Chapter XXI
Business Process Intelligence

M. Castellanos
Hewlett-Packard Laboratories, USA

A. K. Alves de Medeiros
Eindhoven University of Technology, The Netherlands

J. Mendling
Queensland University of Technology, Australia

B. Weber
University of Innsbruck, Australia

A. J. M. M. Weijters
Eindhoven University of Technology, The Netherlands

ABSTRACT

Business Process Intelligence (BPI) is an emerging area that is getting increasingly popular for enterprises. The need to improve business process efficiency, to react quickly to changes and to meet regulatory compliance is among the main drivers for BPI. BPI refers to the application of Business Intelligence techniques to business processes and comprises a large range of application areas spanning from process monitoring and analysis to process discovery, conformance checking, prediction and optimization. This chapter provides an introductory overview of BPI and its application areas and delivers an understanding of how to apply BPI in one's own setting. In particular, it shows how process mining techniques such as process discovery and conformance checking can be used to support process modeling and process redesign. In addition, it illustrates how processes can be improved and optimized over time using analytics for explanation, prediction, optimization and what-if-analysis. Throughout the chapter, a strong emphasis is given to describe tools that use these techniques to support BPI. Finally, major challenges for applying BPI in practice and future trends are discussed.

1. INTRODUCTION

Business Process Intelligence (BPI) refers to the application of Business Intelligence (BI) techniques to business processes (Grigori et al., 2004). In this context, BI refers to technologies, applications, and practices for the collection, integration, analysis, and presentation of business information and also sometimes to the information itself. The purpose of BI is to support better business decision making (Power 2007). The data source for BI is a so-called data warehouse, i.e., a special data base where an organization stores important historical data. Most of the time the data is collected from different information systems as used in an organization. Data analysis and data mining can be performed using this data. The goal is to translate the data to useful business information that can support the decision making process of the organization. If the data warehouse also contains information about the processes within an organization it is called a process data warehouse (Casati at al., 2007) and can be used as source for BPI analysis.

BPI is an emerging area, that is quickly gaining interest due to the increasing pressure companies are facing to improve the efficiency of their business processes and to quickly react to market changes in order to be competitive in this highly dynamic Internet era. In addition, the need to meet regulatory compliance has recently strengthened this trend (e.g., Sarbanes-Oxley (Sarbanes-Oxley Act 2002). The large number of buzzwords like Business Activity Monitoring (BAM), Business Operations Management (BOM), Business Process Intelligence (BPI), Process Mining, and Business Operations Intelligence (BOI) is a good indication of the interest of vendors to monitor and analyze business activities to gain insight into the operation of their business and ultimately on their effect on the business goals. In the past the focus of workflow tools has been mostly on process modeling and automation. However, today most vendors of business process management (BPM) suites have extended their portfolio with BPI functionality (e.g., IBM, SAP, Tibco, Oracle, Pallas Athena, Lombardi, webMethods).

Process-aware information systems (PAIS) such as WFM, ERP, SCM and CRM systems are recording business events occurring during process execution in event logs (Dumas et al., 2005). Typically, event logs contain information about start and completion of activities and the resources that executed them. In many cases relevant data (like the values of data fields linked to tasks) is recorded too. Sometimes, there is no, or only a very primitive process model available. However, in many situations it is possible to gather information about the processes as they take place. For instance, in many hospitals, information about the different treatments of a patient are registered (date, time, treatment, medical staff) for reasons like financial administration. This kind of information in combination with appropriate (mining) techniques can also be used to get more insight in the health care process.

BPI exploits this process information by providing the means for analyzing it to give companies a better understanding of how their business processes are actually executed. It supplies support in the discovery of malfunctions and bottlenecks and helps identifying their causes. Therefore, BPI often triggers process improvement or reengineering efforts. BPI not only serves as a tool for improving business processes performance, but also fosters changes by facilitating decision-making. In addition, BPI is used to monitor the alignment of operational business processes with strategic business goals and to give the visibility that regulatory compliance requires. Furthermore, BPI is not restricted to the analysis of historical data, but can also be used to optimize future efforts (e.g., through predicting future problems). To provide for the above, BPI comprises several application areas, which are detailed in the following.

Process Analysis: refers to the analysis of past and sometimes even current process executions. Process analysis can lead to different kinds of models: explanatory, prognosis and decision models. On the explanatory category, process analysis is helpful for business analysts to find correlations between different workflow aspects and performance metrics (e.g., unproportionally high cost occurs whenever goods are shipped to a particular country). In addition, analysts are supported in identifying the causes of malfunctions or bottlenecks (e.g., waiting times at first level customer support result in costly service level agreement violations). These explanatory models help in identifying opportunities for process **optimization** (Castellanos et al., 2005a**)** whose aim is the generation of decision models that optimize some aspect(s) of the operation of a process. For example, optimizations may include changes in the sizing of resource pools or resource assignment rules. In addition to the analysis of historical data to understand past process behavior (explanatory models) and the identification of opportunities for process optimization (decision models), BPI also aims at building prognosis (a.k.a prediction) models by **predicting** critical situations and undesired behavior (e.g., exceptional situations or delays on a running process instance that bear the risk of an SLA violation). This enables companies to either prevent the occurrence (or at least minimize the damage) of these critical situations by taking corrective actions proactively, or prepare a plan for handle them after they occur.

Process Discovery: refers to the analysis of business events recorded in event logs to discover process, control, data, organizational, and social structures (Aalst et al., 2007b). Like process analysis, process discovery allows users to gain insight into their operations and can be the first step when implementing business processes with a workflow tool. While process analysis primarily focuses on the analysis of business processes in respect to performance metrics, process discovery aims at constructing process models from historical data. This information can be used together with performance metrics to identify malfunctions or bottlenecks.

Process Monitoring: refers to the monitoring of running process instances (e.g., their progress, bottlenecks and times spent in each activity) and their analysis results (e.g., percentage of instances not completing successfully) to inform users about unusual or undesired situations (i.e., alerts) (Grigori et al., 2004). Process dashboards or reporting features provide information about process performance characteristics like average cycle time or number of processes not meeting a Service Level Agreement (SLA). For example, notifications are automatically sent to individuals if critical events can be detected, enabling them to take immediate action (e.g., process instances with long cycle time raise the likelihood of SLA violation).

Conformance Checking: while process monitoring analyses running process instances, conformance checking can be applied to analyze whether a log conforms to a process model and to identify undesired behavior a-posteriori (Rozinat and Aalst, 2008). For instance, (Aalst and de Medeiros, 2005) describes the application of conformance checking to detect security violations (e.g., violations of the separation-of-duty principle) in event logs.

Although the application of business process intelligence can provide companies with substantial benefits and case studies like (Aalst et al., 2007b) clearly show that these techniques have gained a level of maturity that makes them applicable to real-world business processes, the practical use of BPI is still limited. Companies are facing several challenges when applying BPI, and these need to be solved in a practical way before BPI will become mainstream.

In this chapter we first describe the process mining techniques that aid the modeling of business process (cf. Section 2). Then an overview on process optimization including analytics for explanation and prediction, business impact analysis and resource allocation is presented (cf. Section 3). Major challenges for applying BPI in practice are discussed briefly (cf. Section 4).

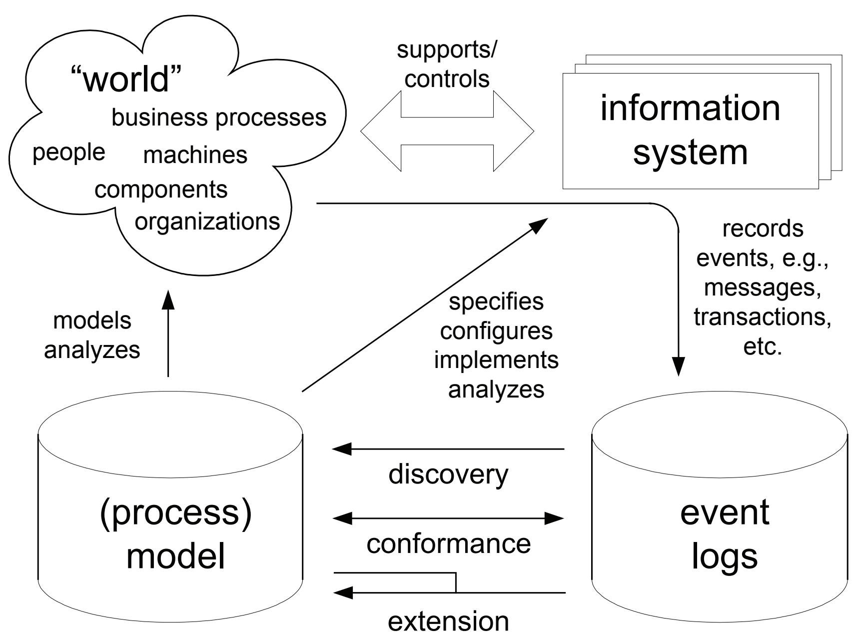
Finally, a conclusion and an outlook close the chapter (cf. Section 5).

2. PROCESS MINING FOR MODELING

Process mining targets the discovery of information based on an event log. As explained in Section 1, nowadays organizations usually are able to register in some log what events have being carried out during the execution of their business processes. Such logs are the starting point of process mining techniques. These techniques assume event logs to minimally contain data about (i) which events belong to the execution of a same process instance and (ii) the ordering of execution for these events. Additionally, some process mining techniques also require data about the event timestamps, performers and data fields. In general, the more information a log contains, the more different process mining techniques can be used.

The analysis provided by current process mining techniques can be seen as from three types: *discovery*, *conformance* and *extensions* (cf. Figure 1). The techniques that focus on *discovery* mine information based on data in an event log only. This means that these techniques do not assume the existence of pre-defined models to describe aspect of processes in the organization. Examples of such techniques are *control-flow mining* algorithms (Aalst et al., 2004, Cook et al., 2004, de Medeiros, 2006, de Medeiros et al., 2007b, Dongen, 2007, Dongen and van der Aalst, 2005, Greco et al., 2006, 2007, Günther and van der Aalst, 2007, Herbst and Karagiannis, 2004, Pinter and Golani, 2004, Schimm, 2004, Wen et al., 2007) that extract a process model based on the dependency relations that can be inferred among the tasks in the log. Another example are social network mining algorithms (Aalst et al., 2005b, Ly et al., 2005) that discover the relations between the performers of certain tasks, like a graph that shows who is handing over work to whom. The algorithms for *conformance* verify if

Figure 1. Perspectives on process mining



the executions registered in logs follow *prescribed* behaviors and/or rules. Therefore, besides a log, such algorithms also receive as input a model that captures the desired property or behavior to check. Examples are the algorithms that assess how much the behavior expressed in a log matches the behavior defined in a model and points out the differences (Rozinat and Aalst, 2008), and algorithms used for auditing of logs (in this case, the model is the property to be verified) (Aalst et al., 2005a). The *extension* algorithms enhance existing models based on information discovered in event logs. Examples include algorithms that automatically discover business rules for the choices in a given model (Rozinat and van der Aalst, 2006).

The remainder of this section describes the *main process mining techniques that aid the modeling of business process*. Other than providing details of given techniques, we focus on *giving an overview of the possibilities and indicating pointers* where the reader can find more details. The descriptions are based on a running example inspired in a real-life situation (cf. Section 2.1). Relevant discovery, conformance and extension techniques are respectively introduced in Sections 2.2, 2.3 and 2.4. All the illustrated techniques (and many other techniques) are implemented in the process mining tool ProM, an open-source tool available at *www.processmining.org*.

2.1 Running Example

The running example is inspired on processes of the Dutch rental housing organizations. These organizations rent houses at cheaper prices than in the private sector. They have many processes, like registration, personal information update, complaints handling, etc. In this section we will focus on the *process to handle requests (or complaints) for house repairs*. The process starts when a tenant contacts the company to file a complaint. If the complaint indeed involves a repair in the house, a ticket is created and an appointment is made to inspect the house such that the actual problem can be detected/confirmed. Additionally, the inspector estimates how much time will be needed to fix the problem. Easy fixes are usually performed together with the inspection. More complicated fixes require a new appointment and can be performed by an internal or external team. When the repair has been performed, the client is informed and the ticket number is communicated to the financial administration so that they can take care of the payment to the appropriate institutions. The process completes whenever the payment is in place.

The next sections show how a designer could use process mining techniques to get more insight about how complaints are actually handled. The results are based on a simulated event log for the running example.

2.2 Discovery Techniques

When (re-)designing business processes models, two aspects are particularly relevant: the *control-flow structure* (i.e., which tasks precede/follow others and how frequently) and the *organizational structure* (i.e., which teams/roles perform which tasks and how they cooperate). These two structures are important because they define the core elements necessary to execute business processes.

Current *discovery* process mining techniques mine information that helps in modeling both the control-flow and the organizational structure. For instance, have a look at figures 2, 3 and 4. Figure 2 shows the EPC (Event-Driven Process Chain) (Keller et al., 1992) model for an event log of our running example. The mined EPC is shown on the right pane. Note that this discovered model, which captures the control-flow structure of the process in our running example, is a very good starting point for the designer because it objectively summarizes the *real behavior* during process execution (as registered in a log). For instance, by looking at the selected part of the mined model (cf. left pane), it is possible to see that, according to the registered behavior, (i) the tasks `SendTicketToFinAdmin` and `ReadyInformClient` can be executed in any order after the task `Repair Ready` has being completed, and (ii) the two branches containing the respective tasks `TicketReady` and `InformClientWrongPlace` are alternative ones. In a similar way, figures 3 and 4 illustrate how the designer can respectively get feedback about how people are cooperating in the organization and possible roles for tasks. The works in (Aalst

*Figure 2. Screenshot containing the result of the **Multi-phase Macro** plug-in for a log of the running example*

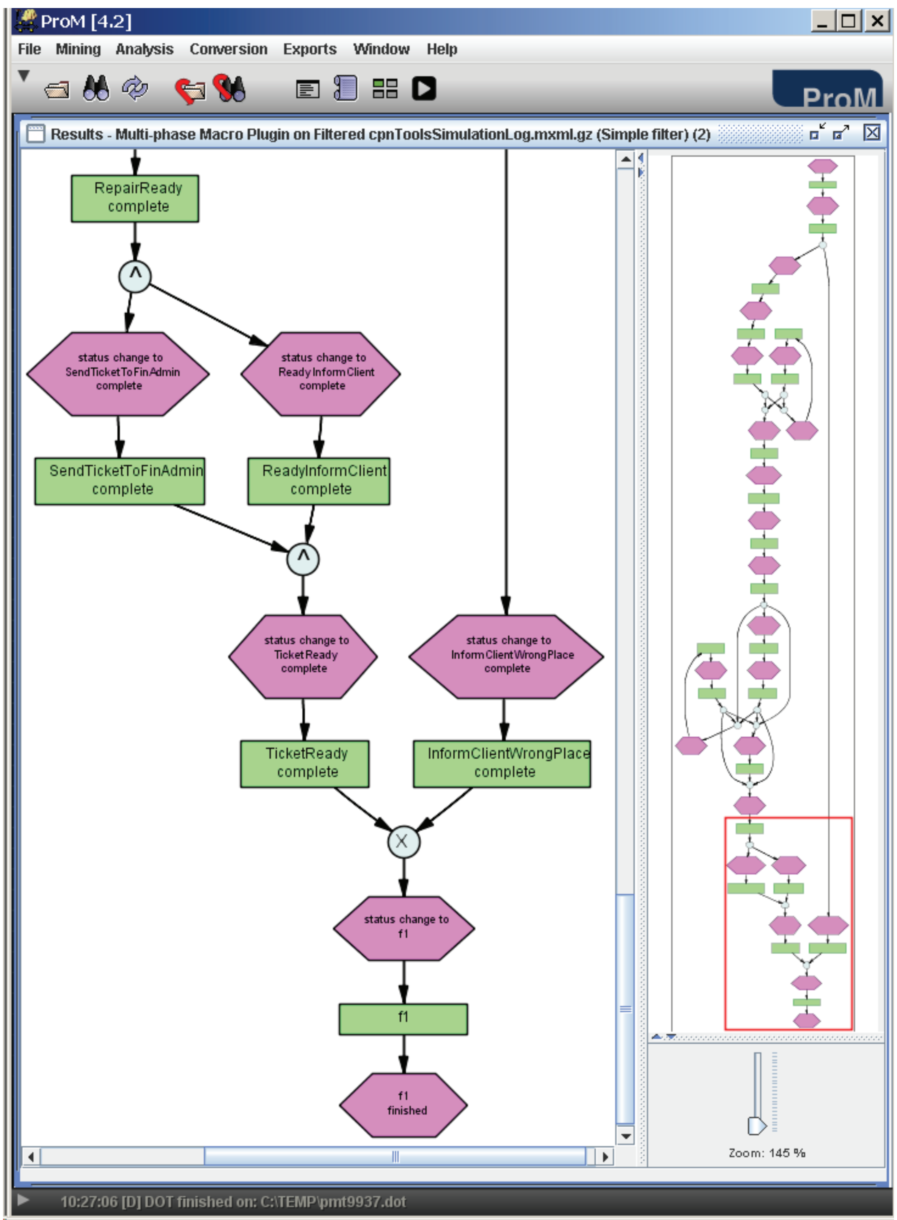


Figure 3. Screenshot showing the Analyze Social Network plug-in in action. The illustrated social network presents the mined handover of work for a log of the running example. Every circle represents a performer (or user). The area of a circle (or "Vertex Size") indicates how often users execute tasks, the stretching direction of a node (horizontally or vertically) indicates the relation between its in (receiving work) and out (handing over) degrees. For instance, note that some performers (like "System") collaborate with many others, while others (like "Monica" and "Dian") collaborate with fewer ones. Furthermore, some users never hand over work (like the external repair companies "DoIt" and "FixIt"). Graphs like this one give the designer insights about how people work together and who are the team players in the organization. This may be useful feedback when defining the scheduling for the distribution of work.

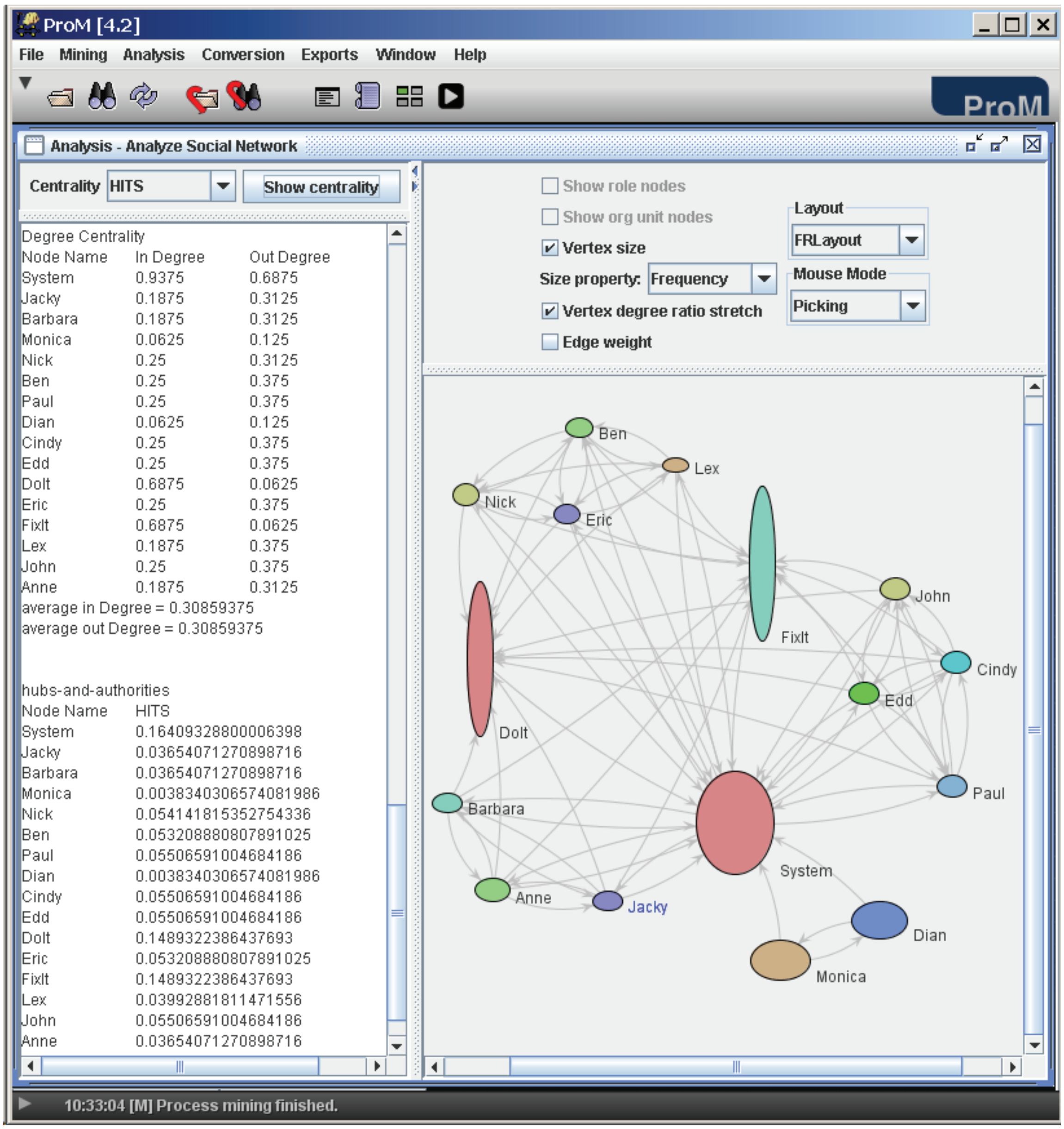
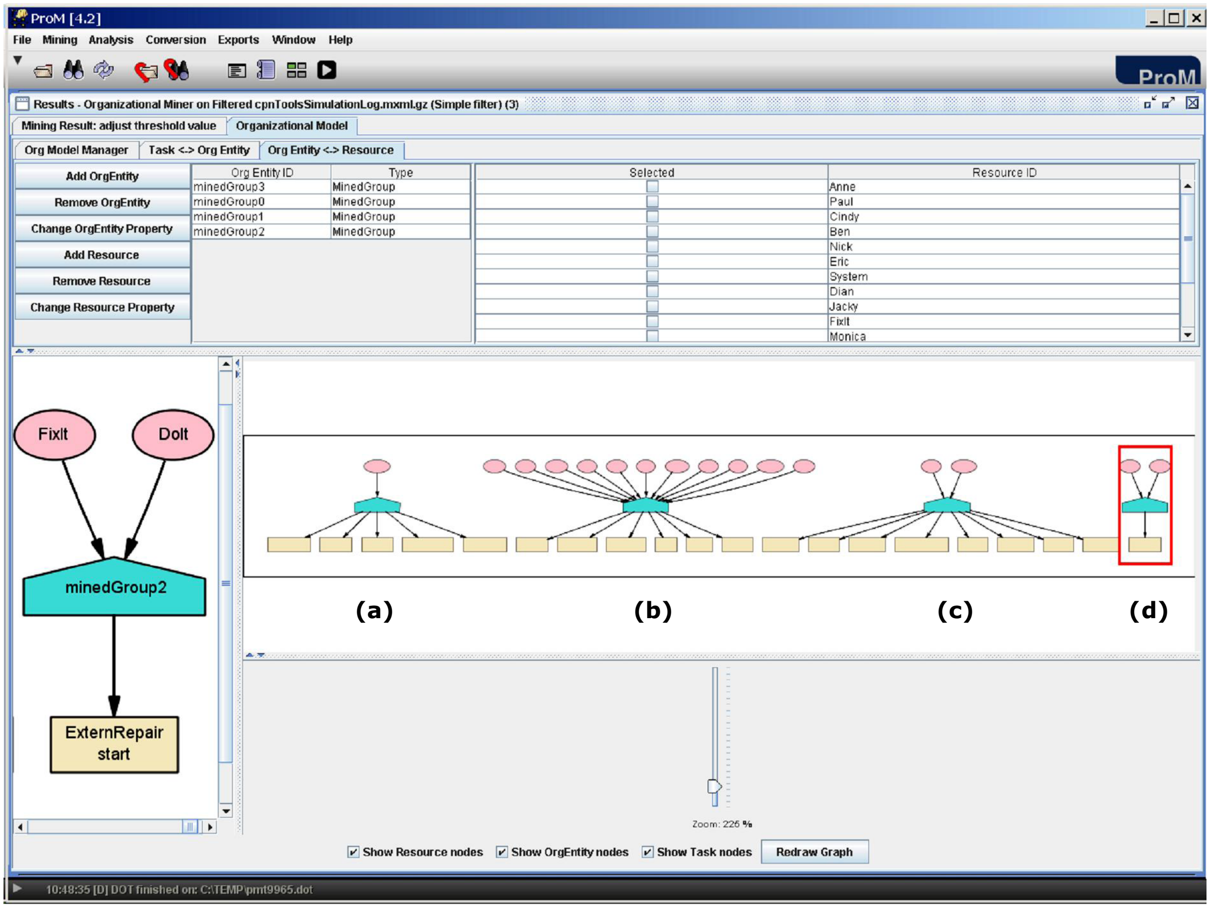


Figure 4. Screenshot with the result of the plug-in ***Organizational Miner****. In this case, four roles have been automatically discovered containing the following users: (a) "System", (c) "Dian" and "Monica" (i.e. front office employees), (d) "FixIt" and "DoIt"' (i.e. external repair companies), and (b) the remaining users (i.e. the technical employees). Note that each mined role relates users to tasks. For instance, the role in the left-pane refers to the task* `ExternRepair`*.*



et al., 2004, Cook et al., 2004, de Medeiros, 2006, de Medeiros et al., 2007b, Dongen, 2007, Dongen and van der Aalst, 2005, Greco et al., 2006, 2007, Günther and van der Aalst, 2007, Herbst and Karagiannis, 2004, Pinter and Golani, 2004, Schimm, 2004, Wen et al., 2007) provide details of different existing control-flow mining techniques. For organizational related techniques, the reader is referred to (Aalst et al., 2005b, Song and van der Aalst, 2007).

2.3 Conformance Techniques

The conformance techniques compare the behavior expressed in models with the one registered in logs. They are useful to check compliance in companies. In a nutshell, conformance techniques focus on two aspects: (i) assessing how much a log matches a model and highlighting the points of discrepancy, like the Conformance Checker (Rozinat and Aalst, 2008), and (ii) verifying if certain properties hold in a log, like the LTL Checker[1] (Aalst et al., 2005a).

Conformance Checker is helpful when comparing prescribed behavior with enacted one. The technique basically measure how much a log fits a model. If the behavior in the log can be fully replayed in the model, the fitness is 100%. The more problems are encountered during the log replay, the lower the fitness value. For instance, Figure 5 shows the results of comparing a model to a log in the context of our running example. The result shows that most of the behavior in the model (about 94%, as indicated by the "Fitness" metric) matches what has been actually executed. However, there are points of mismatch because the model defines that the task InformClientSurvey should happen between the tasks ArrangeSuvey and Survey, but this task has not been executed a single time in the log (see value "0" in the input/output arcs of this task).

LTL Checker is mainly used for auditing purposes. For instance, in the setting of our running example one could inspect if the rule that *immediate fixes that could not be solved should be handled by an internal team again before being sent to an external team* has been followed. Figure 6 shows the result of verifying this property for a given log. In this case, the traces in the log have been pre-processed to keep only the tasks `ImmediateRepair`, `InternRepair`, and `ExternRepair`. The top window in this figure shows the configured property and the bottom one, the returned results. As can be seen, an unfixed immediate repair has been directly sent to an external team in 3.8% (38 out of 1000) of the cases.

2.4 Extension Techniques

Extension techniques enhance existing models by making information that is hidden in the log explicit. A good example is the process mining technique that mines the business rules applying to points of choice in process models (Rozinat and van der Aalst, 2006). Figure 7 shows the results of using this technique to a model of our running example. Note that three rules have been mined to determine if a repair should be executed during the inspection, later by an internal team or by an external one. Actually, these *de facto* business rules could be incorporated in the deployed process model. Another example is the

*Figure 5. Screenshot showing a result of the **Conformance Checker** plug-in. In this case, the "Model" perspective is being illustrated. The task* **`InformClientSurvey`** *should happen between the tasks* **`ArrangeSuvey`** *and* **`Survey`**, *but this task has not been executed a single time in the log.*

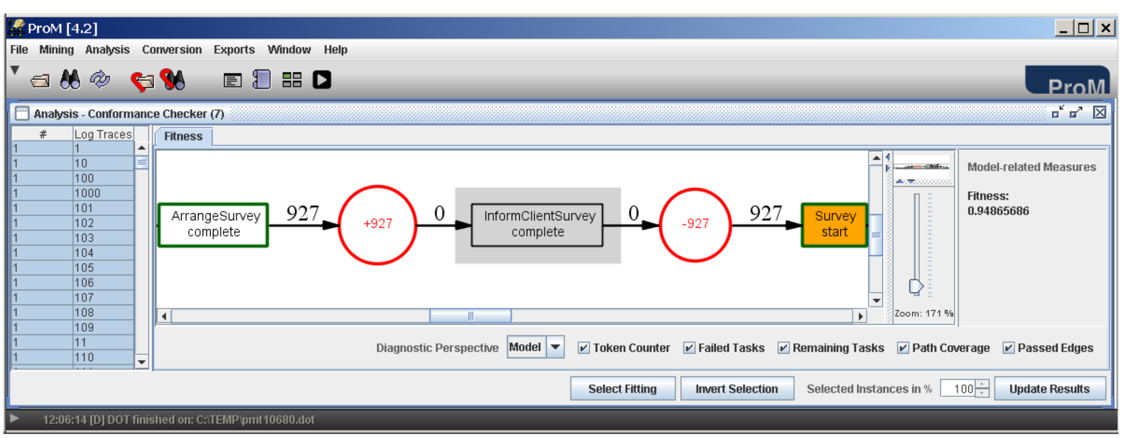
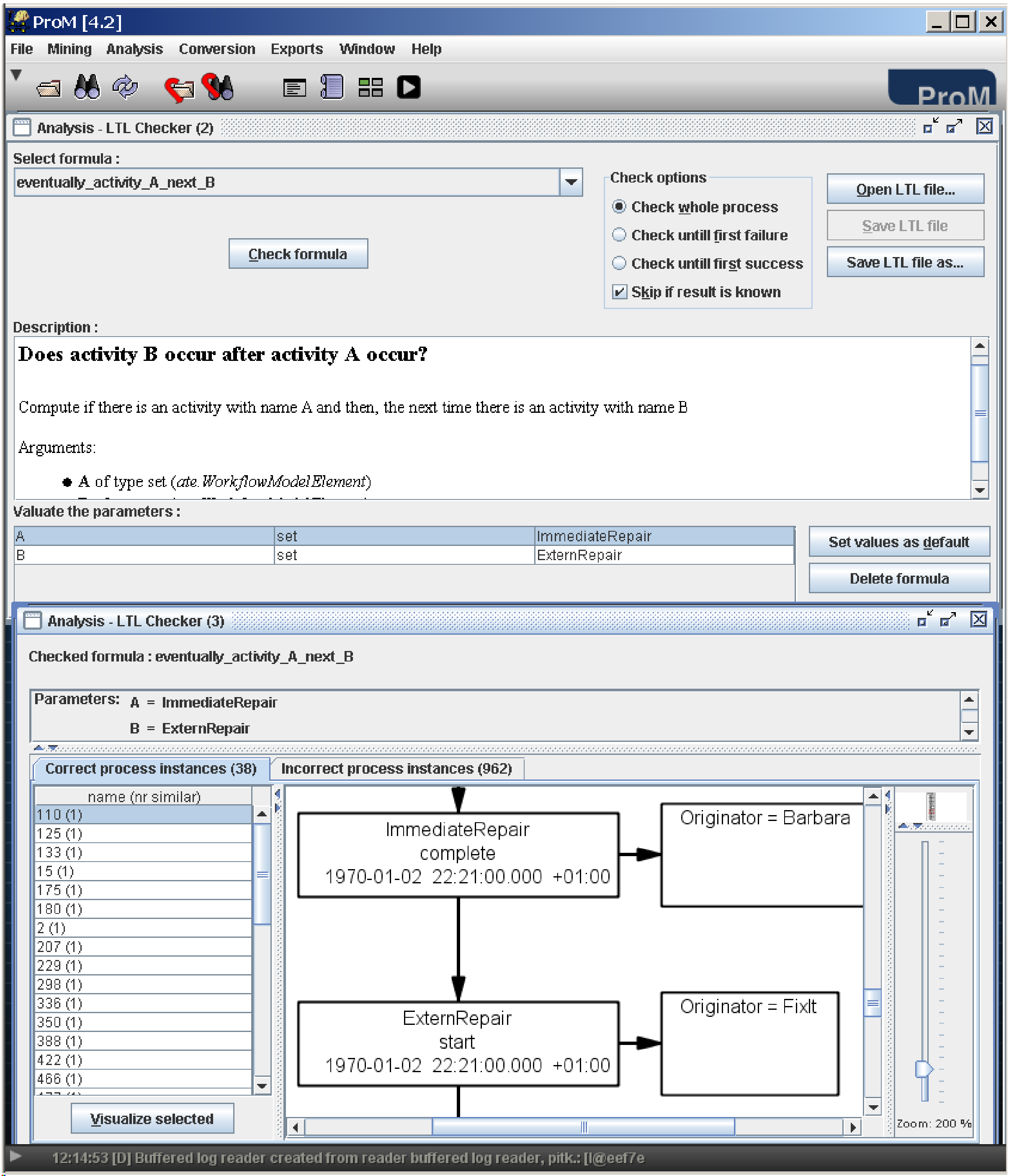


*Figure 6. Screenshot illustrating both interfaces of the **LTL Checker** plug-in. The top window shows the main user interface. The bottom one contains the results of checking the formula "eventually_activity_A_next_B" for a log of the running example*



process mining technique that detects bottlenecks in processes (Hornix, 2007). This technique automatically mines upper bounds for different key performance indicators (like waiting times, execution times etc) of a process. It does so by taking into account both the timestamps of tasks in a log and the overall structure of the process model given as input. The results of the analysis are directly indicated in the process model. Note that this feedback is very important when, for instance, trying to optimize throughput times of processes based by re-design. Figure 8 illustrates a bottleneck point for a model of our running example.

Many of the techniques described in this section have been used to perform the case studies in (Aalst et al., 2007a, de Medeiros, 2006, Dongen, 2007, Rozinat and Aalst, 2008, Rozinat et al., 2007), confirming that process mining is indeed a useful tool to get feedback about how systems are actually being used.

Figure 7. Screenshot with the result of the ***Decision Point Analysis*** *plug-in. In this case, after the survey (or inspection) has been completed, a choice is made on how to proceed. The mined rules indicate which fields values determine this choice.*

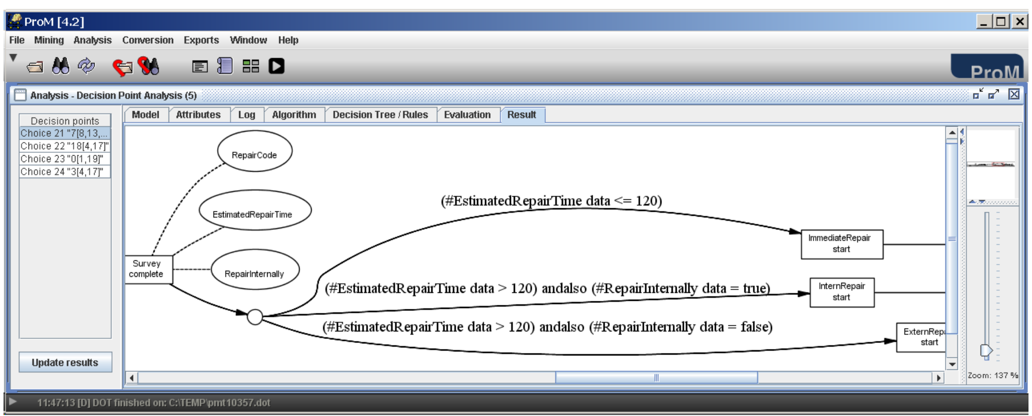
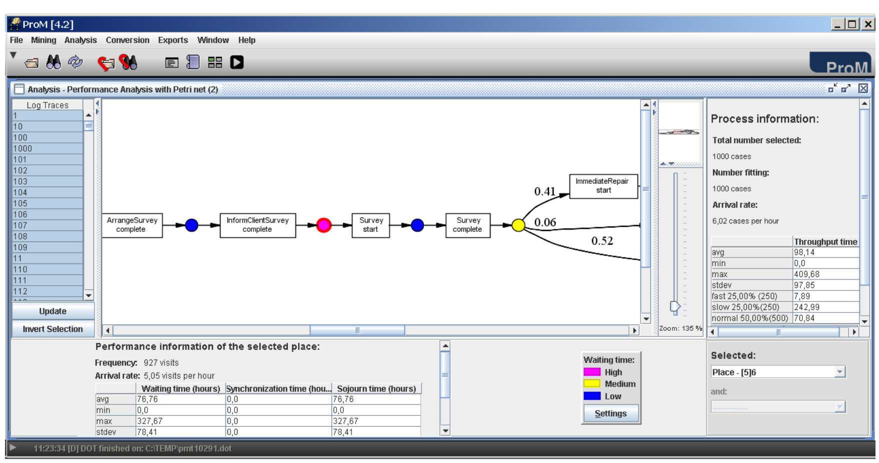


Figure 8. Screenshot of the ***Performance Analysis with Petri Nets*** *plug-in in action. Note that the places are colored based on waiting time thresholds (cf. "Waiting time:" box at the right of the bottom pane). In this case, there is a high waiting time between the tasks* **`InformClientSurvey` and `Survey`.**



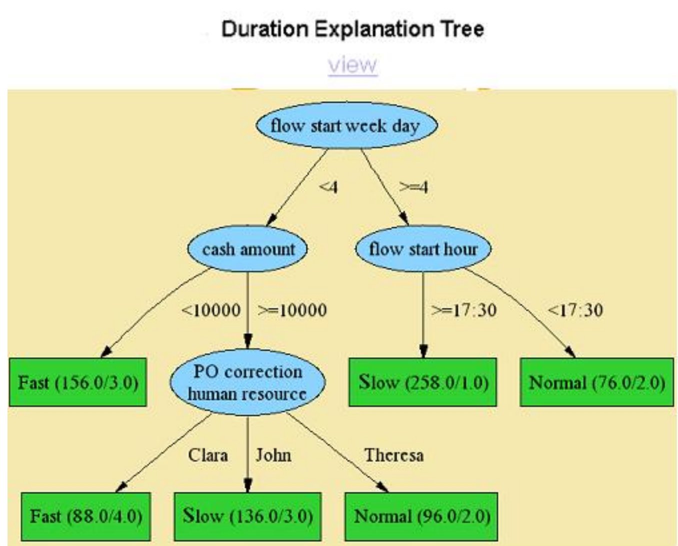
3. PROCESS OPTIMIZATION

In addition to supporting improved modeling (cf. Section 2), Business Process Intelligence also equips companies with functionalities that facilitate the optimization of different quality aspects of their business processes (Casati et al., 2002), either in terms of metrics meaningful to internal operations of the enterprise, or to the external customer perception. Again, process logs are exploited to derive useful information but in this case it is not for modeling purposes, instead it is to compute quality metrics (Casati et al., 2006) and mine process behavior related to them. For example, the performance metric of an order process could be derived from the start and end times of execution of orders processed during a given period of time. By warehousing (Casati et al., 2007) the execution data and the metrics, the business process can be monitored, analyzed and optimized with different kinds of techniques relying on data mining, statistics, simulation, and optimization, among others (Castellanos et al., 2005a, 2005b). There are many opportunities and challenges for analysis and optimization of business processes. Here we give a brief overview of four challenging areas: explanation, also called critical factor analysis (Section 3.1), prediction (Section 3.2), proactive optimization and business impact analysis (Section 3.3), and resource allocation (Section 3.4). The illustrated techniques have been implemented in Business Cockpit, a BPI platform built at HP Labs.

3.1 Critical Factor Analysis

The capability of defining and monitoring metrics (Casati et al., 2006) on a business process can be leveraged by process mining techniques that produce explanatory models to help understand the behavior of a process given by its metrics. In particular, getting insight into the critical factors determining the abnormal behavior of a metric. For example, users may want to know the characteristics of invoices and of the invoice management procedure (the cash out process) that cause a slow execution (*duration* SLA violation). In Figure 9 we observe that when the process (flow) execution starts after 17:30 or when the invoice amount is equal or greater than 10,000 and the

Figure 9. Critical factor tree cash out process duration



person executing the *purchase order correction* step is John, the process execution is *slow* in general. To obtain this functionality, models are mined as soon as a metric has been defined and computed for all the completed process instances in the process data warehouse[2] (Casati et al., 2007). It is important that the models explaining the critical factors affecting metric behaviors be easily interpretable by the business analysts. In Business Cockpit such models take the form of decision trees (Figure 9) that are automatically mined from process execution data labeled with classes corresponding to metric values (e.g., *slow, normal, fast*). To this end, a solution for each step of the data mining lifecycle had to be tailored to business processes and built into the engine. This analysis functionality is readily available to the users without requiring them to write any code. This implied a compromise between generality and ease of use on one hand, and accuracy on the other. The reader is referred to (Castellanos et al., 2005b) and (Grigori et al., 2004) for further details.

3.2 Prediction

Monitoring and explanations on metric values provide valuable visibility into current and past behavior of business processes but equally important is to provide visibility into future behavior. The ability to predict metrics and performance indicators gives the opportunity to proactively optimize the business process to improve its behavior with respect to its metrics. Predictions can be done at the instance level or at the aggregate level. The same applies to optimization. For example, we may want a prediction of the duration metric for a specific order of a customer to see if we will deliver the goods on time, and if not then we may want to increase the priority of the order so that it uses express shipment. This is referred as *instance-based* prediction (the prediction is done for a given instance while it is being executed) and *dynamic* optimization (the optimization is only for that instance and it is done during its execution), respectively. Instead, we may want to know if the average duration of orders on a certain day of next week will exceed the promised 24 hours delivery time (SLA violation) to plan for extra resources if needed. This type of prediction is referred to as *class-based* time series prediction and *static* optimization is applied in this case (Castellanos et al., 2005c). While the first kind of prediction (i.e., instance-based), as its name suggests, is based on the instance properties (e.g., day of the week that the order was submitted, type of product, region, etc), the second one is based on the time series of previous values of the metric. In consequence, suitable techniques for instance-based prediction belong to data mining, while a relaxed form of time series forecasting is used for the second one (Castellanos et al., 2005c).

In instance-based prediction (Grigori et al., 2004) a model is generated from patterns mined from execution and business data associated to process instances. For example, a pattern may indicate that if an order was received on a Friday afternoon and step *check inventory* is performed by server S3, there is an 85% chance that the order won't get shipped in less than 24 hours. Figure 10 shows the display of predictions for instances of a process on the Business Cockpit platform.

Class-based time series prediction (Castellanos et al., 2005c) is a relaxed form of time series forecasting with the goal of predicting whether a given metric (i) will exceed a certain threshold or not, (ii) is within some specified range or not, or (iii) belongs to which one of a small number of specified classes. This relaxation enables complete automation of the forecasting process to enable the analysis of hundreds or even thousands or time series of business process metrics which otherwise would not be possible to incorporate in a BPI platform. The main idea is to characterize a time series according to its components (i.e., trend and seasonality) and then apply

Figure 10. Screenshot of predictions for violation of wait time to Audit step for active cash out process instances

Metric: Wait time to audit

Number of Predictions 3363

Next Page Previous Reset

Instance ID	Current Active Node	Starting Time	Prediction	Confidence
26-1-1-11262	Vendor_mantainance	1/17/04 6:36 AM	Acceptable	0.90
26-1-1-1272	Vendor_mantainance	1/17/04 6:38 AM	Acceptable	0.97
26-1-1-1274	Vendor_mantainance	1/17/04 6:41 AM	Unacceptable	0.71
26-1-1-1274	Vendor_mantainance	1/17/04 6:44 AM	Acceptable	0.79
26-1-1-1275	Vendor_mantainance	1/17/04 6:47 AM	Unacceptable	0.70
26-1-1-1279	Vendor_mantainance	1/17/04 6:50 AM	Unacceptable	0.78
26-1-1-1281	Vendor_mantainance	1/17/04 6:53 AM	Acceptable	0.80
26-1-1-1287	Vendor_mantainance	1/17/04 6:56 AM	Unacceptable	0.70
26-1-1-1289	Vendor_mantainance	1/17/04 6:59 AM	Acceptable	0.97
26-1-1-1292	Vendor_mantainance	1/17/04 7:01 AM	Unacceptable	0.99
26-1-1-1294	Vendor_mantainance	1/17/04 7:04 AM	Acceptable	0.82
26-1-1-1297	Vendor_mantainance	1/17/04 7:07 AM	Acceptable	0.84
26-1-1-1299	Vendor_mantainance	1/17/04 7:10 AM	Acceptable	0.95
26-1-1-1301	Vendor_mantainance	1/17/04 7:13 AM	Unacceptable	0.88
26-1-1-1303	Vendor_mantainance	1/17/04 7:16 AM	Acceptable	0.99
26-1-1-1305	Vendor_mantainance	1/17/04 7:19 AM	Acceptable	0.95
26-1-1-1306	Vendor_mantainance	1/17/04 7:22 AM	Acceptable	0.71
26-1-1-1316	Vendor_mantainance	1/17/04 7:24 AM	Unacceptable	0.71
26-1-1-1317	Vendor_mantainance	1/17/04 7:27 AM	Acceptable	0.92

the most appropriate technique(s) to create a good forecasting model (Castellanos et al., 2005d). Once the model is created it can be applied to obtain a numeric prediction which is mapped to the corresponding class (e.g., exceeds-threshold or not, within-range or not, low/medium/high, or others).

Once a prediction is obtained different actions can be taken to optimize the process to improve the predicted value. When the prediction is made for a specific instance, it is possible to dynamically change things that only affect that instance to improve its execution. Typical actions are to assign a specific resource for a given action, change the priority of the instance, or dynamically change a selected path. In contrast, when the prediction is made for an aggregated metric, the optimization is static in the sense that it changes aspects of the process that are common to all its instances, like the number of resources of a given type that are allocated to a process (cf. Section 3.4). As stated above, prediction opens up the opportunity to proactively optimize aspects of a process upon the alert of undesired predicted values. Furthermore, prediction also proves helpful to business impact analysis (cf. Section 3.3).

3.3 Business Impact Analysis

Business managers need support to assess the impact of malfunctions in the Information Technology (IT) infrastructure in high level business terms. Business Cockpit provides functionality to analyze the impact on business goals (expressed as process metrics) caused by performance degradations in the IT infrastructure. The idea is to leverage the linkage information between the IT and the business layers (part of a process model defined on Business Cockpit), the IT resource monitoring functionality (provided by some infrastructure monitoring tool), and the prediction functionality (cf. Section 3.2). As a failure or degradation in an IT resource is detected, the linkage information is used to identify which nodes and consequently which processes are affected by the failure. For example, the link established between an Oracle database server and the *invoice validation* step of the cash out process makes it possible to identify this step (i.e., node) and this process as the ones affected when the server fails. Moreover, not only the impact of a resource failure can be done at the type process level, but also at the process instance level to indicate which active instances are or will be affected by the failure. Prediction models (cf. Section 3.2) are obtained beforehand to determine whether a node supported by the resource that has failed will be executed or not and whether the time interval to get to the node is larger than the average time that the resource linked to that node takes to recover from a failure. Upon a failure, the appropriate models are applied and if both predictions are positive and with high confidence values (i.e., most likely the process instance will execute that node and will do it before the resource failure is fixed), then the process instance is predicted to be affected by the failure. A confidence value for this prediction is computed as a function of the confidence of both predictions.

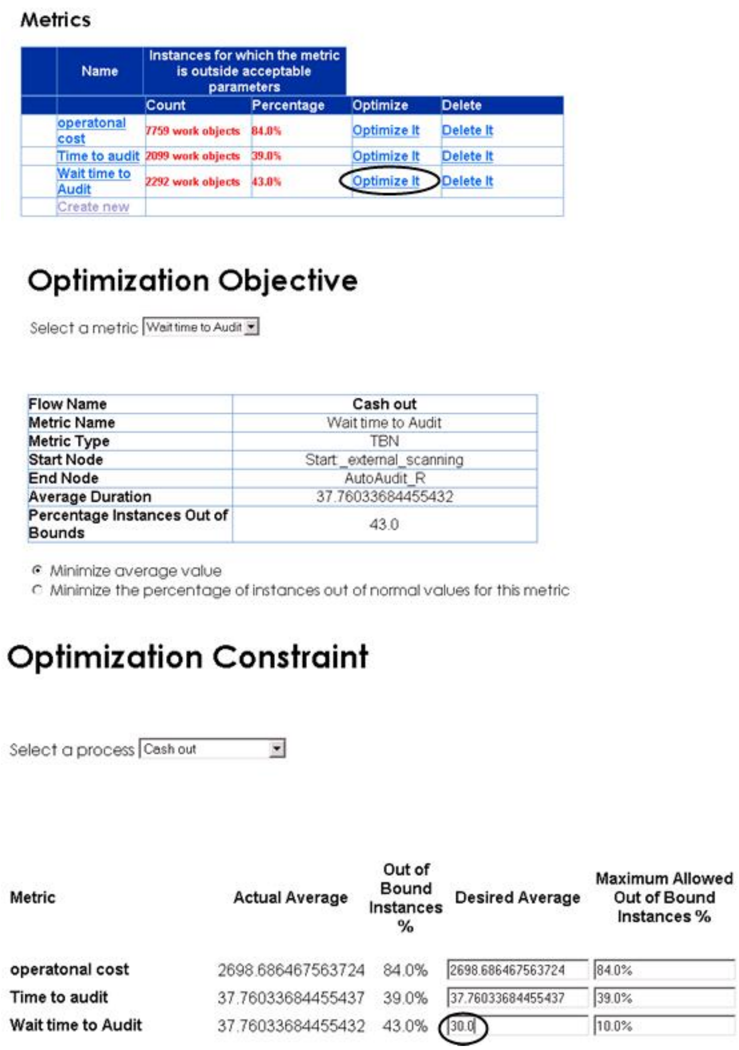
3.4 Resource Allocation

The allocation of resources to tasks can significantly affect the performance and outcome of the business processes, which in turn affects the quality of services and products of an enterprise. Identification of bottlenecks in a business process and proper allocation of resources to critical tasks can help a business meet the deadlines and SLA terms while delivering services and products at a desired quality. Business process simulation tools are used for analyzing the behavior of resources and their effect on the overall performance and outcome of processes. In particular, sensitivity analysis (what-if analysis) (Castellanos et al., 2005b) allows users to analyze outcomes of various simulated scenarios in which the effect of different parameter settings can be observed. For example, the effect of assigning two resources to a particular task, instead of only one, to know how much benefit such an additional resource allocation could bring. Possible parameters for simulation and sensitivity analysis could be not only resource pool sizes for individual tasks, but also inter-arrival rate of entities to be processed, resource behavior (response time to particular tasks), and cost of individual resources (per unit time or total).

Companies are interested not only in understanding the effect of changes in a business process but also in determining the best possible (optimal) allocation of resources in order to achieve certain performance and quality goals. Simulation leveraged with a search technique offers the solution in Business Cockpit (Castellanos et al., 2005b). Here, the objective is (i) to minimize the number of process instances that exceed a certain metric threshold (e.g., the number of invoice payments that are delayed more than 3 days) or (ii) to minimize or maximize the overall value of a given metric (e.g., minimize the average duration of processing an order). The objective is subject to constraints on the cost, other metric values and maximum number of resource elements for one or more resource pools.

Figure 11 shows an example where the goal is to determine the optimal number (within a range) of resources in the pools to minimize the average value of the metric “Wait time to Audit”. A simulation is run for a possible configuration of resource allocation and the resulting simulation execution data are transformed and loaded into the simulation results database so that metric values can be computed on them (just as for execution data of actual processes). These values determine which configuration from the search space to try next. This continues until (i) adding or removing a resource to any pool does not improve the goal, or (ii) a maximum number of simulations is reached. At the end of the process, the best configuration is presented to the user, along with the values reached for the objective and constraint metrics. Other configurations are also presented, ranked by their objective metric value, for users

Figure 11. Wait Time to Audit metric optimization



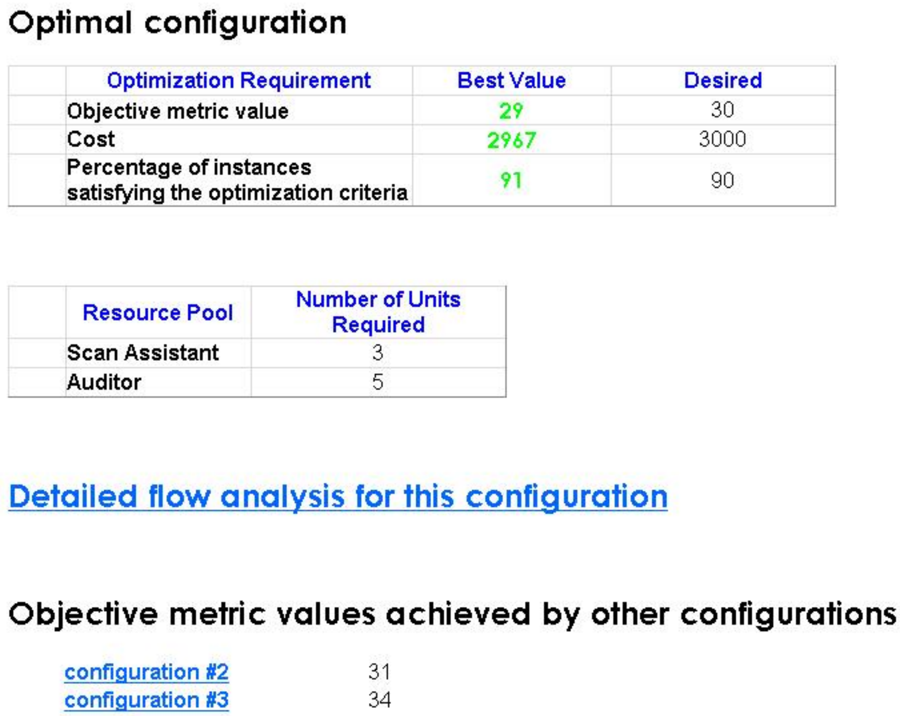
to examine them. Figure 12 shows the results for the optimization request in Figure 11.

Finally, it is also important to automatically identify the resources that perform poorly in certain contexts. Data mining, and in particular classification algorithms, can be used for this purpose (cf. Section 3.1).

4. PRACTICAL CHALLENGES AND FUTURE TRENDS

The analysis of business processes with business process intelligence techniques and tools faces several challenges in practice. In this section, we focus in particular on three types of challenges: technical challenges (Section 4.1), interpretative challenges (Section 4.2), and pragmatic challenges (Section 4.3). These challenges have to be ad-

Figure 12. Optimal number of resources to minimize the Wait time to Audit metric average value



Optimal configuration

Optimization Requirement	Best Value	Desired
Objective metric value	29	30
Cost	2967	3000
Percentage of instances satisfying the optimization criteria	91	90

Resource Pool	Number of Units Required
Scan Assistant	3
Auditor	5

Detailed flow analysis for this configuration

Objective metric values achieved by other configurations

configuration #2	31
configuration #3	34

dressed with care in order to apply BPI successfully in an organization, and they are related to the challenges identified by the CRIPS-DM process which is an acronym for Cross Industry Standard Process for Data Mining (Shearer 2000). Finally, we discuss some future trends (Section 4.4).

4.1 Technical Challenges

Business process intelligence initiatives face several technical challenges, and some of them are analogous to data warehouse challenges Brackett (1996). Most importantly, business process intelligence has to cope with the *heterogeneous systems landscape* of large enterprises. While process discovery tools can be rather easily used on log data of business processes that are executed by a single workflow system (Aalst et al., 2007b), it becomes already difficult to project transactional log data of a single ERP system such as SAP back to high-level business events (Ingvaldsen and Gulla, 2008). Even worse, business systems, for instance, in some financial institutions have been growing over 40 years and contain diverse technologies and systems ranging from classical mainframe systems to message-oriented middleware and from implementation languages such as ancient COBOL to modern object-oriented .NET. The case of a German bank reported in (Genrich et al., 2008) summarizes some of the problems for business process intelligence associated with this systems heterogeneity. Beyond the diversity and sheer complexity of its applications, most of its applications were not developed with process-orientation in mind. This poses considerable challenges to definition and integration of case identifiers across systems, i.e. matching the data fields that uniquely identify the process instance. Furthermore, log files have to be transformed from various formats to one analysis format.

Some systems do not even record log files at all (or at least some human executed steps) such that they cannot be included in an analysis directly (Genrich et al., 2008). Finally, large scale business applications typically record heaps of data. In case of the German bank 40,000 database entries were generated each day. Accordingly, the analysis tools must be able to deal with such a high amount of data in an efficient manner.

4.2 Interpretative Challenges

When the technical challenges have been sorted out, it has to be kept in mind that BPI tools provide evidence to support or falsify certain hypotheses about the business operations. The generated pieces of evidence still have to be interpreted by the persons who understand the business. As Van der Aalst et al. put it (Aalst et al., 2007b):

"It seems crucial to be closely involved with the people of the organization itself to carry out a meaningful analysis. As a small illustration of this point, it would have been impossible to determine the real value of the oddly connected activity 170_Parkeer [that] turned out not to be an activity at all, but rather a WfMS facility to suspend an operation. More importantly, it took the input of the […] process owners to identify and prioritize four locations of the process that seemed of interest to subject to a closer analysis. This certainly helped to speed up the identification of relevant results."

This statement can hardly be underestimated. The interpretative challenge stems, among others, from the fact that BPI analysis techniques can only operate on the set of events that is actually recorded for a process. In practice, not all relevant events are actually logged, and people may find ways to work around the system (Aalst et al., 2007b). Even if data is available, the quality of it is often too poor to use it directly. Given these impediments, it is crucial to understand the mindset and motivations of the various agents involved in the execution of the process (Genrich et al., 2008). Accordingly, it can be recommended to interview process stakeholders to make sure that the data is interpreted correctly.

4.3 Pragmatic Challenges

As soon as technical and interpretative issues are resolved, pragmatic conclusions can be drawn from the interpretations. The findings must be presented in an appropriate manner such that decision makers can translate them into action. It appears that a poor selection of business metrics and performance indicators prevents the effective usage of BPI tools such as management dashboards (Corea and Watters, 2007). Even if the right analysis parameters have been found, they cannot be directly translated into business objectives for staff. The reason for this observation is that some objectives enforce undesired behavior of workforce (cf. Anderson and Oliver, 1987). This is, for example, the case when call center agents hang-up on callers in order to improve their number of handled calls. Therefore, the performance objectives concluded from the BPI analysis must be chosen such that they align the behavior of the workforce with the performance objectives of the business process.

4.4 Future Trends

While the technical challenges are currently addressed by tool vendors and academia, there is only little research around so far, e.g. (Corea and Watters, 2007), that investigates the interpretative and pragmatic challenges of BPI in a systematic way. This stream of research puts a stronger emphasis on behavioral research methods including qualitative interviews and quantitative survey analysis. It is likely that we will see more work following this research paradigm as technical tools and solutions mature, providing valuable feedback for creating new innovations in BPI.

To facilitate the automatic data integration and identification, as well as the interpretation of results, recently there is a trend to embed semantics in BPM systems, yield the Semantic BPM (SBPM) systems (Hepp et al., 2005). Such systems combine Semantic Web and SWS technologies with BPM. In a nutshell, SBPM targets accessing the process space (as registered in event logs) of an enterprise at the *knowledge level* so as to support reasoning about business processes, process composition, process execution, etc. The driving force behind SBPM is the use of ontologies (Gruber, 1993). Actually, the European project SUPER (SUP, 2008) is funding research in this field. In this context, first efforts have appeared for supporting BPI based on the semantic layer of SBPM systems. For instance, the work in (de Medeiros et al., 2007a) presents an outlook on the possibilities for *semantic process mining and monitoring*, and pointers to concrete implementations in the ProM framework.

5. CONCLUSION

In this chapter we introduced business process intelligence by giving an overview of its application areas and discussing its benefits. In particular, we showed how process discovery can be applied to extract information like control-flow or the organizational structure from event logs and illustrated the application of conformance checking to detect discrepancies between a process model and the corresponding event log. In addition to process discovery and conformance checking BPI gives enterprises functionality to optimize their business processes. We introduced techniques for identifying the main factors affecting malfunctions or bottlenecks and for pro-actively optimizing business processes. Although the benefits of BPI have been widely recognized, its application in practice still faces technical, interpretive and pragmatic challenges, which need to be resolved before BPI becomes mainstream. Recent trends focus on making use of Semantic Web technologies to bring the execution and analysis of processes to a semantic level.

REFERENCES

European Project SUPER - Semantics Utilised for Process Management within and between Enterprises. *http://www.ip-super.org/*.

Aalst, W. M. P. van der, Reijers, H. A., Weijters, A. J. M. M., van Dongen, B. F., Alves de Medeiros, A. K., Song, M., & Verbeek, H. M. W. (2007a). Business process mining: An industrial application. *Information Systems*, *32*(5), 713–732.

Aalst, W. M. P. van der, & de Medeiros, A. K. A. (2005). Process mining and security: Detecting anomalous process executions and checking process conformance. *Electronic Notes in Theoretical Computer Science*, *121*, 3–21.

Aalst, W. M. P. van der, Dongen, B. F. van, Herbst, J., Maruster, L., Schimm, G., & Weijters, A. J. M. M. (2003). Workflow Mining: A Survey of Issues and Approaches. *Data and Knowledge Engineering*, *47*(2), 237–267.

Aalst, W. M. P. van der, Weijters, A. J. M. M., & Maruster, L. (2004). Workflow Mining: Discovering Process Models from Event Logs. *IEEE Transactions on Knowledge and Data Engineering*, *16*(9), 1128–1142.

Aalst, W. M. P. van der, Beer, H. T. de, & Dongen, B. F. van. (2005a). Process Mining and Verification of Properties: An Approach Based on Temporal Logic. In R. Meersman, Z. Tari, M.-S. Hacid, J. Mylopoulos, B. Pernici, Ö. Babaoglu, H.-A. Jacobsen, J. P. Loyall, M. Kifer, & S. Spaccapietra (Eds.), *OTM Conferences (1)*, volume 3760 of *Lecture Notes in Computer Science*, (pp. 130–147). Springer, ISBN 3-540-29736-7.

Aalst, W. M. P. van der, Reijers, H. A., & Song, M. (2005b). Discovering Social Networks from Event Logs. *Computer Supported Cooperative*

Work, *14*(6), 549–593.

Aalst, W. M. P. van der, Reijers, H. A., Weijters, A. J. M. M., Dongen, B. F. van, Alves, A. K., de Medeiros, Song, M., & Verbeek, H. M. W. (2007b). Business process mining: An industrial application. *Information Systems*, *32*(1), 713–732,

Anderson, E., & Oliver, R. L. (1987, October). Perspectives on behavior-based versus outcome-based salesforce control systems. *Journal of Marketing*, *51*(4), 76–88.

Brackett, M. H. (1996, July). *The Data Warehouse Challenge: Taming Data Chaos*. John Wiley and Sons.

Casati, F., Castellanos, M., Dayal, U., Hao, M., Shan, M. C., & Sayal, M. Business operation intelligence research at HP Labs. *Data Engineering Bulletin*, *25*(4).

Casati, F., Castellanos, M., Dayal, U., & Shan, M.-C. (2006). A metric definition, computation, and reporting model for business operation analysis. In Y. E. Ioannidis, M. H. Scholl, J. W. Schmidt, F. Matthes, M. Hatzopoulos, K. Böhm, A. Kemper, T. Grust, & C. Böhm (Eds.), *Advances in Database Technology - EDBT 2006, 10th International Conference on Extending Database Technology, Munich, Germany, March 26-31, 2006, Proceedings*, volume 3896 of *Lecture Notes in Computer Science*, (pp. 1079–1083). Springer.

Casati, F., Castellanos, M., Dayal, U., & Salazar, N. (2007). A generic solution for warehousing business process data. In C. Koch, J. Gehrke, M.N. Garofalakis, D. Srivastava, K. Aberer, A. Deshpande, D. Florescu, C.Y. Chan, V. Ganti, C.-C. Kanne, W. Klas, & E.J. Neuhold (Eds.), *Proceedings of the 33rd International Conference on Very Large Data Bases, University of Vienna, Austria, September 23-27, 2007*, (pp. 1128–1137). ACM, 2007. ISBN 978-1-59593-649-3.

Castellanos, M., Casati, F., Sayal, M., & Dayal, U. (2005a). Challenges in business process analysis and optimization. In C. Bussler & M.-C. Shan (Eds.), *Technologies for E-Services, 6th International Workshop, TES 2005, Trondheim, Norway, September 2-3, 2005, Revised Selected Papers*, volume 3811 of *Lecture Notes in Computer Science*, (pp. 1–10). Springer, 2005a. ISBN 3-540-31067-3.

Castellanos, M., Casati, F., Shan, M.-C., & Dayal, U (2005). (2005b). iBOM: A platform for intelligent business operation management. In *Proceedings of the 21st International Conference on Data Engineering, ICDE 2005, 5-8 April 2005, Tokyo, Japan*, (pp. 1084–1095). IEEE Computer Society, ISBN 0-7695-2285-8.

Castellanos, M., Salazar, N., Casati, F., Dayal, U., & Shan, M.-C. (2005c). Predictive business operations management. In S. Bhalla (Eds.), *DNIS*, volume 3433 of *Lecture Notes in Computer Science*, (pp. 1–14). Springer, ISBN 3-540-25361-0.

Castellanos, M., Salazar, N., Casati, F., Shan, M. C., & Dayal, U.(2005d, July) Automatic metric forecasting for management software. In *Proceedings of the 12th HP Openview University Association (HP-OVUA) Workshop*, Porto, Portugal, July 2005d. Hewlett–Packard Corporation.

Cook, J. E., Du, Z., Liu, C., & Wolf, A. L. (2004). Discovering Models of Behavior for Concurrent Workflows. *Computers in Industry*, *53*(3), 297–319.

Corea, S., & Watters, A. (2007). Challenges in business performance measurement: The case of a corporate it function. In G. Alonso, P. Dadam, & M. Rosemann (Eds.), *Business Process Management, 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007, Proceedings*, volume 4714 of *Lecture Notes in Computer Science*, (pp. 16–31), Brisbane, Australia, 2007. Springer.

de Medeiros, A. K. A. (2006). *Genetic Process Mining*. PhD thesis, Eindhoven University of

Technology, Eindhoven..

de Medeiros, A. K. A., Pedrinaci, C., van der Aalst, W. M. P., Domingue, J., Song, M., Rozinat, A., Norton, B., & Cabral, L. (2007a). An Outlook on Semantic Business Process Mining and Monitoring. In R. Meersman, Z. Tari, & P. Herrero (Eds.), *OTM Workshops (2)*, volume 4806 of *Lecture Notes in Computer Science*, (pp. 1244–1255). Springer. ISBN 978-3-540-76889-0.

de Medeiros, A. K. A. Weijters, A. J. M. M., & van der Aalst, W. M. P. (2007b). Genetic Process Mining: an Experimental Evaluation. *Data Mining Knowledge Discovery*, *14*(2), 245–304.

van Dongen, B. F. (2007). *Process Mining and Verification*. PhD thesis, Eindhoven University of Technology, Eindhoven.

van Dongen, B. F., & van der Aalst, W. M. P. (2005_, A. K. A.). Multi-phase Process mining: Aggregating Instance Graphs into EPCs and Petri Nets. In *Proceedings of the Second International Workshop on Applications of Petri Nets to Coordination, Workflow and Business Process Management (PNCWB)*.

van Dongen, B. F., Mendling, J., & van der Aalst W. M. P. (2006). Structural Patterns for Soundness of Business Process Models. In *EDOC '06: Proceedings of the 10th IEEE International Enterprise Distributed Object Computing Conference (EDOC'06)*, (pp. 116–128), Washington, DC, USA. IEEE Computer Society.

Dumas, M., ter Hofstede, A. H. M., & van der Aalst, W. M. P. (Eds.), *Process Aware Information Systems*. Wiley Publishing.

Genrich, M., Kokkonen, A., Moormann, J., zur Muehlen, M., Tregear, R., Mendling, J., & Weber, B. (2008). Challenges for business process intelligence: Discussions at the BPI Workshop 2007. In A. H. M. ter Hofstede, B. Benatallah, & H.-Y. Paik, (Eds.), *Proceedings of the BPM 2007 Workshops*, volume 4928 of *Lecture Notes in Computer Science*, Brisbane, Australia. Springer-Verlag.

Greco, G., Guzzo, A., Pontieri, L., & Saccà, D. (2006). Discovering expressive process models by clustering log traces. *IEEE Transactions on Knowledge and Data Engineering*, *18*(8), 1010–1027. ISSN 1041-4347.

Greco, G., Guzzo, A., Manco, G., & Saccà, D. (2007). Mining unconnected patterns in workflows. *Information Systems*, *32*(5), 685–712.

Grigori, D., Casati, F., Castellanos, M., Dayal, U., Sayal, M., & Shan, M. (2004). Business process intelligence. *Computers in Industry*, *53*(3), 321–343.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199–220. ISSN 1042-8143.

Günther, C. W., & van der Aalst, W. M. P. (2007). Fuzzy mining - adaptive process simplification based on multi-perspective metrics. In G. Alonso, P. Dadam, & M. Rosemann, editors, *BPM*, volume 4714 of *Lecture Notes in Computer Science*, (pp. 328–343). Springer. ISBN 978-3-540-75182-3.

Hepp, M., Leymann, F., Domingue, J., Wahler, A., & Fensel, D. (2005). Semantic Business Process Management: a Vision Towards Using Semantic Web services for Business Process Management. In *IEEE International Conference on e-Business Engineering (ICEBE 2005)*, (pp. 535 – 540). ISBN 0-7695-2430-3.

Herbst, J., & Karagiannis, D. (2004). Workflow Mining with InWoLvE. *Computers in Industry*, *53*(3), 245–264. ISSN 0166-3615. http://dx.doi.org/10.1016/j.compind.2003.10.002.

Hornix, P. T. G. (2007). *Performance Analysis of Business Processes Through Process Mining.* Master's thesis, Eindhoven University of Technology, Eindhoven.

Ingvaldsen, J. E., & Gulla, J. A. (2008). Preprocessing support for large scale process mining of sap transactions. In A. H. M. ter Hofstede, B. Benatallah, & H.-Y. Paik, (Eds.), *Proceedings of the BPM 2007 Workshops*, volume 4928 of *Lecture Notes in Computer Science*, Brisbane, Australia. Springer-Verlag.

G. Keller, M. Nüttgens, and A.W. Scheer. Semantische Processmodellierung auf der Grundlage Ereignisgesteuerter Processketten (EPK). Veröffentlichungen des Instituts für Wirtschaftsinformatik, Heft 89 (in German), University of Saarland, Saarbrücken.

Ly, L. T., Rinderle, S., Dadam, P., & Reichert, M. (2005). Mining Staff Assignment Rules from Event-Based Data. In *Business Process Management Workshops*, (pp. 177–190).

Pinter, S. S., & Golani, M. (2004). Discovering Workflow Models from Activities' Lifespans. *Computers in Industry*, *53*(3), 283–296.

Power, D. J. (2007, March). *A Brief History of Decision Support Systems*. dssresources.com, http://DSSResources.COM/history/dsshistory.html, version 4.0 edition.

Rozinat, A., & van der Aalst, W. M. P. (2008). Conformance checking of processes based on monitoring real behavior. *Inf. Syst.*, *33*(1), 64–95.

Rozinat, A., & van der Aalst, W. M. P. (2006). Decision Mining in ProM. In S. Dustdar, J. L. Fiadeiro, & A. P. Sheth (Eds.), *Business Process Management*, volume 4102 of *Lecture Notes in Computer Science*, (pp. 420–425). Springer, 2006. ISBN 3-540-38901-6.

Rozinat, A., de Jong, I. S. M., Günther, C. W., & van der Aalst, W. M. P. (2007). *Process Mining of Test Processes: A Case Study*. BETA Working Paper Series, WP 220, Eindhoven University of Technology, Eindhoven.

Sarbanes-Oxley Act. Sarbanes-Oxley Act of 2002. http://www.sarbanes-oxley-forum.com/.

Schimm, G. (2004). Mining Exact Models of Concurrent Workflows. *Computers in Industry*, *53*(3), 265–281. ISSN 0166-3615. http://dx.doi.org/10.1016/j.compind.2003.10.003.

Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, *5*, 13-22.

Song, M., & van der Aalst, W. M. P. (2007). *Towards Comprehensive Support for Organizational Mining.* BETA Working Paper Series, WP 211, Eindhoven University of Technology, Eindhoven.

Wen, L., van der Aalst, W. M. P., Wang, J., & Sun, J. (2007). Mining process models with non-free-choice constructs. *Data Mining Knowledge Discovery*, *15*(2), 145–180.

KEY TERMS AND DEFINITIONS

Business Process Intelligence: Refers to the application of Business Intelligence (BI) techniques to business processes (Grigori et al., 2004)

Conformance Checking: Compares an event log with a (process) model to check for undesired behavior.

Critical Factor Analysis: Analyses past process executions to identify the main factors determining specific process behaviors (with respect to the process metrics).

Event Log: An event log records business events from process-aware information systems (PAIS) such as WFM (Workflow Management), ERP (Enterprise Resource Planning), SCM (Supply Chain Management) and CRM (Customer Relationship Management) systems. Typically, event logs contain information about start and completion of activities, their ordering, resources which executed them and the process instance they belong to.

Prediction: It is the application of data mining and forecasting techniques to estimate future behaviors of a process.

Process Analysis: Refers to the analysis of past process executions with respect to process performance metrics.

Process Discovery: Refers to the analysis of business events recorded in event logs to discover process, control, data, organizational, and social structures (Aalst et al., 2007b).

Process Mining: Is the discovery of information based on event logs. Process discovery,, conformance checking, critical factor analysis and prediction qualify as process mining techni

Process Monitoring: Refers to the monitoring of running process instances to inform users about critical events.

EXERCISES

Exercise 1: As illustrated in Section 1 BPI comprises several application areas. Recently most BPM vendors have extended their portfolio with BPI functionality, but not everyone is supporting the entire spectrum. Browse the website of selected vendors of BPM suites (e.g., IBM, SAP, Tibco, Oracle, Pallas Athena, Lombardi, webMethods, Savvion) and try to find out which BPI functionality they support. Which application areas are supported by which BPM suites?

Exercise 2: In the context of BPI a large number of buzzwords like Business Activity Monitoring (BAM), Business Operations Management (BOM), Business Process Intelligence (BPI), Process Mining, and Business Operations Intelligence (BOI) exist. Although these buzzwords all relate to BPI slight differences exist and they sometimes refer to different applications areas of BPI.

Browse the website of selected vendors of BPM suites (e.g., IBM, SAP, Tibco, Oracle, Pallas Athena, Lombardi, webMethods, Savvion).

a. Which buzzwords are used by which vendors?

b. Which application areas of BPI are usually covered by which buzzwords?

Exercise 3: In Section 1 different synonyms for BPI and different application areas are described. Create a mind map to organize all these terms and concepts. The mind map should have BPI in its center and the different synonyms and application areas should be organized around the central node in branches. In case you are not familiar with mind maps you can find a description of this technique as well as guidelines at http://en.wikipedia.org/wiki/Mind_Map. Examples for mind maps are provided at http://www.buzanworld.com/mindmaps/.

Exercise 4: To familiarize you with the analysis techniques provided by process mining tools, we like you to use the open source ProM tool that can be downloaded from www.processmining.org to analyze an event log for the running example used in this chapter: a Dutch rental housing organization. The event log is located at http://tabu.tm.tue.nl/wiki/_media/tutorial/EventLogDutchRentalHouseOrganization.zip. Additionally, you may want to have a look at the ProM tutorial provided at the above URL. Your analysis should cover the following points:

What are the five most frequent paths for this process? How much of the log do they account for? What are their average throughput times?

How does the process model that describes the behavior in the log look like? Does this model completely fit the log? If not, how many instances fit this model and how many do not? Where are the problems for the non-fitting process instances?

What are the roles in the organization? Which employees can perform both immediate and internal repairs? Who is handing over work to whom? Who are the central employees for this process?

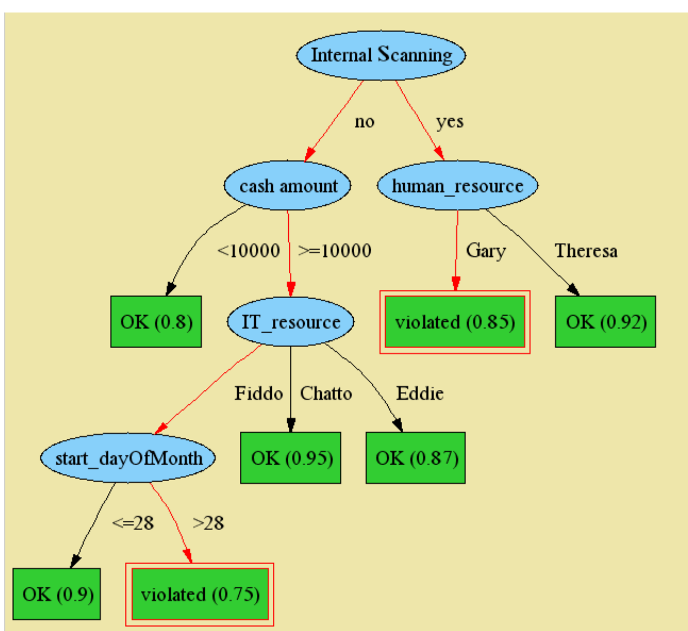
Is the rule "immediate repairs that could not be solved should be handled by an internal team again before being sent to an external team" being indeed obeyed? Which percentage of process instances complies with this rule? Which percentage does not comply with this rule?

What kind of repair is more common in the organization? Which percentage of repairs can be fixed in a first attempt?

Where are the bottlenecks in the system? Which bottlenecks are due to long waiting times and which are due to long execution times? Which three tasks have the longest average Sojourn times?

What are the business rules that usually apply for the moments of choice in the model? (Note: At least one rule should be reported!)

Exercise 5: The decision tree in figure below shows the critical factors affecting the violation of an SLA on the maximum duration for processing an invoice. Notice that duration is the metric on which this SLA is defined. Interpret the decision tree and indicate which factors lead to the violation of the SLA, what could be a possible reason and how some of these violations could be eliminated. (note: Fiddo, Chatto and Eddie are Unix servers)



Exercise 6 (advanced): Think of a business process of your choice and model it with a simple diagram of nodes and arcs. Then, think of the possible predictions that would be useful to have for this process and the opportunities that these predictions would open up for optimizing this process. Create a list of the predictions and the corresponding optimization actions to proactively eliminate the occurrence of the predicted behavior (assuming undesired behavior –metric values- is predicted) or at least to minimize its negative effect.

Exercise 7: A naïve way to do business impact analysis when a resource fails is to mark all the activities supported by the resource as potentially impacted. A smarter way to do it is using the technique explained in Section 3.3, where intelligence is injected to narrow down the set to only those activities with high probability of being impacted. What is the technique used to make business impact analysis more intelligent and what is the benefit of doing it in terms of the actions that need to be done to cope with the impact? (think as if you were the IT manager, what would you need to do to keep running your process and meeting SLAs while the resource's failure is solved? How would that be different if you get a long list of activities potentially impacted versus a short list of highly probable impacted activities?)

Exercise 8: Search the web for the MXML format for process mining data. Draw a UML class diagram of the MXML format. Furthermore, explain why log information from database systems and web servers cannot be directly mapped to MXML.

Exercise 9: The results of a process mining project show that the cases where employee Peter is involved take on average 50% longer than the other cases. Give four different explanations for this fact. Consider the different cases that Peter is a highly-qualified employee, a bottleneck of the process, a lazy employee, or an employee that only works at a certain time of the day. What does this variety of potential explanations imply for process mining?

Exercise 10: A process intelligence project reveals that the average call time in a call center of a bank takes 30% longer than the average across the financial industry. The bank plans to introduce a performance measurement system to keep track of the average call time. In a speech to the work force the CEO announces that in the next year the call center agents should reduce the average call time by 10%. What is the risk of taking average call time as a performance indicator for call center agents? Consider the different strategies a call center agent might consider to improve the performance in terms of this metric.

ENDNOTES

1. LTL stands for *Linear Temporal Logic*.
2. A process data warehouse is the repository designed specifically to store all process execution related data.