multi-perspective data model of the CPM. The refinement primitives are reported in Table 1 and are inspired by top-down/bottom-up database design strategies [17]. The resulting data model is shown in Fig. 6. In the figure, attributes that are added through entity refinement (primitive P3) are not shown for clarity purposes.

In the following, we describe in more detail how primitives are applied to refine the three perspectives. The application of refinement primitives is a step forward with respect to [11].

Product perspective refinement In the product perspective, the Product entity is enriched by adding product details (primitive P3). Different kinds of BoM are introduced, depending on the specific phase of the product lifecycle that is being considered (primitive P6). Each type of BoM is further refined by adding details on part codes (primitive P3). Moreover, the recursive hierarchy corresponding to the MBoM is re-factored (primitive P1). Part codes belonging to different types of BoM can be connected to each other, e.g., in order to manage change propagation over the whole product lifecycle (primitive P9). The connections of part codes across different kinds of BoM is a crucial aspect of what is recently referred to as Digital Thread [4]. Finally, the Sales Order is added and connected to the corresponding product. To this aim, the Client is created first (primitive P8), and then connected to the product (primitive P9). Finally, the newly created relationship is refined by introducing the Sales Order and the Sales Order Item entities (primitive P4).

Fig. 5 An example of sequence diagram designed during the functional requirements analysis for the *Production Order Scheduling* use case

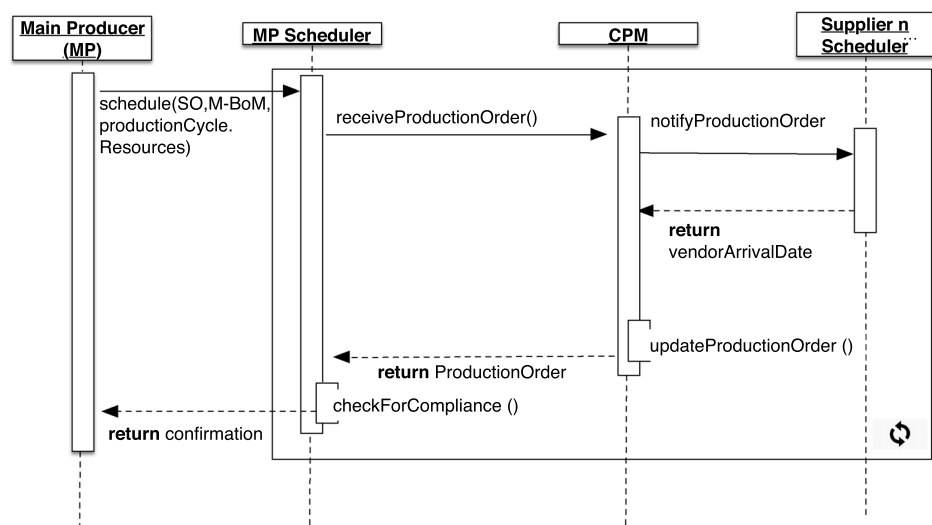


Table 1 Refinement primitives for extracting the final multi-perspective data model from the data model template

Primitive	Description
(P1) Hierarchy refactoring	Refactoring of hierarchy or transformation of a single entity into hierarchy (e.g., the MBoM is refactored as hierarchy)
(P2) Hierarchy differentiation	Different types of the same hierarchy are specified and connected (e.g., the BoM is distinguished into Manufacturing BoM and Engineering BoM)
(P3) Entity refinement	An entity is refined by adding attributes or other connected entities
(P4) Relationship refinement	A relationship between two entities is refined by adding attributes, that is, creating new entities connected to the one originally connected by the relationship (e.g., the Product Phase Execution entity, that refines the relationship between the production phases and the work centers)
(P5) Relationship refactoring	Relationships between entities are refactored by merging or moving them (e.g., the Product Phase Execution entity refactors the relationships between the production phases and the work centers)
(P6) Relationship differentiation	Different types of a relationship are specified and refined (e.g., the relationship between the Product and the BoM is differentiated to take into account different types of BoM)
(P7) Entity specialisation	A general entity is specialised into more specific ones (e.g., different kinds of resources are added, such as Operator and Equipment)
(P8) Additional entity	A new entity is added to the model (e.g., the Sales Order and the Sales Order Item) and related to other entities in the model
(P9) Additional relationship	A new relationship is added to the model (e.g., the relationship between Product and the client who issued the sales order)
(P10) Data format differentiation	Depending on the nature of the parameter (e.g., calculated measure, data streams), a different data storage system is adopted (e.g., relational or NoSQL databases)

Process perspective refinement In the process perspective, the main entity is represented by the production phase which is in turn organised according to a recursive hierarchy, where each phase is composed of a set of sub-phases, reaching a level of detail ranging from macro to micro industrial processes, and re-factored according to primitive P1. Production phases are related to the produced part code through a Production Order, that refines the relationship between the Product and the Process (primitive P4). Similarly, the relationship between the process phase and the work center is refined by introducing the Production Phase Execution entity which contains all details about the execution of a given production phase on a work center, such as the amount of setup and execution times (primitive P4). Indeed, the part code can be further distinguished among: (1) part codes that must be produced internally, connecting to a Production Order (PO); (2) part codes that must be bought externally, by one of the suppliers of the supply chain, connecting to a Purchase Order. According to this distinction, the following refinements are performed from the process perspective: (a) the Supplier entity is created (primitive P8); (b) the Supplier entity is related to the MBoM partcode using a new relationship (primitive P9); (c) the newly created relationship is refined by introducing Purchase Order and Purchase Order Item entities (primitive P4). Furthermore, the execution of some production phases can be outsourced to one of the suppliers using contract work orders. To include also this scenario in the model, a new relationship is created between the Production Phase Execution and the Supplier (primitive P9), further refined through the Contract Work Order entity

(primitive P4). In this way, the information model of the CPM brings together the viewpoints of all the actors of the production network. Finally, the production order is related to the client that issued the sales order for the corresponding product (primitive P9).

Asset perspective refinement The asset configuration perspective includes the resources involved in the realisation of the product (machinery, equipment, information systems, human resources). We distinguish between: (1) work center groups, representing categories of work centers (e.g., pumps, ovens, etc.) to be used in specific production phases, but not yet instantiated on a specific asset; (2) work centers, that is, machines that are used in the manufacturing process and are hierarchically organised, according to the RAMI 4.0 Reference Architectural Model for Industry 4.0 [7]; (3) resources, such as personnel or machine equipment, used or involved during the execution of the production phases. To refine the asset perspective, the following primitives are applied: (a) the relationship between the production phase and work centers is split (primitive P6), thus distinguishing between the production phase execution, which refers to a specific instance of work center, and the work center group to be used and not yet instantiated for a given production phase; (b) the relationship between resources and production phases is refactored, connecting resources directly to the work centers in which they are used (primitive P5); (c) work centers are hierarchically organised to represent the recursive composition of complex plants or shop floors into machines, in turn composed of other components (primitive

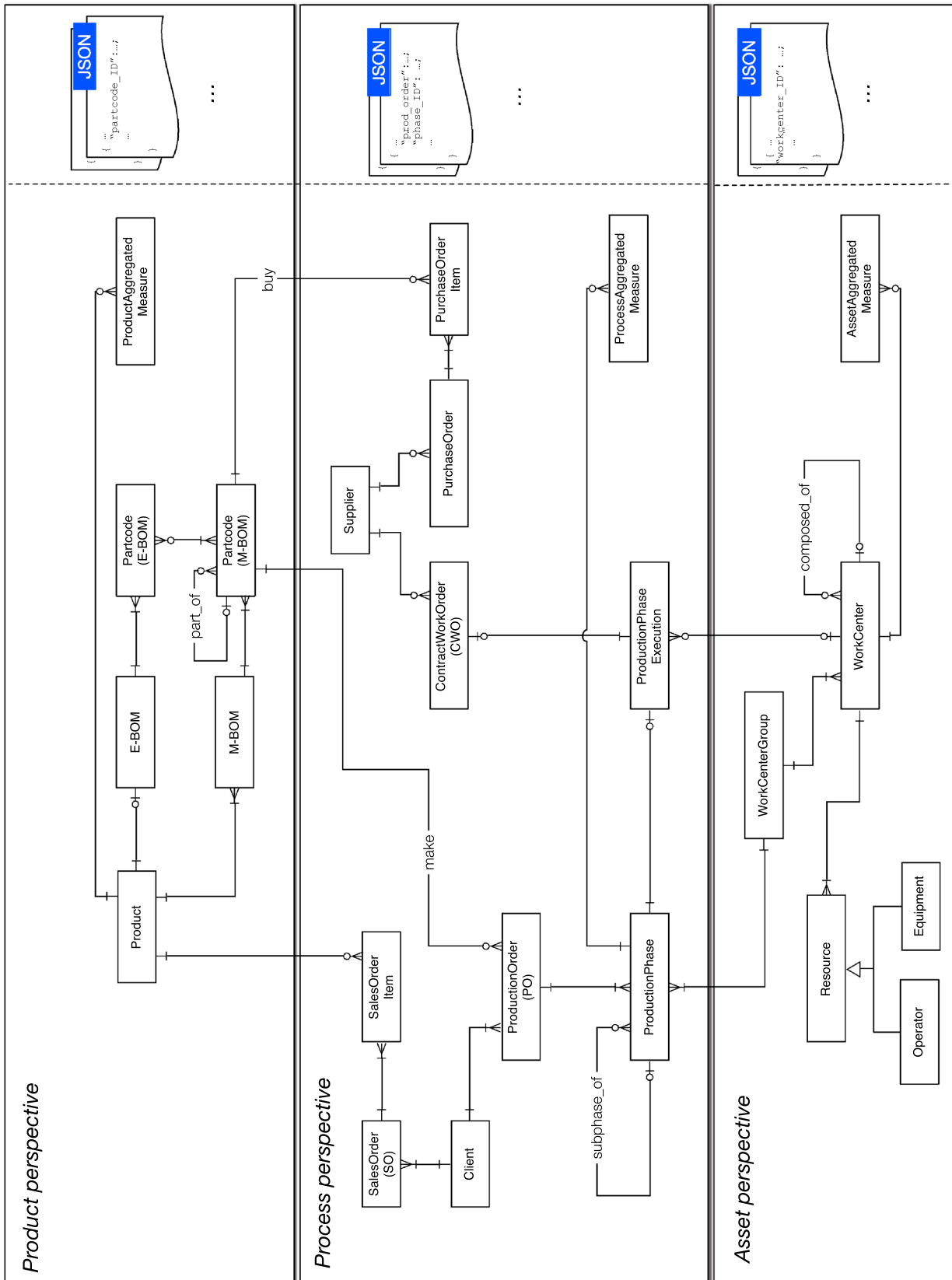


Fig. 6 Multi-perspective data model of the CPM

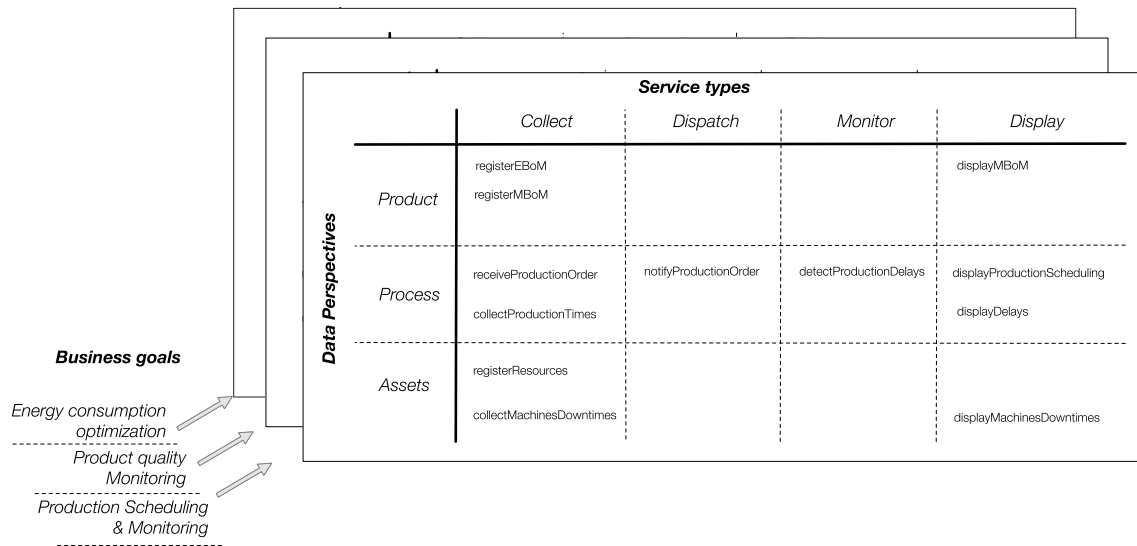


Fig. 7 Multi-perspective service map: services are categorised among collect, dispatch, monitor and display services and are intersected over the product, process and assets perspectives; furthermore, ser-

vices are grouped according to the target business goal. Actors in the production network have the permissions to access and invoke services according to access policies

P1); (d) resources are further specialised into different subtypes such as operators and equipment (primitive P7).

Parameters are distinguished among aggregated measures (e.g., KPI) and data streams collected as-is from the field (e.g., sensors data acquisition). Two options bring to the adoption of different storage systems, according to primitive P10. In the case of aggregated measures a new entity is created for each measure (fact table) containing records with the calculated value, a time reference (depending on the granularity of the aggregation, e.g., monthly/weekly/daily and so forth) and relationships with the target entities of the multi-perspective data model. For example, the number of non compliant parts is a calculated measure, that refers to the product from the corresponding perspective. Another example of a calculated measure is the overall equipment effectiveness (OEE), referring to a specific work center, quantifying its availability over time, its efficiency and quality. Data streams are stored within a document-store NoSQL database (MongoDB, see Sect. 7). For each measurement, a JSON document is registered, reporting the value of the measure, the timestamp, and the ID of the target entities. The data stream collected from a vibration sensor on a specific work center or component is an example of this kind of parameter. JSON documents can be organised in different collections, e.g., with respect to the physical parameter that is being measured (vibration, electrical current, temperature).

5 Multi-perspective Services Design

A repository of multi-perspective services is designed on top of the multi-perspective data model of the CPM. Services are categorized according to different service types and over the three perspectives of the product, process and assets, as shown in Fig. 7. Moreover, services that operate on high volume data streams are labelled as data-intensive services.

For what concerns service types, we distinguish among *collect services*, *monitor services*, *dispatch services* and *display services*. Collect services are used by the production network actors to transfer data from the physical side of the production network toward the CPM on the cloud. In the running example, the `receiveProductionOrder` service in the production scheduling business goal is an example of a collect service. Monitor services are used to proactively raise warnings/errors about the phenomena occurring on the physical production network such as threshold-based anomaly detection services [18]. Dispatch services are used to share data across the actors of the production network. For instance, the `notifyProductionOrder` dispatch service is used to propagate the reception of a production order from the main producer to the suppliers, in order to require the confirmation/communication of the suppliers' delivery date. Finally, display services are exposed to visualise data on the dashboard, such as the service to display the production scheduling and the services implemented in the production status monitoring use case.

Collect, monitor, dispatch and display services rely on the perspectives of the CPM data model and are classified accordingly. For example, the

Table 2 List of CRUD actions on entities and relationships of the data model, implemented within data-oriented services

Action	Description
Create(E_i)	Creates an instance of entity E_i
Create &Connect($E_i, R_1 \dots R_n$)	Creates an instance of entity E_i and connects it to other related instances through relationships $R_1 \dots R_n$
Connect(R_i)	Creates an instance of relationship R_i
Read($E_1 \dots E_n$)	Retrieves instances of entities $E_1 \dots E_n$ and mutual relationships
Update(E_i)	Modifies an instance of entity E_i
Update(R_i)	Modifies an instance of relationship R_i
Delete(E_i)	Deletes an instance of entity E_i
Disconnect &Delete ($R_1 \dots R_n, E_i$)	Deletes an instance of entity E_i after disconnecting it from other instances related through relationships $R_1 \dots R_n$
Disconnect(R_i)	Deletes an instance of relationship R_i

notifyProductionOrder service is focused on the process perspective, while the registerMBoM service is focused on the product perspective. Such a service organisation is promoted to foster the reuse of services in multiple use cases, within the same production network or across different and interleaved ones. Furthermore, services can be accessed by the actors through the definition of proper access policies as detailed in the next section.

We formally define a multi-perspective data-oriented service as follows:

Definition 1 (Data-oriented service) A data-oriented service S_i is described as a tuple

$$S_i = \langle n_{S_i}, t_{S_i}, P_{S_i}, url_{S_i}, m_{S_i}, A_{S_i}, IN_{S_i}, OUT_{S_i} \rangle \quad (1)$$

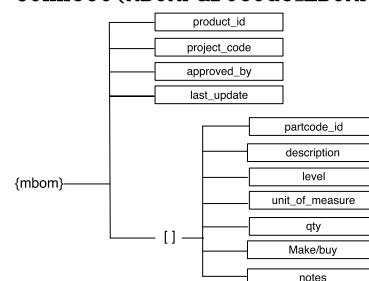
where: (1) n_{S_i} is the name of the service; (2) t_{S_i} is the service type, namely: collect, monitor, dispatch, display; (3) P_{S_i} is the set of perspectives on which the service is focused (product, process, assets); (4) url_{S_i} is the endpoint of the service for its invocation; (5) m_{S_i} is the HTTP method (e.g., get, post) used to invoke the service; (6) A_{S_i} is the list of CRUD actions implemented in the service on the entities and relationships of the data model (see Table 2); (7) IN_{S_i} is the representation of the service input; (8) OUT_{S_i} is the representation of the service output. We denote with S the overall set of data-oriented services.

Services in S are implemented as RESTful services; therefore, they are also described through actions implemented in the service, corresponding to the Create-Read-Update-Delete actions, on the entities and relationships of the multi-perspective data model. Moreover, input/output data can be parameters (i.e., calculated measures or data-streams collected from the physical side of the production network). Services dealing with data-streams are called data-intensive services and need advanced techniques to

deal with data-streams. For instance, a predictive maintenance service is a data-intensive monitor service if it detects concepts drifting from massive data streams collected on the monitored machines. Service inputs/outputs are represented in JSON format and modelled in terms of information model entities and their relationships. In the following, we report the example of service specification designed for MBoM registration.

```

Name      register_MBoM
Type      Collect
PSi     {Product}
urlSi    $baseUrl$/api/register/mbom/
mSi     POST
ASi     Create&Connect (MBoM,
                MBoM_Product)
                Create&Connect (MBoMPartcode,
                MBoMPartcode_MBoM,
                MBoMPartcode_MBoMPartcode)
                Connect (MBoMPartcode_EBoMPartcode)
    
```



IN_{S_i}
 OUT_{S_i} -

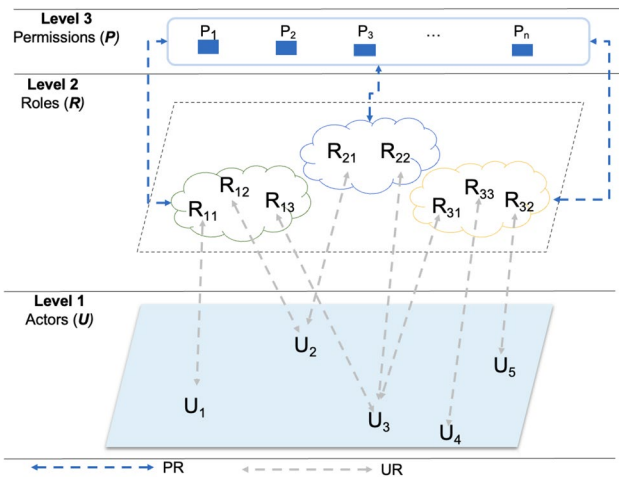


Fig. 8 Role-Based Access Control model for the CPM

6 Data and Services Access Policies

Each actor must have access only to the data and services for which he/she is authorized, i.e., related to his/her own internal company performance and his/her role in the production network. For what concerns the invocation of collect and dispatch services, according to the sequence diagram that models the interactions between actors, token-based mechanisms (e.g., OAuth) can be adopted. Monitor services process collected measures, under the three perspectives, and raise alarms/warnings in case of values overtaking predefined thresholds or in case of concept drifts in the data streams. These warnings/errors are properly notified to target actors identified at design time (e.g., a problem on a

work center may be notified to the owner of the machine, while a delay on a process phase execution may be notified to the actors responsible for the phase and possibly to the main producer). For what concerns display services, they are provided through a web-based dashboard, where users' authentication and authorisation mechanisms have been implemented. Specifically, an authorisation states which actor (e.g., the supplier) can perform which action (e.g., read, write) on which object (e.g., a table, a set of attributes or a set of records). The introduction of access policies for data and services is another contribution with respect to [11].

A large number of research efforts have been devoted to surveys on access control models [19]. Among them, the most used model is the Role-Based Access Control model RBAC [20]. According to this model, the concepts of users, roles, permissions and actions are conceived. Users may belong to different roles, while permissions to execute actions are applied on roles, thus avoiding replicating the same permissions for all users belonging to the same role. The fundamental constraint behind the RBAC model is that both role assignment and permissions assignment must be authorised, avoiding a single actor to control himself/herself.

Figure 8 shows the RBAC model applied to the CPM, which considers three levels, each defining the users *U*, the roles *R* and the permissions *P*. At the first level, the users in the set *U* correspond to one of the actors of the production network. At the second level, actors are assigned to one or more roles in the set *R*, determining their involvement in the production network. Examples of roles are the main producer, the suppliers of mechanical manufacturing tasks, the suppliers of raw materials, the client. This separation of concerns eases the management of complexity in

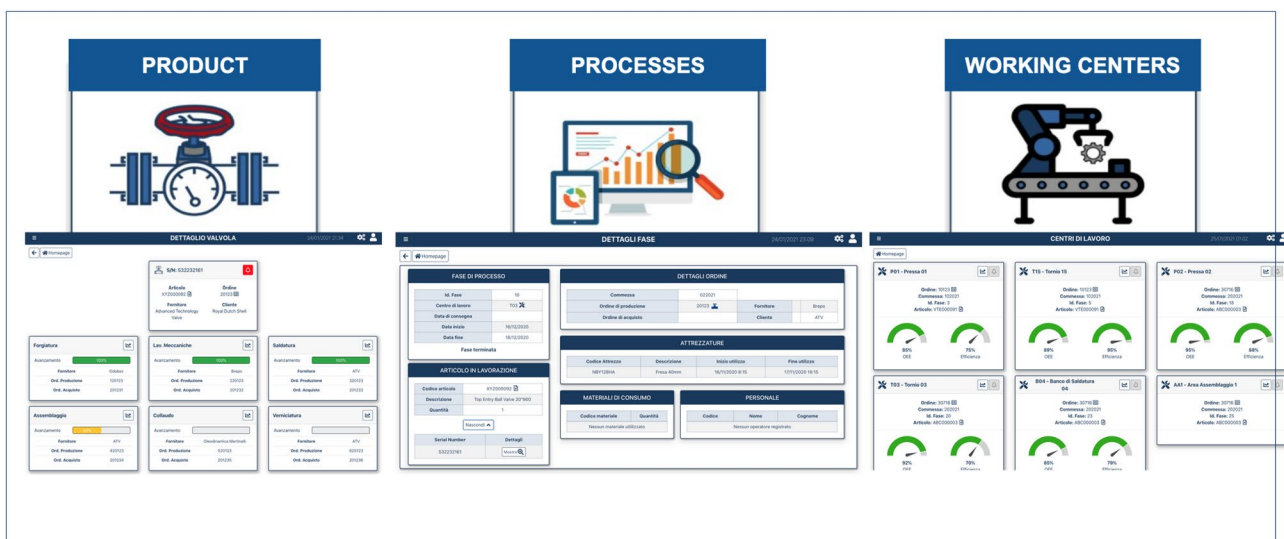


Fig. 9 Screenshots of the multi-perspective dashboard of the CPM

interleaved production networks, where a single actor may be involved in different networks, being a main producer in one of them and a supplier in another one. A permission $p_i \in P$ identifies which object and what properties/fields of the object can be accessed, by which actor and what is the allowed operation on the object and its properties/fields (in terms of CRUD actions). Objects can be entities in the multi-perspective data model of the CPM or collections of documents stored within NoSQL databases in case of data

streams. In the case of an entity, attributes and filtering criteria on the instances of the entity are also specified in the permission. In the case of JSON documents in a collection, filtering criteria are specified in order to filter out not allowed documents. Permissions specifications correspond to views on top of the data model and are implemented within display services. Therefore, the RBAC model for the CPM is formally defined according to the following definition. Figure 9 shows the access to the dashboards of the CPM including different perspectives according to the users access control.

Definition 2 (Role-based Access Control Model of the CPM)

The RBAC model for the CPM is formally defined through the tuple $\langle U, R, S, P, PR, UR \rangle$, where:

- U is the set of actors in the production network;
- R is the set of roles;
- S is the set of display services;
- P is the set of permissions; each permission $p_i \in P$ is defined as $\langle obj_i, \pi_i, \sigma_i, op_i, s_i \rangle$, where: (1) obj_i is the entity/document to access; (2) π_i is the set of attributes that must be accessed (not specified in case of JSON documents); (3) σ_i is the filtering condition to select records of the entity or JSON documents; (4) op_i is the CRUD action to

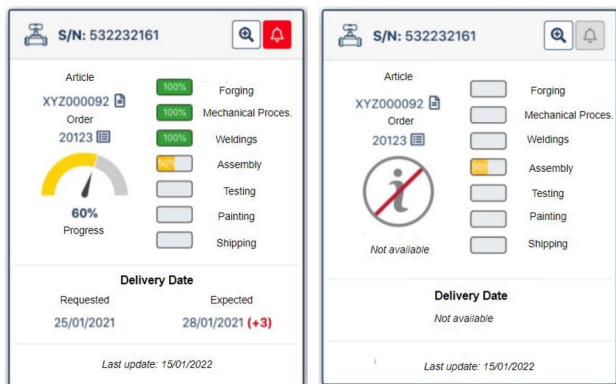


Fig. 10 The visualization of the product synoptic tile (valve production) as seen by the main producer (left) and by one of the suppliers with restricted access permissions (right)

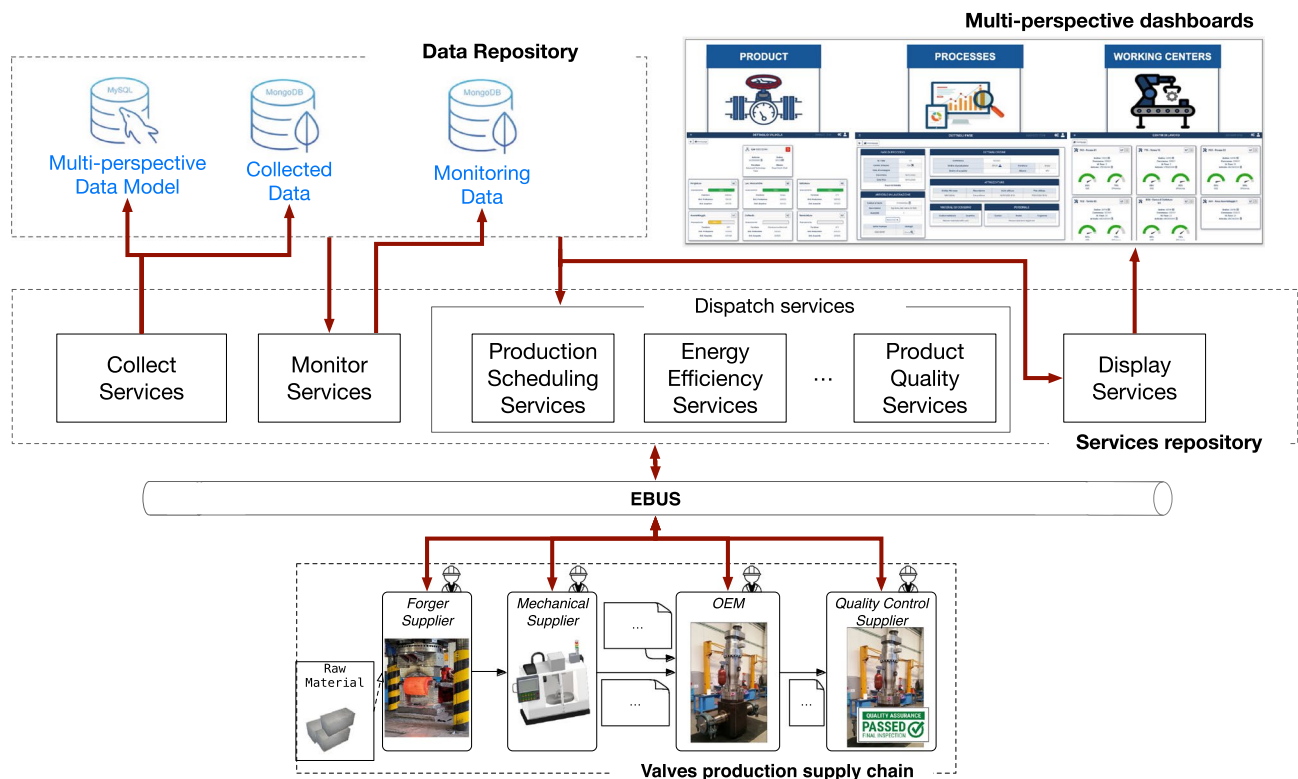


Fig. 11 Architecture that implements the CPM (composed of the data repository, the service repository and the multi-perspective dashboard)

be allowed; (5) $s_i \in S$ is the display service implementing the permission;

- PR is the permission assignment function that assigns permissions to roles (i.e., $RS \subset P \times R$);
- UR is the actor assignment function that assigns actors to roles (i.e., $UR \subseteq U \times R$).

6.1 Web-based Dashboard

Display services populate the web-based dashboard, that the actors use for data exploration. From the home page of the dashboard, it is possible to start the data exploration by following one of the three perspectives, namely, product, process and industrial assets. Each perspective brings to a UI component (tile) implemented using ReactJS libraries: (1) the *product synoptic* tile allows an exploration from the product perspective; (2) the *process phases* tile allows an exploration from the process perspective; (3) the *working centers* tile allows exploration from the industrial asset perspective. The management of different access permissions through display services leads to tiles customization. Therefore, it is possible through the dashboards to configure what to show and what to hide for each actor according to his/her role. For example, considering the main producer, the product synoptic tile is visualized without any limitations as shown in Fig. 10 (left). On the other hand, the mechanical supplier actor has a limited view, restricted to his/her own production phase, without displaying details about the other phases, the delivery date and the warning on the final product quality issues, as shown in Fig. 10 (right).

7 Implementation

The approach presented in this paper brings to the architecture shown in Fig. 11. The architecture implements the four business goals of the project, namely production scheduling and monitoring, product quality monitoring and energy consumption optimisation. However, it can be seamlessly extended to include other business goals by adding further collect, dispatch, monitor and display services, following the same methodological steps. The architecture is composed of three main parts: the *service repository*, the *data repository* and the *multi-perspective dashboard*. The services interact with the actors of the physical production network, who expose their own services to exchange data with the CPM. Services in the repository have been implemented using the RESTful technology and JSON for input/output representation. The data repository includes structured data, stored within a relational database (MySQL) and semi-structured data, corresponding to data streams collected from the production network, stored within a MongoDB NoSQL installation.

Monitor services are implemented to detect anomalies that may lead to higher consumption or breakdown/damage of the work centers, production process failures or delays, and product quality issues. Monitor services implement either threshold-based techniques or advanced data stream processing solutions to detect anomalies within data streams in Industry 4.0 applications, by applying incremental clustering and data relevance evaluation techniques to manage the volume and velocity of incoming data. To this aim, the IDE-AaaS approach [21] has been integrated within the architecture. IDEAAaaS is a suite of advanced techniques specifically conceived for anomaly detection and predictive maintenance applications in presence of massive data streams. The aim of IDEAAaaS is to promptly guide exploration towards the portion of the data stream where critical conditions or problems over industrial assets occurred. The IDEAAaaS approach can benefit from the partition of incoming data streams according to different perspectives. For this reason, the application of IDEAAaaS techniques has been fostered through its integration on top of the multi-perspective data model of the CPM, as demonstrated in the next section on experimental results.

8 Experimental Evaluation

In this section, we describe the experimental validation of the methodological approach in the real case study, to demonstrate the effectiveness of the dashboard in supporting data exploration (Sect. 8.1). Moreover, in Sect. 8.2, we present a proof-of-concept validation of the performance of the advanced monitor services, dealing with Big Data, since these services are crucial to determine the applicability of the whole approach.

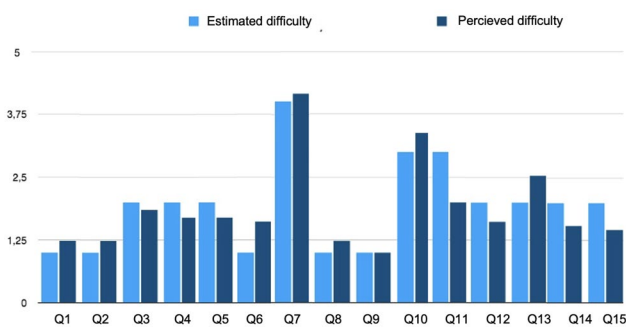
8.1 Usability Experiments

Data exploration is very limited in a platform covering one perspective at a time and it is too time-consuming and complex, since it requires accessing and querying distinct perspectives separately and then mixing up and combining the intermediate query results. Combining information across multiple perspectives is difficult, as a result of the lack of knowledge about how these perspectives influence each other in order to analyse different monitored phenomena. Therefore, our usability study will be focused on the multi-perspective data exploration. The main goal of this evaluation is to prove that multi-perspective exploration through the dashboard is better than single perspectives explored separately.

Experimental setup. We performed a user interaction experiment based on the *within-subjects design* approach, where all the participants take part in all the experiments in every condition. The users will perform an activity in an

Table 3 Task and questions defined for each task having one, two or three perspectives, issued during usability experiments

Tasks	Questions	Perspective
Check production scheduling	Q1 What is the expected delivery date for a given valve?	Product, Process
	Q2 On which work center the turning operation of the valve is scheduled?	Product, Process, Asset
	Q3 When is the production of the two valves belonging to that order is expected be delivered with respect to the requested date?	Product, Process
	Q4 When did the refinement operation (which is a sub-step of the welding process) begin for order number 20123?	Process
	Q5 When did the refinement operation (which is a sub-step of the welding process) end for order number 20123?	Process
Check process monitoring	Q6 How many valves are in production for a given item?	Product, Process
	Q7 What are the problems, if any, that occurred on industrial assets during the production of the valve VSS000799?	Product, Process, Asset
Check working center monitoring	Q8 Show the value of the Overall Equipment Effectiveness index calculated on the working center	Asset
	Q9 Which item is currently produced in a given working center?	Product, Process, Asset
	Q10 What is the maximum value of the temperature detected on the press 02 in the forging phase for the order number 30716	Product, Process, Asset
	Q11 What is the code of the cutter used in the turning phase of the order 20213?	Process, Asset
Check product quality	Q12 Indicates which of the following employees participated in the quality control operation (part of the testing phase) for the given order number?	Process
Check product monitoring	Q13 Which are the components that are part of the BOM items?	Product
Check actors	Q14 Who is the supplier for the cutter used during the turning phase of order 20213?	Process, Asset
	Q15 Who is the customer to whom the valve with serial number X refers to?	Product, Process

**Fig. 12** Comparison between the degree of estimated and perceived difficulty of tasks required to users participating to the usability experiments

environment similar to the one in which they actually operate. We provide tasks to be executed by the users that include multiple perspectives e.g., *When is the delivery date (process perspective) expected for the production of the valve having serial number 532232161?* (product perspective) or *Which work center or components (assets perspective) have been affected by anomalies in their electrical current absorption, given that the produced valve (product perspective) presented quality parameters out of admissible range?*

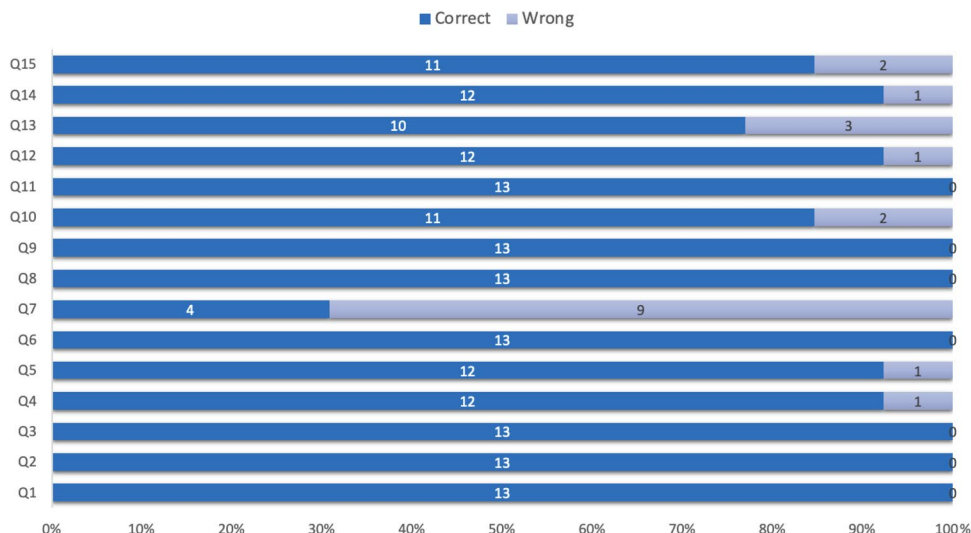
The experiment includes four steps: (1) *test preparation*, to verify and check that everything is working correctly before submitting to users; (2) *test presentation*, to provide a

welcome message to the users and explain the purpose of the test; (3) *test execution*, to let the examiner remotely monitor the test, in which the user operates, for not influencing the user during the experiment; (4) *debriefing*, performed through questionnaires, to collect the opinions of the users about the usability of the dashboard. During the experiments, 15 questions/tasks have been prepared (as shown in Table 3), some of them involving just one perspective (e.g., *Q13—Which are the components that are part of the BoM items?*), two perspectives (e.g., *Q6—How many valves are in production for a given item?*) or all the three perspectives together (e.g., *Q9—Which item is currently produced in a given working center?*).

Experiments participants The profile of involved participants ranges from less experienced to high qualified users. In particular, users belong to three categories based on their level of expertise, or familiarity with data visualisation tools, and with management systems for knowledge issues relating to industrial processes, as follows:

- *Beginner user* category comprises users who are not familiar with the topics covered in the project;
- *Intermediate user* category comprises users who are fairly familiar with the issues of the management systems and related software, but without having a great experience;

Fig. 13 Comparison between correct and wrong answers in usability experiments



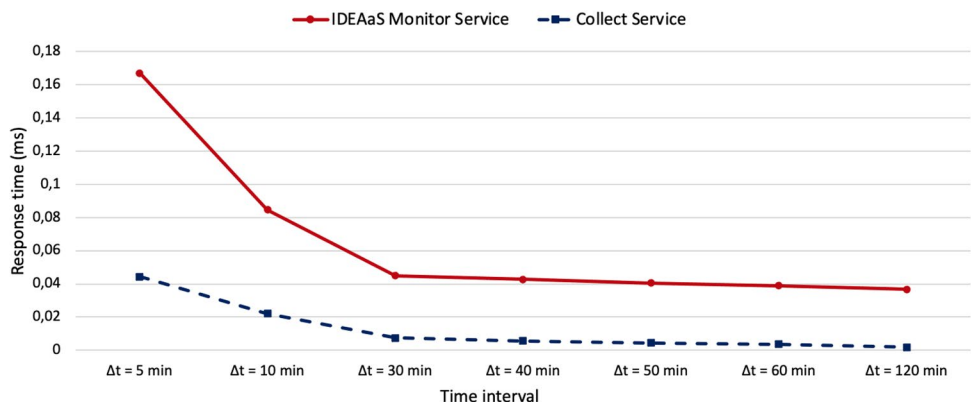
- *Expert user* category comprises users who work daily with the management systems and are familiar with the use of the software.

The participants involved in the experiments are 13 users, taken from industrial production contexts, divided in 2 beginner users, 3 intermediate users and 8 expert users.

Results of usability experiments. The results obtained from the experiments have been analyzed considering three features: (1) the degree of the task difficulty from the examiner’s viewpoint (*estimated difficulty*); (2) the correctness of answers; (3) the degree of the task difficulty as evaluated by the user (*perceived difficulty*). Both the estimated and perceived difficulty answer the questions *How do you rate the requested operation?* and are evaluated using a five-point Likert scale [22] where 1 refers to *very easy* and 5 refers to *very hard*. A comparison has been made between the estimated and the perceived difficulty. Figure 12 shows that the difference between the estimated and the perceived difficulty degree of the questions is low, meaning that the estimated values reflect the reality.

Figure 13 shows the correctness of the answers having a high rate of 89% of correct answers. Among the wrong answers, an exception regards the question Q7 (*What are the problems, if any, that occurred on industrial assets during the production of the valve VSS000799?*), which has been mistaken by most of the users. This question consists of identifying problems during a processing phase in a working center. The high difficulty in this case is due to the fact that, although the question implicitly involves all three perspectives (in fact, to answer the question, the user must start from the product, explore the corresponding production order and involve work centers and browse possible warnings/errors raised during the production), the interpretation of problems is left to the user’s subjective evaluation. These experimental results clarify the outcomes of usability experiments performed in [11]. In fact, in the usability experiments in [11] the most difficult tasks were associated to the involvement of multiple perspectives while using the exploration dashboard. Nevertheless, in this experiment, the difficulty perceived by users is not strictly related to the number of involved perspectives (see Table 3). Therefore, multi-perspective

Fig. 14 Average response time of data collect and monitor services. Response time is normalized with respect to the number of collected measures



dashboard has been able to guide the exploration across all the perspectives, but the data visualisation of the problems (i.e., making them more visible and easily identifiable by the operators) in question Q7 relied too much on the expertise of users.

8.2 Performance Evaluation

In this section we focus on the validation of the advanced monitor service based on data streams, considering that dealing with data streams is a time sensitive task when dealing with real time collected data. The category of monitor services has been added here with respect to [11]. The IDEAAAS approach [21] is engaged for this task, due to its efficacy in reducing the volume of data streams and identifying concept drifts in the monitored streams. Roughly speaking, IDEAAAS applies an incremental clustering algorithm to the multi-dimensional data stream, where each dimension corresponds to an observed parameter, in order to extract a summarized representation of the stream as a set of clusters (referred to as *snapshots*). This extraction is performed every Δt seconds. Concept drift is identified by calculating a distance between snapshots, based on differences in positions of clusters centroids and differences in their radii, obtained through pair-wise comparisons between clusters belonging to the compared snapshots. The Δt parameter can be properly tuned to reduce the volume of data (higher Δt values), thus improving performances, and to augment the promptness in identifying the concept drift (lower Δt values), in a balanced manner. Based on the case study scenario of the project, we monitored a data stream of ~ 140 millions of measures corresponding to 8 parameters about absorbed electrical currents and rotation speed of a group of spindles used for mechanical processing of a part of the valves. Measures are collected every 500 ms (Δt parameter) leading to an acquisition rate of 144 measures per second.

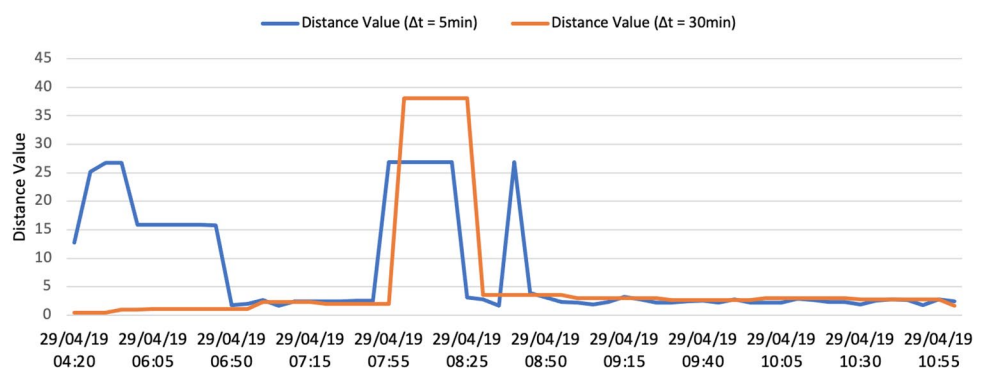
Figure 14 reports the average response time of both collect services and IDEAAAS monitor services for each measure, with respect to the Δt interval. Massive data streams, in fact, have a direct impact on the performance of services used to

collect data from the physical world toward the CPM and on the performance of monitor services, which might hamper the application of the proposed approach. Once data stream volume and processing times have been properly managed, display and dispatch services can be efficiently performed.

We run experiments on a MacBook Pro Retina, with an Intel Core i7-6700HQ processor, at 2.60 GHz, 4 cores, RAM 16GB. In an apparently counter-intuitive way, lower Δt values require more time to process data. This is due to the nature of the IDEAAAS incremental clustering. In fact, for each algorithm iteration (i.e., every Δt seconds), a certain amount of time is required for operations such as opening/closing the connection to the data storage layer and retrieval of the set of clusters previously computed. Therefore, increasing Δt values mean better distribution of this overhead over larger portions of the data stream. On the other hand, higher Δt values decrease the promptness in identifying anomalous events, as the frequency with which the data checking is performed is lower. However, even in the worst case ($\Delta t = 5$ min), ~ 2290 measures per second can be processed by the IDEAAAS monitor service, which is the most critical kind of service, thus demonstrating how the approach is able to face satisfying data acquisition rates for the application domains like the ones in which the approach has been adopted. Moreover, in such applications the deterioration of monitored machines may increase the occurrence of anomalies, and for this reason, the machines should be monitored at a higher frequency. Indeed, as the machines get older, Δt should be set as lower as possible, based on the computational resources, in order to have a near real-time detection of anomalous events.

In Fig. 15 the IDEAAAS monitor service has been applied every 5 min ($\Delta t = 5$ min) and every 30 min ($\Delta t = 30$ min). In this figure, the value of the distance between a current snapshot and the one extracted when the monitored system operates normally is plotted, to detect concept drifts from stable working conditions of the industrial assets. Operators, who supervised the monitored machines, reported anomalous behaviors on April 29th, 2019. Figure 15 shows how IDEAAAS has been able to detect anomalies on the real-data

Fig. 15 Values of the distance between snapshots for anomaly detection applied every 30 min ($\Delta t = 30$ min) and every 5 min ($\Delta t = 5$ min) through the application of IDEAAAS monitor services for massive data streams



stream, in correspondence with the portions of the stream highlighted by operators as critical. However, setting $\Delta t = 30$ min, anomalies before 8 am have not been detected. The reason can be attributed to the presence of more rapid changes in the data stream. Figure 15 shows how by reducing the Δt value, as expected, the promptness in identifying anomalous conditions increases with respect to $\Delta t = 30$ min.

9 Related Work

In this section, a comparison with related work is described taking into account different research directions concerning the design of CPPN infrastructures, data models for CPPN and the adoption of service-oriented architectures in these kinds of systems for Industry 4.0 applications.

CPPN design. The design of CPPN infrastructures has been driven by the collection, organization, analysis and exploration of data for implementing both vertical and horizontal integration of different smart factories participating in the production network [8, 9]. Vertical integration has been extensively studied for Cyber-Physical Systems, defined as “physical and engineering systems that monitor, control, coordinate and integrate physical elements by utilizing computing and communication technologies” [23–25]. The research on Cyber-Physical Production Systems (CPPS) [26], defined as “a physical and engineering composition system, which aggregates resources, equipment, and products by using the interface for the connection and interaction between the physical world and the cyber world”, moved the attention toward data management over the whole production process, to monitor, control, coordinate and integrate resources equipment and products. The above definition indicates that any study on CPPS must focus on the composition of a complex system and that the modularity and interoperability of technology and applications with various levels, layers, and scopes are core issues [27, 28]. The approach described in this paper is more focused on modelling data and services across multiple smart factories, thus further moving the target to horizontal integration in a Cyber Physical Production Network. With respect to recent research on Digital Threads [4], which are more focused on product lifecycle management, the goal here is to balance data management over multiple perspectives, going beyond a single viewpoint, either the product, a single machine or even a set of machines and human resources as done in CPPS approaches. This raises data management issues in CPPN.

Data management issues in CPPN. From an information systems perspective, openness is enabled by standardized interfaces and autonomous data exchange, connecting formerly isolated companies [29]. Significant effort has been invested in creating a coherent standardised information

meta model to enable the exchange of sensitive and valuable data [30, 31]. In this paper, we also addressed the problem of defining proper access policies, i.e., the definition of permissions at the application level similarly to the idea of the role-based access control proposed in [20], which was inspired by the work of creating access rules according to the different actors’ roles [32].

Authors in [33] describe a model-based approach (and a corresponding web-based GUI) to compose CPPS based on predefined building blocks, abstracted as smart services. Smart services are connected to each other and hierarchically organized, instead of assuming a holistic view of CPPS. Ontologies have been also proposed in [34] to face interoperability issues. These papers focus on CPPS within a single production line.

In [25] authors model Digital Twins behind CPPS for product customisation. An information model is proposed to provide data about product, process, plan, plant, resource. In our information model, process and plan are associated with the different concepts of production phases and phase execution. Moreover, resources are properly organized within plants in a hierarchical way. Hierarchies are also used for the product perspective (BoM). The approach in [25] provides five services: production planning, automated execution, real-time monitoring, abnormal situation notification and dynamic response. Our proposal is agnostic about the services to implement in the supply chain, providing a methodology to model services on top of the product, process and industrial assets perspectives.

In [4] the authors focus on the use of models for designing smart products along their lifecycle, being agnostic about the specific technologies, and binding to specific implementations of such features only when needed. A case study for the design of Universal Robots UR3 controller is proposed in [35] within a model-driven integrated development environment, independently from any implementation programming language, operating system, or runtime platform. The focus in the latter papers is on the smart product, whose representation evolves during the product lifecycle, but no data are collected on the design, production or maintenance stages. Similarly, the notion of Digital Thread proposed by commercial solutions such as PTC Windchill is implemented as a sequence of interleaved BoMs (e.g., EBoM, MBoM, as-built, as-maintained) without collecting data on the process and on the work centers during the product lifecycle. Our approach, although mainly focused on the production phase, could be fruitfully extended also during the other stages of the product lifecycle, such as the design or the maintenance ones, enabling a fruitful connection also with the process and industrial assets perspectives.

CPPN and service-oriented architectures Service-oriented architectures (SOA) proved to be an efficient approach for CPS architectures, being able to cope with the

heterogeneity of the industrial systems, to assure interoperability [25, 36, 37]. The authors in [37] propose manufacturing services that are also the trend in the manufacturing industry. As SOA is designed and operates based on data from heterogeneous sources, they enable distributed applications to realize adaptive, flexible and extensible development, integration, management and replacement through loosely coupled connections [23]. Thus SOA is the most widely used design framework in IIoT-based application architectural designs [23]. The adoption of SOA in Industry 4.0 promotes modular architectural design, where loosely coupled and reusable modules, implementing one or few distinct functions, are connected through a simple interface [38]. Design strategies of the services in the portfolio should take into account several challenges, that are very common in service-oriented applications, but have not been properly addressed in existing approaches providing advanced Industry 4.0 solutions, where the focus is on specific needs in the production network (e.g., energy efficiency [12], anomaly detection [13] and predictive maintenance [14] on work centers, production scheduling [15] and process monitoring [16]).

Overall the analysis of the literature indicates that there are a few tentative working on the information modelling to guide the service design of CPPS and service-oriented architectures have not been explicitly investigated in CPPN yet, where access policies on data and services, data sovereignty and smart data mobility are critical aspects. A methodology for guiding the design of data-oriented services is also missing, thus leading to a need for research contributions.

10 Concluding Remarks

In this paper, we presented a methodology for the design of a portfolio of multi-perspective data-oriented services, meant for the horizontal integration of smart factories in cyber physical production networks (CPPN). Services are designed to collect, monitor, dispatch and display data according to three perspectives, namely the production process, the industrial assets and the product. The result is a cyber physical model (CPM), composed of the multi-perspective data model, the portfolio of data-oriented services and a multi-perspective dashboard that enables data access for the actors of the production network upon properly defined access policies. The research described here is part of an industrial research project to move the implementation of Industry 4.0 services at the production network level in the application domain of deep and ultra-deep water valves. Experimental validation in the real case study demonstrated the benefits of the approach both in terms of usability of the data navigation through the services and in terms of service performances in presence of Big Data. Future research

investigation is still required on all the components of the CPM architecture. At the data repository level, methods and techniques for conceptual modelling and design of data lakes (central architectures to store data regardless of formats and sources) or polystores (architectures to elaborate queries on heterogeneous data sources, transparent transformation of just-in-time data and support to multiple query interfaces), as well as a careful choice between these two kinds of architectures to ensure data sovereignty is fundamental. At the service level, the semantic enrichment of the service library to enable reuse and (semi-)automatic composition is required to implement adaptable and resilient CPPN. Finally, the study of CASE tools for supporting the composition of multi-perspective dashboards starting from a set of reusable graphical components on top of display services is another research direction that will be considered in future.

Acknowledgements Not applicable.

Author contributions Not applicable.

Funding Not applicable.

Data availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

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. Firmani D, Leotta F, Mandreoli F, Mecella M (2021) Editorial: big data management in industry 4.0. *Front Big Data* 4:788491. <https://doi.org/10.3389/fdata.2021.788491>
2. Harrison R, Vera D, Ahmad B (2021) A connective framework to support the lifecycle of cyber-physical production systems. *Proc IEEE* 109(4):568–581

3. Nunes D, Silva J, Boavida F (2018) A practical introduction to human-in-the-loop cyber-physical systems. Wiley IEEE Press, New York
4. Margaria T, Schieweck A (2019) The digital thread in industry 4.0. In: Proceedings of international conference on integrated formal methods (IFM), pp 3–24
5. Gould LS (2018) What are digital twins and digital threads? Gardner Business Media's, Cincinnati
6. Gerhard D (2017) In: Biffi S, Lüder A, Gerhard D (eds.) Product lifecycle management challenges of CPPS. Springer, Cham, pp 89–110
7. Hankel M, Rexroth B (2015) The reference architectural model industrie 4.0 (rami 4.0). ZVEI
8. Pennekamp J, Glebke R, Henze M, Meisen T, Quix C, Hai R, Gleim L, Niemietz P, Rudack M, Knape S, Eppele A, Trauth D, Vroomen U, Bergs T, Brecher C, Bührig-Polaczek A, Jarke M, Wehrle K (2019) Towards an infrastructure enabling the internet of production. In: 2019 IEEE international conference on industrial cyber physical systems (ICPS), pp 31–37. <https://doi.org/10.1109/ICPHYS.2019.8780276>
9. Hawkins M (2021) Cyber-physical production networks, internet of things-enabled sustainability, and smart factory performance in industry 4.0-based manufacturing systems. *Econ Manag Financ Mark* 16(2):73–83
10. Hribernik K, Cabri G, Mandreoli F, Mentzas G (2021) Autonomous, context-aware, adaptive digital twins—state of the art and roadmap. *Comput Ind* 133:103508
11. Bagozi A, Bianchini D, Rula A (2021) A multi-perspective model of smart products for designing web-based services on the production chain. In: Zhang W, Zou L, Maamar Z, Chen L (eds.) Web information systems engineering—WISE 2021—22nd international conference on web information systems engineering, WISE 2021, Melbourne, VIC, Australia, October 26–29, 2021, Proceedings, Part II. Lecture notes in computer science, vol 13081. Springer, pp 447–462
12. Matsunaga F, Zytowski V, Valle P, Deschamps F (2022) Optimization of energy efficiency in smart manufacturing through the application of cyber-physical systems and industry 4.0 technologies. *J Energy Resour Technol* 144:1–8
13. Qi L, Yang Y, Zhou X, Rafique W, Ma J (2021) Fast anomaly identification based on multi-aspect data streams for intelligent intrusion detection toward secure industry 4.0. *IEEE Trans Ind Inform.* <https://doi.org/10.1109/TII.2021.3139363>
14. Nordal H, E-Thalji I (2020) Modeling a predictive maintenance management architecture to meet industry 4.0 requirements: a case study. *Syst Eng* 24(1):34–50
15. Jiang Z, Yuan S, Ma J, Wang Q (2021) The evolution of production scheduling from Industry 3.0 through Industry 4.0. *Int J Prod Res* 60:3534–3554
16. Nica E, Stehel V (2021) Internet of Things sensing networks, artificial intelligence-based decision-making algorithms, and real-time process monitoring in sustainable industry 4.0. *Int J Prod Res* 3:35–47
17. Atzeni P, Ceri S, Paraboschi S, Torlone R (1999) Database systems—concepts, languages and architectures. McGraw-Hill Book Company, New York. <http://www.mcgraw-hill.co.uk/atzeni/>
18. Hsieh R-J, Chou J, Ho C-H (2019) Unsupervised online anomaly detection on multivariate sensing time series data for smart manufacturing. In: 2019 IEEE 12th conference on service-oriented computing and applications (SOCA), pp 90–97. <https://doi.org/10.1109/SOCA.2019.00021>
19. Bertino E, Ghinita G, Kamra A (2011) Access control for databases: concepts and systems. *Found Trends Databases* 3(1–2):1–148. <https://doi.org/10.1561/1900000014>
20. Kashmar N, Adda M, Atieh M (2020) From access control models to access control metamodels: a survey. In: Arai K, Bhatia R (eds) Advances in information and communication. Springer, Cham, pp 892–911
21. Bagozi A, Bianchini D, Antonellis VD, Garda M, Marini A (2019) A relevance-based approach for big data exploration. *FGCS* 101:51–69
22. Jebb AT, Ng V, Tay L (2021) A review of key likert scale development advances: 1995–2019. *Front Psychol* 12:1590
23. Lee J, Bagheri B, Kao H-A (2015) A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf Lett* 3:18–23
24. Monostori L, Kádár B, Bauernhansl T, Kondoh S, Kumara S, Reinhart G, Sauer O, Schuh G, Sihn W, Ueda K (2016) Cyber-physical systems in manufacturing. *CIRP Ann* 65(2):621–641
25. Park KT, Lee J, Kim HJ, Noh S (2020) Digital twin-based cyber physical production system architectural framework for personalized production. *Int J Adv Manuf Technol* 106:1787–1810
26. Napoleone A, Macchi M, Pozzetti A (2020) A review on the characteristics of cyber-physical systems for the future smart factories. *J Manuf Syst* 54:305–335
27. Ribeiro L, Björkman M (2018) Transitioning from standard automation solutions to cyber-physical production systems: an assessment of critical conceptual and technical challenges. *IEEE Syst J* 12(4):3816–3827. <https://doi.org/10.1109/JSYST.2017.2771139>
28. Lee J, Ardakani HD, Yang S, Bagheri B (2015) Industrial big data analytics and cyber-physical systems for future maintenance & service innovation. *Procedia CIRP* 38:3–7. In: Proceedings of the 4th international conference on through-life engineering services
29. Brettel M, Friederichsen N, Keller M, Rosenberg M (2014) How virtualization, decentralization and network building change the manufacturing landscape: an industry 4.0 perspective. *Int J Inf Commun Eng* 8(1):37–44
30. Jarke M (2020) Data sovereignty and the internet of production. In: International conference on advanced information systems engineering. Springer, pp 549–558
31. Bader S, Pullmann J, Mader C, Tramp S, Quix C, Müller AW, Akyürek H, Böckmann M, Imbusch BT, Lipp J, et al (2020) The international data spaces information model—an ontology for sovereign exchange of digital content. In: International semantic web conference. Springer, pp 176–192
32. (EFSA), E.F.S.A. (2015) The EFSA data warehouse access rules. Technical report, Wiley Online Library
33. Stock D, Schel D, Bauernhansl T (2020) Middleware-based cyber-physical production system modeling for operators. *Procedia Manuf* 42:111–118
34. Stock D, Schel D (2019) Cyber-physical production system fingerprinting. *Procedia CIRP* 81:393–398
35. Bosselmann S, Frohme M, Kopetzki D, Lybecait M, Naujokat S, Meubauer J, Wirkner D, Zweihoff P, Steffen B (2016) DIME: a programming-less modeling environment for web applications. *ISO-La* 2021:809–832
36. Sacala IS, Pop E, Moisescu MA, Dumitrache I, Caramihai SI, Culita J (2021) Enhancing cps architectures with soa for industry 4.0 enterprise systems. In: 2021 29th mediterranean conference on control and automation (MED). IEEE, pp 71–76
37. Zhang H, Yan Q, Wen Z (2020) Information modeling for cyber-physical production system based on digital twin and automationML. *Int J Adv Manuf Technol* 107(3):1927–1945
38. Catarci T, Firmani D, Leotta F, Mandreoli F, Mecella M, Sapio F (2019) A conceptual architecture and model for smart manufacturing relying on service-based digital twins. In: Bertino E, Chang CK, Chen P, Damiani E, Goul M, Oyama K (eds.) 2019 IEEE international conference on web services, ICWS 2019, Milan, Italy, July 8–13, 2019. IEEE, pp 229–236