Process model as a basis for intellectual control in real time by the processes of the production enterprise

O Novikova¹, A Kotsyubinsky², N Yudina², S Livshits², T Dunaeva²

¹Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
²Kazan State Power Engineering University KSPEU, Kazan, Russia

novikova-olga1970@yandex.ru

Abstract. Modern enterprises are working in a complex and volatile business environment. In order to adapt to this situation, enterprises are trying to realize horizontal and vertical integration of the enterprise processes (managerial and production processes) leading to the creation of a real-time enterprise. In such an enterprise, integral data is used with different time horizons, starting from the millisecond, generated by automatic devices, and ending with daily, generated business applications. However, these integral data are rarely used for operational management of time by enterprise processes of the enterprise. To overcome this limitation, a process model for identifying and assimilating knowledge is proposed that encompasses four steps: (i) analyzing and (re-)designing the enterprise's controlled processes, (ii) developing a data model and enterprise data flow diagrams for automatic devices and business applications, (iii) identification of knowledge based on the discovery of knowledge in working with databases, (iv) continuous monitoring and management of enterprise processes, using complex event processing.

1. Introduction
The difficult, changeable and uncertain business environment of a modern enterprise exerts on the last (enterprise) increased pressure to provide quality services, produce high-quality components, shorten the time for services and production, and reduce costs. Due to this pressure, the business and production processes of creating value added must be flexible, adaptable and constantly monitored. In this connection, vertical and horizontal integration of enterprise processes becomes necessary for survival. For business process management (supply chain management), customer relationship management (CRM) can be available in the market of the system of enterprise applications integration (EAI) [1]. Along with horizontal integration, vertical integration of various levels of the enterprise is a necessary condition for continuous monitoring of enterprise processes and creation of a real-time enterprise (RTE - Real-Time Enterprise) [2].

The enterprises of production, gas and electric transport - can be divided into three levels of manufacturing execution systems (MES): the level of enterprise management, the level of production management and the production level, as shown in figure 1.
At the enterprise management level, it is business processes that allow to achieve strategic targets, so business processes must be developed, launched and analyzed in accordance with the four steps of the BPM (BPM - Business Process Management): design, configuration, execution, diagnostics (monitoring and optimization). During the configuration of the business process at the enterprise management level, are laid transactional planned performance values (i.e., meanings "as expected" - TO-BE) based on the strategic objectives of the enterprise and with a measurement period of weeks and months.

Production processes are used to achieve the goals and indicators set at the enterprise management level and use automatic devices of the production level. A huge amount of data (so-called sensory data) is generated by these devices in real time (i.e., seconds and millisecond intervals). In addition, operators define data related to automatic devices, such as predetermined causes of equipment failure, instructions defining the order of supply of parts at the time the operation is started, and so on. In general, these values (i.e., the value "as is" - AS-IS) reflect the real productivity at the production level. MES solutions and process applications allow the creation of computerized and automated vertical integration. However, the problem of interaction between the enterprise management level and the production level remains unresolved [3]: implementation of multi-loop control within the enterprise and through all levels from the level of enterprise management to production level. Insufficient vertical integration prevents the accumulation of knowledge of the enterprise, the creation of training cycles [4] and the use of accumulated knowledge. First, data from various levels of MES are not integrated and, therefore, cannot be used to gain new knowledge. Secondly, if new knowledge has been received, it is not used in the continuous management of enterprise processes. As a consequence, the concepts of RTE: feel-and-react and learn-and-adapt are little used in enterprise processes.
2. Current status

Since the mid-1990s, several reference enterprise integration architectures (CIMOSA, PERA, ARIS and GRAI / GIM) have become available to guide the design and implementation of the integrated enterprise. However, these reference architectures differed in theoretical basis, understanding of the enterprise, approaches to modeling and goals. To overcome these problems, a generalized enterprise reference architecture and methodology (GERAM-Generalized Enterprise Reference Architecture and Methodology) was developed [5]. The mentioned reference architectures contributed to the development of GERAM, which was later standardized as ISO 15704 - requirements for the enterprise reference architecture and methodologies. It is still largely taken into account in the new standards (for example, ISO 19439: 2006 "Enterprise integration Framework for enterprise modeling").

The GERAM scheme provided for four groups of aspects of the enterprise architecture, called Views - the types of models ("functions", "data", "resources", "organization"), purpose, implementation and "physical representations" (hardware), but also the possibility identify additional aspects. It included a description of all aspects or some part of them in each of the seven or eight phases of the architecture and operation of the enterprise, and also included the specification of the architecture model at three levels - generalized, level of partial models and specific models. GERAM defines common adjacent enterprise concepts recommended for application in enterprise engineering and implementation of integration projects. Such concepts can be classified as:

1) concepts that are human-centered.
2) concepts aimed at describing the business process of the enterprise;
3) concepts focused on the description of technology to support the business process, involved in the work process and strengthening the enterprise engineering (modeling and application of the model).

Figure 2. Levels of manufacturing systems of the enterprise.
The company's reference architectures do not specify how to implement them in terms of specific technologies, i.e. are platform-independent. In addition to the company's reference architectures, some software vendors have developed solutions to overcome bottlenecks of vertical integration between MES-levels, for example, MES-HYDRA. But data exchange between MES-levels continues to be done manually or, at best, semi-automatically due to inflexible and closed interfaces [3].

An agent-oriented production monitoring and control system (PMC) was developed and based on the JADE (JADE framework) [6] PMC Provis.Agent, which integrates various information systems and machine control devices, determines the use of information between different systems (for example, for visualization). The NIIP-SMART architecture provides horizontal and vertical integration and interoperability, using workflow, enterprise rules, agents and STEP.

In the service-oriented architecture (SOA) used to integrate enterprise applications, business applications offer their functionality as loosely coupled services (services) that are combined into complex work flows. Due to the loosely coupled structure, the implemented IT architecture is flexible and adaptable. European projects SIRENA and SOCRADES [7] use the SOA paradigm for seamless integration of heterogeneous production-level resources with business applications at the enterprise management level. Unlike the prevailing "request-response" communication approach, an enterprise must react to current events, which requires the implementation of the "publication-subscription" mechanism. SOA and event handling are complementary concepts that provide modularity, weak connectivity and flexibility.

The enterprise data model is a necessary part of the EI for correlating the TO-BE and AS-IS criteria of the various MES levels. IEC 62264-2 [8] describes models and terminology, which allows to realize interaction between the level of enterprise management and the level of production management. For system integration, the structure of the MES database with respect to resources, system plans, system statuses and system configuration was analyzed [9]. The production data model is defined for represent strategic objectives, the organization structure and the enterprise's behavior. This data model consists of the classes "strategy", "object", "process", "resource", "token" and "flow". The "token" class is a physical flow of material (for example, a workpiece, an account). The "flow" class is a reference to the "token" and "process" classes. For example, The Data Reference Model (DRM) is one of the five major reference models in Enterprise Architecture, which gave rise to a comprehensive description and planning of the development of the business device of the enterprise. It is necessary for any leader, at least in order to know that his organization is a whole, to maintain an expedient order of its arrangement, and then to proceed to its current development or transformation taking into account all important recommendations. It is a structure whose main purpose is to ensure the exchange of information and its reuse through the federal government through a standard description and detection of common data and the promotion of common data management methods. In other words, it is a model for describing information.

Not explicit and context-specific knowledge is embedded in the processes of the enterprise by employees of the enterprise, i.e., "know-who" in the form of “know-what”, “know- why” and "know-how". This knowledge is specified by the employees of the enterprise and can be expressed using the data (attributes) of the enterprise process (i.e., TO-BE values). This data as a process feedback is constantly updated during each process execution (i.e., AS-IS values). TO-BE and AS-IS values from different MES levels are integrated for the operational management of enterprise processes [4]. These integrated values are stored in the relational database as historical data that are periodically used in the processes of calculating key performance indicators (KPIs), overall equipment effectiveness (OEE) and the like.

The process of decision-making related to the management of enterprise processes covering different MES levels is a complex task, depending on the quantity and quality of the information. A framework for the organization and application of knowledge in decision-making in the field of production was developed [10] in which the decision-making process was supported by knowledge obtained using data mining algorithms.
The European project K-NET (sub-project of the FInES project cluster - Future Internet Enterprise Systems) presented a conceptual approach to improving, monitoring and reusing knowledge in a networked enterprise. In addition, the modeling and enterprise integration framework based on Knowledge Discovery in Databases (KDD) with the extension of CIMOSA views. In most enterprises, explicit knowledge is codified as rules in managed systems based on rules (RBS - Rule-Based Systems).

RBS Drools Expert, for example, often lacks time and cause-effect relationships between events, therefore, studies were conducted on the use of explicit knowledge to create patterns of events used in the complex event processing (CEP) program. The concept of unified event management architecture for the simplest and complex events for monitoring and control of production processes was developed [11]. The architecture is positioned at the enterprise management level and integrates data in real time from production level, but neglects the integration of transaction data from the enterprise management level. Later, the architecture for the expandable event-driven production system was refined. This architecture is built on a MES solution with a tight integration of enterprise-level management, and uses the CEP to manage events generated by the production level. However, the approach presented does not take into account the knowledge identification required to define the event templates.

The Insights system is an open source framework for operational monitoring and analysis of manufacturing enterprises [12], which contains components for the delivery, collection and analysis of data. Data delivery from various devices is achieved through the MTConnect standard and the MTConnect data bus. Data is stored in high-speed databases. To monitor enterprise processes, data analysis is performed online, using the services of Esper Tech or Drools Fusion [13] CEP. Saved data is used to calculate various indicators.

3. Methodology
For digital integration of the enterprise, an IT framework was developed (figure 3), which uses tracking objects along with RBS.

![Figure 3. IT-framework for digital integration of the enterprise.](image-url)
To take into account temporal and cause-effect relationships between events launched at different levels of the enterprise, it is necessary to supplement RBS with modern CEP. This framework is based on the developed methodology for the identification and use of knowledge for the operational control of enterprise processes, including the components of the analysis of enterprise processes, enterprise data models and diagrams of data flows, knowledge identification, and assimilation of knowledge.

3.1. Overview of the IT framework for digital enterprise integration

The architecture of the IT framework digital enterprise integration is shown in the figure 2. The flow of process data and control data between the various components of this IT framework is shown in figure 4.

![Figure 4. Process data flow and control data IT enterprise integration digital enterprise.](image)

The IT framework is based on current standards (ISO15704, IEC 62264), technologies and paradigms (SOA paradigm) and includes various IT systems. Enterprise processes are defined earlier by the created workflow templates (step 1 in figure 4) and are supported with the required planned performance or process criteria (TO-BE values) such as the ERP system (figure 3 and Step 2 in figure 4). Business applications or their critical functionality in a specific context (access to planned performance values) are available as services (services) within the SOA.

For continuous monitoring of enterprise processes, as one of the goals of the IT framework, a publishing-subscription mechanism used in event-driven architectures (EDA) is used. It is a software architecture template that allows the creation, definition, consumption and response to events.

This architectural template can be used to develop and implement applications and systems that transfer events among loosely coupled software components and services. An event-driven system
typically contains sources of events (or agents) and event consumers (or sinks). The drains are responsible for the response as soon as the event has occurred. The reaction can be fully or not completely created by runoff. For example, the runoff can only be responsible for filtering, converting and delivering an event to another component, or it can create its own response to this event. The first category of waste can be based on traditional components, such as middleware for messaging, and the second category of wastewater (which forms its own reaction in the process of operation) may require a more suitable transaction execution platform. Event-driven architecture corresponds to a service-oriented architecture (SOA), because services can be activated by triggers triggered by incoming events.

Service-oriented architecture, driven by events, develops SOA and EDA architectures to provide a deeper and more reliable service level by using previously unknown cause-effect relationships to form a new event model. This new business intelligence template leads to further automation of the processing, adding to the enterprise previously unachievable performance by adding valuable information to the recognized pattern of activity.

The EI layer subscribes to events generated by production resources (automatic devices) at the production level. To collect data from these production resources and forward them to all subscribers, a three-level physical integration architecture is used (step 3 in figure 4). The received real-time data from the production level contain the actual performance (AS-IS values). The enterprise management level plan values together with the actual production level performance values are combined according to the enterprise data model and stored as a historical data in the relational database (step 6 in figure 4).

The IT-framework for digital integration of the enterprise is expanded by the operational monitoring and control of the enterprise processes using the tracking objects, which are objects such as orders, products and resources that are the entities of the corresponding workflow in the WMS (step 2 in figure 4). These objects change during execution of the process with the values of different MES levels (step 4 in figure 4). Changes in the state of the tracking objects are analyzed in a constant mode (online) using RBS (for example, Drools Expert). The use of RBS implies the absence of time and cause-effect relationships between events. It should be noted that only a few WMSs support the collection and interpretation of real-time data, hence, it becomes necessary to use CEP (for example, Esper Tech) instead of RBS. SERE-continuously analyzes the tracking objects and directs the control data to the required levels (step 7 in figure 3).

In addition to monitoring enterprise processes using CEP, historical data are periodically used to calculate various indicators, including key performance indicators. However, historical data are rarely used to acquire new knowledge, and in the future this knowledge is not used for operational monitoring and management of enterprise processes. To overcome these shortcomings, a methodology for identifying and assimilating knowledge for operational monitoring and management of enterprise processes has been developed.

3.2. Methodology of identification and assimilation of knowledge

The model of the process of implementing the digital integration of the enterprise is shown in figure 5.
Figure 5. Model of the process of realizing the operational control of enterprise processes.

This process model consists of four stages, each of which is unique to a particular enterprise and they need not necessarily be performed sequentially; moreover, individual stages of the process can be performed from time to time to improve the processes of the enterprise.

Before the implementation of the IT infrastructure of enterprise digital integration, it is necessary to analyze and redesign the enterprise processes, similar to the life cycle of BPM (Phase I in figure 5). The company data model based on industry standards (IEC 62264) is extended by technical models (DIN EN 61512-2) and relates the corresponding AS-IS and TO-BE values of different MES levels (step II in figure 5). Data link diagrams (DFDs) should be created to identify the relationships between business applications, resources, and associated events generated at different MES levels.

The knowledge (pressure, temperature) generated in the data of the enterprise's processes generated before and during the execution of these processes are TO-BE and AS-IS values. These process data are mapped to the mentioned enterprise data model and stored in the relational database as historical data. Subsequently, the standalone KDD process can be run on historical data to determine new knowledge (phase III in figure 5). Reuse of time-consuming queries to the database, data is not optimal for the on-line monitoring of enterprise processes. Event streams (TO-BE and AS-IS values) created during the execution of the process must be analyzed and processed operatively (online) using CEP (step IV in figure 5). Explicit knowledge must be codified as event templates and their corresponding rules, and these event patterns should be used to detect complex events. Along with the detection of complex events, event processing should be used for operational control of enterprise processes.

3.2.1. Step 1. Analysis and (re-) design of enterprise processes

The analysis and redesign of enterprise processes is part of the enterprise reference architecture and business process reengineering (BPR) [14]. BPR can be described in four main steps: (i) identifying critical enterprise processes, (ii) reviewing, updating and analyzing enterprise processes (AS-IS analysis), (iii) (re-)designing enterprise processes based on AS IS-analysis and (iv) implementation (re-engineered enterprise processes).

Business processes are very diverse, but there are more definite requirements that all of them must meet. One can single out the following principles of organization of business processes formed during the reengineering:

1. Integration of business processes. The most characteristic property of the redesigned processes is the lack of assembly pipelines as a way of coordinating the work of personnel with relatively simple labor functions. When performing complex labor functions, a different organization
of work is required. In practice, of course, it is not always possible to reduce all the stages of the process to the work performed by one person. In this case, a team is created that is responsible for this process. There may be malfunctions and errors, but the losses will be much less than with the traditional organization of work.

2. Horizontal compression of business processes. Comparative evaluations performed by companies that reengineered show that the transition from the traditional organization of work to the implementation of the process by one person allows to reduce the number of personnel and accelerate the process approximately 10-fold. Reduces the number of errors and eliminates the need to keep professionals to eliminate these errors. By reducing the number of employees and a clear distribution of responsibilities between them, manageability is improved.

3. Decentralization of responsibility (vertical compression of business processes). Performers make independent decisions in cases in which they have traditionally had to turn to management.

4. The logic of the implementation of business processes. Linear execution of work is replaced by a logical order (that is, often the work is carried out in parallel). This saves time, which was spent on interconnection of works on different sites.

5. Diversification of business processes. There are different versions of the execution process. A traditional process, oriented to the production of mass production, must be performed equally for all inputs, leading to concerted outputs. Traditional processes are usually very complex, as they are very detailed and in many respects are calculated for exceptions and special cases.

6. Development of various versions of business processes in an ever-changing market is necessary for processes to have different options depending on situations, entrances and market conditions. New processes that have different versions begin with a check step, which determines which version of the process is most appropriate for the current situation. Therefore, new processes, unlike traditional ones, are simpler and more understandable, since each option is oriented only on one, the corresponding situation.

7. Rationalization of management impact. It is a matter of reducing the number of inspections and reducing the degree of managerial influence that does not lead directly to the acquisition of material values. Therefore, the task of reengineering is to implement them only to the extent that it is economically feasible.

8. Culture of the solution of the problem. It is assumed that the harmonization is minimized, since they too have no material value. The task of reengineering is to minimize reconciliations in the course of execution of the process by reducing external contacts.

9. Rationalization of "company-customer" ties. Improvement of the company's organizational structure should create conditions under which the authorized manager provides a single channel of communication.

10. Preservation of positive moments of centralization of management. In practice, this is achieved by improving the information support of the divisional management organization. Modern IT enables the company's units to operate autonomously, while retaining the ability to use centralized data. Thus, the company can eliminate the bureaucratic regional structures necessary to serve a geographically dispersed clientele, and simultaneously improve the quality of service.

11. Rationalization of horizontal links. Creation of linear functional units. Work is carried out in the place where it is most appropriate. Previously, in companies, the work was organized according to the "thematic" principle in the relevant divisions: the settlement department, the transport department, the supply department, etc., so if the accounting department needed pencils, then he applied to the supply department with the application. This department found the manufacturer, negotiated the price, placed the order, inspected the goods, paid for it, and transferred it to the settlement department. This process is long and uneconomical. An analysis conducted in one of the companies showed that the cost of buying a battery for $ 3. amounted to 100 dollars. During reengineering, horizontal management connections between departments are most often created. This eliminates unnecessary integration.
12. Authorized leader. This principle is applied in those cases when the steps of the process are either complex or distributed in such a way that their integration by the forces of a small team is impossible. The authorized manager is a buffer between the complex process and the customer. The manager in the relationship with the customer is responsible for the entire process. To play this role, the manager should be able to answer customer questions and solve his problems. The content of the task makes it necessary to provide the manager with access to all the information systems used in this process, as well as to his executors.

Understanding the processes of the enterprise and their integration within the organizational structure of the enterprise is crucial for the implementation of the strategy of operational (online) management of enterprise processes. Thus, the analysis and redesign of processes involves a corporate organizational structure at the same time as the process-oriented organization itself.

Activities and functions of the enterprise process are performed by various resources (IT systems) responsible for ensuring the operation of the organizational units of the enterprise, which are modeled by organizational charts. Functions and activities of the enterprise process accept several types of inputs, this can be data (printed documents) and intangible resources, as implicit knowledge of the participants of the enterprise. Each function of the enterprise process can be associated with the organizational unit with various inputs and outputs. Functions are combined into enterprise processes using logical connectors "and", "or", "exclusive or".

Several modeling languages and methodologies are used to model enterprise processes, both on the basis of an event-driven process chain (EPC), and models and notations of business processes (BPMN - Business Process Model and Notation). In addition to these modeling languages that focus on material inputs and outputs (data and documents), the descriptive knowledge management languages (KMDL - Knowledge Management Description Language) can be used to describe "knowledge" processes with the creation and use of knowledge in enterprise processes.

3.2.2. Step 2. Enterprise data model and data flow diagram
Today's business applications and automatic devices are complex and require a lot of data to determine the processes of the enterprise, in real time automatic devices produce a huge amount of data, such as feedback, product position, alerts. For operational monitoring and control of these processes, it is necessary to analyze business applications, automatic devices and their corresponding processes to identify important parameters. In this regard, the modeling of enterprise data is an important step and affects the quality of information needed to perform the processes, achieve EI, and improve operational monitoring and management of enterprise processes.

In addition to the static structure of the enterprise data model, DFDs identifies the interdependencies between process systems and production resources, either alone or in combination. In addition, DFDs identifies the flow of data describing the dynamic behavior of enterprise processes. Data is generated and managed by several resources at runtime. DFDs can be used to show the relationships between different information systems and resources. As the number of systems and resources increases, DFDs are organized in the DFDs hierarchy. Large-scale DFDs can display, for example, a trading floor and its resources, while small-scale DFDs is created for a particular trading room sub-process.

3.2.3. Step 3. Identification of knowledge
In an environment where information and knowledge - these entities that do not receive adequate objectification beyond their own power - become the main production factor, the problem loses its former economic character and to some extent becomes a sociological and even psychological problem. This is confirmed, in particular, by the fact that it identifies and develops problems.

Knowledge can be defined from different points of view. In the current context, the following definition is used: "Data is raw numbers and facts, information is processed data, and knowledge is verified information".
From the historical data, new knowledge is used, used in online monitoring and process management of the enterprise. Knowledge identification should be performed using a KDD process defined as "the process of mapping low-level data to other forms that can be more compact or abstract, or useful." It is the process of finding useful knowledge in "raw" data. KDD includes questions: preparing data, selecting informative features, cleaning data, applying Data Mining (DM) methods, post-processing data, and interpreting the results. Of course, the "core" of this whole process is the DM methods, which allow us to discover knowledge. These knowledge can be rules that describe the relationships between data properties (decision trees), frequently encountered patterns (associative rules), and classification results (neural networks), data clustering, etc. The KDD process consists of the following steps:

1) Prepare the original data set. This stage consists in creating a data set, including from various sources, the selection of the training sample, and so on. To do this, there must be developed tools for accessing various data sources. It is desirable to have support for working with data warehouses and the existence of a semantic layer that allows you to use business concepts to prepare the initial data, not technical terms. The system includes Deductor Warehouse - a multidimensional data warehouse, focused on solving problems of consolidating information from heterogeneous sources and quickly retrieving the data set of interest.

2) Preprocessing of data. In order to effectively apply the methods of Data Mining, you should pay attention to the issues of data preprocessing. Data may contain omissions, noise, abnormal values, etc. In addition, the data may be redundant, insufficient, etc. In some problems it is required to supplement the data with some a priori information. It is naive to assume that if you submit data to the input of the system in its existing form, then at the output you will get useful knowledge. The data must be qualitative and correct in terms of the DM method used. Therefore, the first stage of KDD is the preprocessing of data. Moreover, sometimes the dimension of the original space can be very large, and then it is desirable to apply special algorithms for reducing the dimension. This is both the selection of significant features, and the mapping of data into a space of smaller dimensions. It also includes a Deductor that contains a large set of mechanisms for preprocessing and cleaning data: filling in blanks, editing anomalies, removing noise, smoothing, filtering and many others with the ability to combine preprocessing methods.

3) Transformation, normalization of data. This step is necessary to bring information to a form suitable for later analysis. For what it is necessary to do such operations as casting, quantization, and reduction to a "sliding window" and so on. In addition, some analysis methods that require that the source data be in some specific form. Neural networks, for example, work only with numerical data, and they should be normalized. Here Deductor includes a large set of data transformation mechanisms that allow all the preparatory work for the subsequent analysis. In addition, the system contains a wide range of normalization mechanisms for all types of data: numeric, string, date / time and logical.

4) Data Mining. At this step, various algorithms are used to find knowledge. These are neural networks, decision trees, clustering algorithms, associations, etc. The package includes algorithms that implement popular and effective Data Mining methods: neural networks, decision trees, Kohonen self-organizing maps, associative rules and so on.

5) Post processing the data. Interpretation of results and application of acquired knowledge in business applications. The results of any processing can be displayed using a large set of visualization mechanisms: OLAP, tables, diagrams, trees and many others. For some mechanisms, specialized visualizers are provided, which make it easy to interpret the results. Results can be exported for later processing with other applications. This makes it possible to effectively use the acquired knowledge or models on other data.

The KDD process is depicted in Fig. 6. The KDD process receives historical data, and at the output we get templates with some definite quality, known as a measure of interest. This measure of interest can be objectively measured by statistical methods, based on the properties found by patterns, can be based on subjective indicators that are derived from user beliefs or expectations. A template is an
abstract representation of a subset of data and should be evaluated by domain experts to identify knowledge.

![Diagram of Data Discovery Process in Databases (KDD Process)](image)

**Figure 6.** Data discovery process in databases (KDD process)

Understanding the domain is necessary for the successful implementation of the KDD process. This should be done as described in the analysis and redesign of the process, enterprise data models and DFDs (steps I and II in Fig. 5). Based on the main activities of manufacturing and gas distribution companies (production (transportation), maintenance, quality management and production preparation), it is necessary to determine the objectives of the KDD process. Depending on the goals of the KDD process, the specific data (target data) from which historical templates are searched will be extracted from the historical data.

For various reasons, historical data may be inaccurate, influencing the identified knowledge. During data collection, the latter come from operators through consoles, analog equipment and digital meters with programming logic controllers (PLCs - Programming Logic Controllers) (Fig. 3). In the AS-IS IT framework, the values become available for the Enterprise Integration layer through Object Linking and Embedding for OLE for Process Control (Fig. 3 and Step 3 in Fig. 4). The collected data consists of noise, inaccuracies, missed values, which complicates the search for templates. This can be caused by limitations of measuring instruments, operator misprints or errors in PLCs logic. To overcome these inaccuracies, it is necessary to study and understand the subject domain (stage I in Fig. 5). Understanding the domain along with the enterprise data model will help determine appropriate statistical methods for removing noise, a strategy to fill in the missing values and remove duplicate data. Data collection should be supplemented by verification of collected data from production resources in coordination with operators. In general, cleaning and pre-treatment should be performed on the selected data for further processing.

To achieve the above goals, the KDD process requires a subset of pre-processed data. Data conversion causes a reduction in the number of parameters in the target data or the presentation of the target data in a more general or acceptable format. To reduce the number of parameters, filtering and data convolution should be used [15]. Understanding production processes, operations and constraints will help in the transformation process supported by the enterprise data model and DFDs.
Data mining is a specific sub-process of the KDD process and is based on proven techniques of machine learning, pattern recognition, statistics, artificial intelligence, knowledge collection, data visualization and high-performance computing. Data mining consists of three stages: selecting a method for searching for data, determining the appropriate algorithms for searching for data, and using these algorithms to find a template.

Data mining methods should be selected in accordance with the objectives of the KDD process, as they determine the type of knowledge that will be extracted, i.e., description of the concept, classification, association, clustering and forecasting. The knowledge necessary to support decision-making processes must be identified. To determine new knowledge, it is necessary to use methods of classification and regression. Classification is a way of categorizing new instances of data into one of several predefined classes. It consists of two steps: constructing a model (classification) based on the analysis of tuples of a database (training set) described by attributes, and using this constructed (classification) model to classify new instances of data. Regression is a function that maps data to a predictable real variable.

KDD includes multi-disciplinary events. This includes storing and accessing data, scaling algorithms to massive data sets, and interpreting the results. The process of data cleaning and data access included in the data warehouse. Artificial intelligence supports KDD, revealing the empirical laws of experimentation and observation. Patterns recognized in the data must be valid with new data and have some degree of certainty. These templates are considered new knowledge.

In accordance with the previously selected methods of data extraction (for example, classification), data mining algorithms should be chosen to search for templates; Decision trees, decision rules, programmable inductive logic can be used to define rules [10]. Selected methods and data mining algorithms are used to detect templates.

The detected patterns can be large enough or require a subset of the detected patterns. To improve the quality of the detected patterns, the measure of interest should include subjective and objective approaches [15]. The detected template can be obtained using structured interviews of experts in the subject area. If necessary, all or individual steps of the KDD process must be repeated in order to obtain more accurate knowledge. If significant integrated data is not available for the KDD process, structured interviews with domain experts should be conducted to identify the primary knowledge. Later, the collected historical data can be used in the KDD process to increase and expand the knowledge base.

3.2.4. Step 4. Knowledge-Based Operational Control of Enterprise

Repeated and time-consuming requests to the database (for example, the use of online analytical processing requests, OLAP - OnLine Analytical Processing) result in non-operative (offline) monitoring of enterprise processes. Consequently, real-time data must be processed in real time-online (or almost real-time) to monitor enterprise processes, using pre-identified knowledge, tracking objects, and CEP, as shown in Fig. 3 and 4.

Tracking objects represent the entities (products, orders, resources) of a particular enterprise process that are monitored by the workflow of the WMS process route (step 1 in Fig. 3) and the associated performance targets (TO-BE values). During the execution of enterprise processes, AS-IS production-level values are available in the OPC servers and simultaneously directed to the EI layer using the publish-subscribe mechanism (step 3 in Fig. 4). These AS-IS values are integrated with the corresponding TO-BE values from the ERP system obtained using the request-response mechanism.

The EI layer manages the integrated process data simultaneously in several ways. First, tracking objects are updated with the corresponding integrated process data (step 4 in Fig. 4) and therefore contain the latest status information about the entities of the particular enterprise in its process. Secondly, integrated process data is delivered by subscription to all subscribed clients with a graphical user interface (GUI) for operational monitoring of enterprise processes (Fig. 3). Finally, integrated process data is stored in the relational database as historical data for offline analysis (Fig. 3 and Step 6 in Fig. 4). Tracking objects are constantly analyzed at the level of production management and the
result is used for the on-line monitoring of enterprise processes using the CEP to improve the main lines of business activity (step 5 in Fig. 4). The CEP is responsible for scheduling the control data to the production resources (step 7 in Fig. 4), for updating the tracking objects with the control data.

The event is characterized by its source (a certain automation device), type, attribute (data) and time stamp or time interval and, in addition, the receiver (operator, enterprise manager). Events are initiated at various levels of the MES during the execution of enterprise processes and form an event cloud [12]. Events can be classified as simple and complex (composite), described by their levels of abstraction. Simple events are run at different levels of the MES and have no abstraction. Consequently, a simple event does not provide enough information for the on-line monitoring of enterprise processes [11]. For example, a simple event is triggered each time a document is printed on the printer.

Conversely, a complex event with a higher abstraction can be described by an event model based on simple events [11]. For example, a complex event is triggered whenever the total number of printed pages of documents exceeds the corresponding article of the department's monthly budget. Higher abstraction events can be obtained from composite events, as shown in Fig. 7. A composite event can be created by combining basic events using a specific set of constructors, such as disjunction, join, sequence, and so on. An event flow is a linearly ordered sequence of events that is ordered by arrival time or is limited to a certain time interval. Event streams consist of simple events that indicate the creation or update of tracking objects at runtime.

![Figure 7. Abstraction of a hierarchical event.](image-url)

An event template is a template that contains event patterns, relational operators, and variables. The relationships between various (simple and composite) events can be basic, (temporal and spatial) [12], logical, (temporal and causal) [11, 16]. Simple events are created at a certain time period and have timestamps associated with them. The participants of the enterprise from the level of enterprise management and the level of production management are interested in aggregated events for the on-line monitoring and management of products, orders or resources. To make event aggregation available, it is necessary to use temporal event patterns with time slots to support slippery time boundaries [11, 16], as shown in Fig. 8. Temporal event patterns include overlap, match, content and
event patterns before or after [12]; and also defined the operators (concatenation, sequence) associated with templates of temporal events.

**Figure 8.** Processing of time events using sliding time boundaries.

CEP performs operations on complex events, including reading, creating, converting, or abstracting [16]. The main task of CEP - continuously analyzing the flow of events (tracking objects), monitor the processes of the enterprise. Tracking objects contain the current status information of entities of enterprise processes. When updating new incoming integral data, the tracking objects are analyzed by CEP (step 5 in Fig. 4). Inside the CEP, capable of analyzing logical, temporal and cause-effect event templates, event templates written in the Event Processing Language (EPL) are used. Rules of event templates (reactive rules) determine how the CER reacts to the occurrence of a specific event template. Incoming events are analyzed using event template rules and the necessary control data. Events with higher abstractions are created and added to the already existing event cloud for further processing.

For effective monitoring and management of enterprise processes, it is necessary to identify and characterize events. Previously, certain knowledge is assimilated by creating event templates and rules of event templates with EPL programming. Structured interviews with domain experts are used to refine and improve event templates and corresponding rules of event templates. Event templates and rules for event templates are stored in a centralized database. Participants in the enterprise or decision-makers do not need all-events. Event receivers (or event users) can be configured in accordance with the roles of the enterprise participants (manager, plant manager) and the appropriate privileges (defined in the LDAP server protocol, LDAP - Lightweight Directory Access Protocol). The event configuration provides the necessary functionality for configuring event templates.

There are two realizations of how CEP influences or controls enterprise processes. First. CEP uses the interfaces and services provided by the EI layer (Fig. 3), for automatic dispatch control commands. The second. Before manipulating the processes of the enterprise, CEP exposes the envisaged solution as a proposal for customers with a GUI. The enterprise participant accepts or rejects the offer. Human interaction is used in cases when the participants of the enterprise must take responsibility. Access to the above functionality of dispatching control data depends on the role of members of a particular enterprise and their respective privileges.

SOA software vendors will extend their proposed CEP features, enabling them to support an event-based architecture as an alternative or complement to SOA. More and more manufacturers will implement CEP support in their BAM tools for BPM environments in order to be able to react almost
instantly to changing business conditions. Corporate bus manufacturers will invest in CEP to offer user-friendly means of combining events, correlating and visualizing them for their publishing and subscription environments. But the CEP market does not realize its full potential until the producer community prepares a coherent platform for interoperability that uses open SOA standards. In this respect, the creation of the Event Processing Technical Society (EPTS) community, a group of manufacturers and other interested parties at the end of 2006, is very encouraging in order to popularize CEP applications, clarify the relevant terminology and define a reference platform to support interoperability.

However, EPTS representatives made it clear that they do not intend to become a standards-setting organization, although they do not deny the possibility of subsequently collaborating with groups dealing with standards. EPTS has not yet submitted open proposals for the reference platform, nor has it accessed other community groups, such as the Organization for Advancement of Structured Information Standards (OASIS), which is developing a standard for CEP. The WS-Notifications standard supports event-driven SOA interaction templates and provides notification exchange.

With the advent of open standards in the next few years, the CEP market will begin to develop very quickly. This will also be facilitated by the creation of freely available alternatives and the acquisition by the leading producers of SOA, ESB, BI and EII of the most promising companies specializing in CEP. By the end of this decade, the situation in the CEP market will radically change, and enterprises will be able to implement standards-based CEP infrastructure, calculated for different applications and not depending on the manufacturer.

4. Practice
The mentioned IT framework and the corresponding model of the process of digital integration of the enterprise and achievement of operational control of the enterprise processes are developed and adapted to manage the innovative activity of the gas transportation and gas distribution enterprise. Expected results are applicable to grid companies in the energy sector.

Acknowledgment
The research carried out with the financial support of the grant from the Program Competitiveness Enhancement of Peter the Great St.Petersburg Polytechnic University, Project 5-100-2020

References
[1] D. Linthicum, Enterprise Application Integration, Addison-WesleyLongman, Amsterdam, 2000.
[2] S. Karnouskos, D. Guinard, D. Savio, P. Spiess, O. Baecker, V. Trifa, and L. de Souza, “Towards the Real-Time Enterprise: Service-basedIntegration of Heterogeneous SOA-ready Industrial Devices with Enterprise Applications,” Proc. 13th IFAC Symposium on Information Control Problems in Manufacturing (INCOM ’09), June 2009, pp. 2127-2132.
[3] H. Panetto and A. Molina, “Enterprise Integration and Interoperability in Manufacturing Systems: Trends and Issues,” Computers in Industry, vol. 59, no. 7, Sept. 2008, pp. 641-646.
[4] M. Grauer, D. Metz, S. Karadgi, W. Schäfer, and J. W. Reichwald, “Towards an IT-Framework for Digital Enterprise Integration”, Proc. 6th International Conference on Digital Enterprise Technology (DEIT 2009), AISC, vol. 66, Springer, Berlin, Dec. 2009, pp. 1467-1482.
[5] P. Bernus and L. Nemes, “The Contribution of the Generalised Enterprise Reference Architecture to Consensus in the Area of Enterprise Integration,” Proc. of ICEIMT97, K. Kosanke and J. Nell,Eds., Springer, 1997, pp. 175-189.
[6] O. Sauer and G. Sutschet, “Agent-Based Control”, Computing & Control Eng. J., vol. 17, no. 3, June 2006, pp. 32-37.
[7] L. M. S. de Souza, P. Spiess, D. Guinard, M. Köhler, S. Karnouskos, and D. Savio, “SOCRADES: A Web Service Based Shop Floor Integration Infrastructure,” Internet of Things, C. Floerkemeier, M. Langheinrich, E. Fleisch, F. Mattern, and S. Sarma, Eds., Springer, Berlin, LNCS 4952, 2008, pp. 50-67.
[8] IEC 62264, Enterprise-control system integration.
[9] B. Zhou, S. Wang, and L. Xi, “Data Model Design for Manufacturing Execution System,” J. of Manufacturing Technology Management, vol. 16, no. 8, 2005, pp 909-935.
[10] A. Kusiak, “Data Mining: Manufacturing and Service Applications,” International J. of Production Research, vol. 44, nos. 18-19, Sept./Oct. 2006, pp. 4175-4191.
[11] K. Walzer, J. Rode, D. Wünsch, and M. Groch, “Event-Driven Manufacturing: Unified management of Primitive and Complex Events for Manufacturing and Control”, IEEE International Workshop on Factory Communication Systems, 2008.
[12] A. Vijayaraghavan, “MTConnect for Realtime Monitoring and Analysis of Manufacturing Enterprises,” Dec. 2009; http://www.systeminsights.com/, 02.04.2016.
[13] Drools Fusion, http://www.drools.org/, 02.04.2016.
[14] M. Hammer and J. Champy, Reengineering the Corporation: A Manifesto for Business Revolution, Harper Business, 1994.
[15] A. A. Freitas, “A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery,” Advances in Evolutionary Computing: Theory and Applications, Springer-Verlag, New York, 2003, pp. 819-845.
[16] K. Walzer, A. Schill, and A. Löser, “Temporal Constraints for Rule Based Event Processing”, Proc. ACM first Ph.D. workshop in CIKM, Nov. 2007, pp. 93-100.
[17] George L. Bagiev, Olga A. Pogrebova, Evgenii A. Konnikov “Marketing platform of organization of system-spatial complex «PSE» (Production — Science — Education)” Published in: "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS). 2017. DOI: 10.1109/ITMQIS.2017.8085864