Process-Mining-enabled audit of Information Systems: methodology and an application

Pierluigi Zerbino*, Davide Aloini, Riccardo Dulmin, Valeria Mininno

Department of Energy, Systems, Territory and Construction Engineering, University of Pisa
Largo Lucio Lazzarino 1, 56122, Pisa, Italy

E-mail addresses: pierluigi.zerbino@for.unipi.it (*corresponding author); davide.aloini@unipi.it; riccardo.dulmin@ing.unipi.it; valeria.mininno@dsea.unipi.it;

Abstract. Current methodologies for Information Systems (ISs) audits suffer from some limitations that could question the effectiveness of such procedures in detecting deviations, frauds, or abuses. Process Mining (PM), a set of business-process-related diagnostic and improvement techniques, can tackle these weaknesses, but literature lacks contributions that address this possibility concretely. Thus, by framing PM as an Expert System (ES) engine, this paper presents a five-step PM-based methodology for IS audits and validates it through a case in a freight export port process managed by a Port Community System (PCS), an open electronic platform enabling information exchange among port stakeholders.

The validation pointed out some advantages (*e.g.* depth of analysis, easier automation, less invasiveness) of our PM-enabled methodology over extant ESs and tools for IS audit. The substantive test and the check on the PCS processing controls and output controls allowed to identify four major non-conformances likely implying both legal and operational risks, and two unforeseen process deviations that were not known by the port authority, but that could improve the flexibility of the process. These outcomes set the stage for an export process reengineering, and for revising the boundaries in the process flow of the PCS.

Keywords: Information Systems audit; error detection; Process Mining; Business Process Management (BPM); Risk Management; Port Community System (PCS).

1. Introduction

The current pervasiveness of Information Systems (ISs) for supporting the firms' business processes has reached the extent that the digital and the physical flows are completely intertwined and impossible to separate (van der Aalst, 2016). Major corporate and accounting scandals (*e.g.* Enron and WorldCom) and the consequent normative advancement, such as the *Sarbanes-Oxley Act* of 2002 and the *Basel II Accord* of 2004, has entailed raising concern for sharper IS risk management approaches and improved audit systems (Rozinat & van der Aalst, 2008; van der Aalst, 2016).

Typically, IS risk analysis involves the systematic operational control of processes within the scope of specific audits, which are one of the most spread techniques of analysis in this field (Mock & Corvo, 2005). An audit is a planned, systematic, and independent examination for evaluating to which extent some criteria, enforced by law or by internal policies, are met, and any non-conformance observed in the evidences from auditing an IS may be flagged as inefficiency, fraud, or abuse (van der Aalst et al., 2010).

Unfortunately, audits are time-consuming and they still fail often in detecting violations quickly (ACFE, 2010). The "*expectation gap between internal auditing and its stakeholders is widening*" (Erasmus & Coetzee, 2018, p. 91; *cf.* Institute of Internal Auditors, 2014) and, in the IS field, this could be due to two weaknesses of current audit tools and methodologies. First, auditors usually have to test the controls by relying on a small dataset of offline samples (Accorsi & Stocker, 2012; Carlin & Gallegos, 2007), which forces them to express judgments and to make time-costly decisions based on a limited overview of the processes under analysis. Second, there is a dearth of specific tools supporting the automatic execution of the audits, especially in workflow systems, which entails huge amounts of manual work (Accorsi & Stocker, 2012; Hosseinpour & Jans, 2016).

Scientific literature has investigated the exploitation of Expert Systems (ESs) for improving such audits. ESs for audit purposes has been a prolific area of research until the late nineties (*e.g.* Akoka & Comyn-Wattiau, 1996; Lee & Jeong, 1995; Richard Ye, 1995). After a plateau of research, it has gained new momentum (Issa, Sun, & Vasarhelyi, 2016), mostly because of the acknowledged advantages such as better understanding of task processes, increased knowledge, and knowledge transferability (Omoteso, 2012).

More recently, academics have argued that the exploitation of Process Mining (PM) could overcome the above-mentioned weaknesses by allowing auditors to analyze whole process datasets effectively and mostly in an automated way, using historical and / or current data (van der Aalst et al., 2010). PM is a set of techniques for monitoring and improving business processes on the basis of data from the event logs (van der Aalst et al., 2012). Since it links process analysis to machine learning and data mining (van der Aalst, 2016) – two well-known ES categories (see Liao, 2005) – it can be framed within the ES scope as an engine for extracting the actual business rules of the IS to be compared to those set by the decision maker.

Although some efforts linking PM and auditing have been spent, they have mostly been focused on the accounting (Jans, Alles, & Vasarhelyi, 2014) and the financial fields (Werner, 2017; Werner & Gehrke, 2015). The few contributions on PM-based IS audits are strongly limited in both scope and considered dataset (*e.g.* Accorsi & Stocker, 2012; Tawakkal, Kurniati, & Wisudiawan, 2017) and they clarify neither how to practically exploit PM for such objective nor what its application could likely imply.

Accordingly, in this paper, we propose a five-step PM-enabled methodology for IS audits, and we validate it by means of a case developed in a Mediterranean port on the freight export process managed by a Port Community System (PCS), an open IS that handles and streamlines the information exchange among port stakeholders. The application of the methodology singled out those process deviations, within the IS, which could represent relevant legal and operational risks. The analysis of the results led us to suggest possible solutions for limiting the occurrence of the detected non-conformances.

This manuscript presents the following contributions:

- It integrates the Expert and Intelligent systems and Risk Management streams by defining and successfully validating a thorough methodology for conducting IS audits enabled by PM – a set of techniques encompassed by the ES scope. We structured it by proposing specific inputs, outputs, ad-hoc decision points / rules / criteria, sub-tasks and procedures for obtaining and prioritizing the audit evidences. This methodology presents some advantages over the extant ESs for IS audit in terms of deeper and wider scope of analysis, easier automation, lesser invasiveness, and higher level of detail. In addition, it takes into consideration some aspects, *e.g.* a quantification of the process structuredness through an ad-hoc index, whose relevance is recognized albeit overlooked in PM literature.
- The validation of the methodology through an IS audit on a PCS is an answer to the need, highlighted by Omoteso (2012), to deepen the exploitation of ES-enabled audits in the public sectors and the practical implications of using ESs on real audits.
- It provides an operational guide that bridges the gaps of the current approaches and Information Technology (IT) tools for off-line IS auditing.
- It contributes to the PM literature by elaborating on an unprecedented, in-depth case study about a thorough PM-enabled IS audit.

The remainder of the paper is structured as follows: section 2 presents the related work; section 3 describes the methodology we propose, while section 4 details the validation; sections 5 covers both results and discussion of the case; section 6 debates the methodology in light of its validation; section 7 expounds conclusions and future work.

2. Related work

This section starts with an overview about IT tools and ESs for IS audit (sub-section 2.1). Thus, it details the PM topic and its suitability for audit activities (sub-section 2.2).

2.1 Information Systems audits

IS audits often rely on the review of past assessment reports, system logs, and audit trails (Gibson, 2014), and are usually conducted by means of both qualitative and quantitative tools, for instance documentation standards, process narratives, test of application controls (Bellino & Hunt, 2007). Audit frameworks, such as ISO 27001, COBIT, ITOL, FISCAM, CONNECT, are used as a reference which IS audit processes and procedures should be aligned with (Wright, 2008). IS audit frameworks are largely accepted because of their structured scope and comprehensiveness, but they do not suggest any operational tool, transferring this decision to the auditors according to the nature of the process / activity to be assessed. Two widespread families of IS audit solutions are Computer-Assisted Audit Tools and Techniques (CAATTs) and ESs.

2.1.1 CAATTs

CAATTs are computer-based solutions that improve efficiency and flexibility of an auditor's work, enabling the execution of tasks that would be excessively time-consuming to perform manually (Coderre, 2015). Although there are several types of CAATTs (*cf.* Braun & Davis, 2003; Pedrosa & Costa, 2014; Sayana, 2003), auditors still prefer data extractions, analytics, and sampling tools rather that those requiring a strong background in statistics, mathematics, and artificial intelligence (Pedrosa & Costa, 2012). Table 1 illustrates some of the most spread CAATTs, along with their pros and cons.

Table 1. Description, Pros, and Cons of CAATTs

CAATT	Description	Pros	Cons
Integrated Test Facility (ITF) (Braun & Davis, 2003; Coderre, 2015)	Entry of test items into an IS for creating transactions to be applied against ad-hoc-created dummy accounts, organizational entities, or	ITF is designed during the IS development; the obtained information does not rely on the client.	High IS corruption risk; removing the effect of dummy transactions may not be simple; high

	departments. The results yielded by the applications are compared to the expected outcomes.		expertise required for designing the audit modules.
Parallel simulation (Braun & Davis, 2003; Coderre, 2015)	Development of a partial or total duplication of the client's application to replicate and check the results using client-supplied data.	Parallel simulation runs independently from the client's application and does not affect it.	The needed advanced programming skills often require computer specialists and may be time-consuming.
Embedded Audit Module (EAM) (Braun & Davis, 2003; Zhao, Yen, & Chang, 2004)	Development of embedded audit modules compiled within the application. They can be turned on and off at intervals.	EAMs enable continuous monitoring, and are strongly effective for data-intensive, online systems and critical applications.	Programming expertise; client interaction is needed when the application is revised; the continuous screening of the transactions can negatively affect the processing speed.
Generalized Audit Software (GAS) (Ahmi & Kent, 2012; Braun & Davis, 2003; Debreceny, Lee, Neo, & Shuling Toh, 2005)	Software packages for data extraction and analysis, and for performing a set of audit procedures and statistical routines.	Relatively easy to use; almost all the processing occurs outside the client's IS.	Exclusive focus on client's end data; general, not-specific audit purposes.

In IS audits, control tests and substantive tests are needed. A control test checks for violations of the internal controls, while a substantive test detects errors in the yielded data (Rezaee, Elam, & Sharbatoghlie, 2001). Besides all the specific weaknesses reported in table 1, most CAATTs provide only a substantive test of the IS, potentially leading to poor data quality, mask-up effect, and incorrect decision-making (Huang, Yen, Hung, Zhou, & Hua, 2009). EAMs can perform control tests, but their embedded nature makes them invasive and excessively specific (Huang et al., 2009). Moreover, the current wider data volumes exacerbated the difficulties in analyzing entire datasets through CAATTs. This is a strong limitation because, although appropriate sampling is a reliable procedure, overlooked outliers may conceal non-conformances in the IS that are unacceptable due to internal or enforced-by-law regulations.

2.1.2 Expert Systems

ESs are a branch of Artificial Intelligence and make use of advanced, task-specific knowledge, transferred from expert humans to a computer for solving complex problems or giving advice (Liao, 2005). Historically, ESs have been problem-specific, hard-to-update, and with limited adaptability (Lombardi & Dull, 2016). Yet, they have largely been used for auditing because they present reduced development costs, increased availability of expertise from multiple sources, time saving, steady and complete response at all times, automation of procedures (Giarratano & Riley, 2005). Research about their exploitation in auditing has mostly been developed in the accounting field and has almost been steady in the years (Issa et al., 2016; Omoteso, 2012; Sutton, Holt, & Arnold, 2016). Table 2 presents an overview of ES-enabled IS-audit-related works.

Table 2. Overview of ESs used for IS audits

Source	ES	Pros	Cons
Tsudik & Summers (1990)	AudES, for computer security auditing	Easy to customize; one of the first ESs for audit purposes;	Lack of possibility to add uncertainty management ability;
Sanchez & Rodriguez (1994)	ZYANYA, a rule-based ES for auditing the lifecycle of IS development	Training tool that interacts with the user; covers the whole implementation, from feasibility to installation;	Qualitative approach only; narrow scope;
Akoka & Comyn-Wattiau (1996); Comyn-Wattiau &	INFAUDITOR, an ES that presents the audit domains and controls as a hierarchical tree,	Extensive scope including both managerial and technical aspects of the IS; rules of	All the AHP-related limitations (*e.g.* Dyer, 1990); no substantive tests;

Akoka (1996);	adopting an analytical hierarchy process	customization for enhanced adaptability;	
Atymtayeva, Bortsova, Inoue, & Kozhakhmet (2012)	Fuzzy ES for auditing information security	Combination of the knowledge from experts and information security standards	Need for improved methods for the optimal choice of the recommendations and for a refined system of rules; narrow scope;
Kanatov, Atymtayeva, & Yagaliyeva (2014)	ES for information security audit and management, based on web-applications and fuzzy tools	The web-based application module fosters the resolution of security-related problems	Low number of rules; narrow scope;
Piech & Grodzki (2017)	ES for auditing communication security in ISs	Increase of investigated threaten structures; possibility of threat prediction;	Need to enrich and strengthen the rules; narrow scope;

As suggested by table 2, initial research focused on ESs for comprehensive IS audit tools. After a decline in favor of an interest towards artificial neural networks for audits (Issa et al., 2016), IS-audit-related research on ESs has come back on narrower aspects, *e.g.* information security. Thus, the spreading of process-aware IS (Weber, Reichert, & Rinderle-Ma, 2008), the need for auditing in workflow systems, and the current, indissoluble bond between information and physical flows in IS have been overlooked by the ES-related research for IS audits. PM, a set of techniques we framed as an ES engine and that may cope with these shortages in IS auditing, is the subject of the following sub-section.

2.2 Process Mining

This sub-section presents an overview on PM (sub-section 2.2.1) and on its employment in IS audit (sub-section 2.2.2).

2.2.1 Process Mining overview

PM is a set of tools for discovering, monitoring, and improving actual business processes drawing from the event logs largely available in today's ISs. The event logs are arranged in process instances (*cases*), organized in *activities* consisting in a series of *events* (van der Aalst, 2016). Event logs can be exploited for three kinds of analysis (van der Aalst et al., 2012): (I) process *discovery*, whose purpose is to produce the actual model of a process without the need to rely on a-priori information; (II) process *conformance*, which checks if the information in the log conforms with a process model, business requirements, or policies to comply with; (III) process *enhancement*, which aims to improve or extend an existing process model. The scope of PM analyses covers a variety of application domains – ranging from smart maintenance to quality management – as long as the event logs are available (van der Aalst, 2016). In supporting operational activities, PM comes up beside other data mining approaches (*e.g.* Kamsu-Foguem, Rigal, & Mauget, 2013; Ruiz, Kamsu-Foguem, & Grabot, 2014), but its spectrum of applications is wider.

Each PM project is structured according to a general lifecycle, roughly defined in five stages by van der Aalst et al. (2012) without specifying instructions, sub-tasks and their order, practices and procedures for conducting such projects. PM projects are affected by the PM perspective that the investigator adopts, which could focus on a whole business process (*process perspective*), on the relationship among the resources performing the tasks (*organizational perspective*), or on a single case (*case perspective*) (van der Aalst, 2016).

Scientific literature proposes several PM algorithms, *e.g.* α-algorithm (van der Aalst, 2016), genetic algorithm (De Medeiros, Weijters, & van der Aalst, 2005), heuristic miner (Weijters, van der Aalst, & De Medeiros, 2006). A fundamental criterion in choosing a PM algorithm is the structure of the process under analysis (De Weerdt et al., 2012). Van der Aalst (2016) claims that the structure of a process is a continuum, whose extremes are unstructured processes (*spaghetti-like*) and structured processes (*lasagna-like*), but also that there is not a formal rule for categorizing a process precisely. On one hand, he argues that a process is *spaghetti-like* when the number of unique traces (variants)

in a log represents a relevant part of the total number of cases: this definition is rather qualitative, and no quantitative metrics are provided. On the other hand, a process is *lasagna-like* when "*more than 80% of the events happen as planned and stakeholders confirm the validity of the model*" (p. 387). Nonetheless, PM literature lacks papers presenting a univocal guideline about which could be the best process discovery or process conformance algorithm to apply, although some works provide hints for a reasoned choice (*e.g.* De Weerdt, De Backer, Vanthienen, & Baesens, 2012; Lang, Bürkle, Laumann, & Prokosch, 2008).

In case of *lasagna* processes, given their streamlined structure, most PM algorithms can be suitable (van der Aalst, 2016). Instead, the low-structured backbone of a *spaghetti* process offers better chances for PM-enabled improvements (van der Aalst, 2016) but, in this case, most PM algorithms could produce tangled, hard-to-interpret process maps (Günther and van der Aalst, 2007). For highly unstructured processes, the Fuzzy Miner (Günther and van der Aalst, 2007) or the Declare Miner algorithms (Burattin, Maggi, & Sperduti, 2016) yield easier-to-handle results.

2.2.2 *Process Mining in IS audits*

PM literature has mostly focused on process discovery rather than on Conformance Checking (CC) (De Leoni & Marrella, 2017). Yet, CC is relevant to auditing because it evaluates if a business process is performed within specific boundaries enforced by the firm and / or by external actors. Event logs are yielded through a *business process provenance approach* – a set of activities for guaranteeing that the logs cannot be modified or obscured – which makes them a reliable basis for auditing (van der Aalst, 2016).

Van der Aalst et al. (2010) propose a quite general framework for PM-enabled audits (figure 1), which relies on four main components: historical log data; current log data; the *de jure* model, describing how a process is supposed to perform; and the *de facto* model, depicting how the process performs actually.

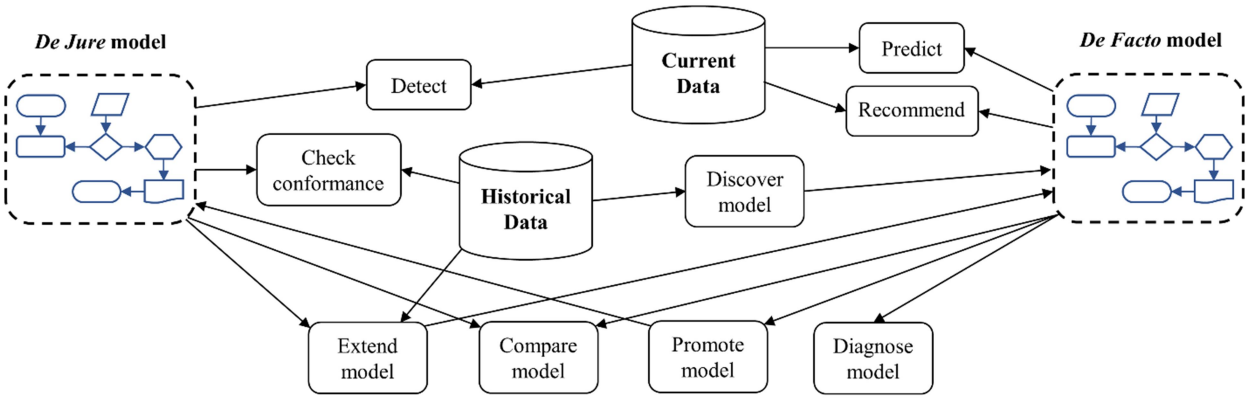

Figure 1. PM-enabled audit 2.0 framework (re-adapted from van der Aalst et al., 2010)

According to figure 1, the interplay among the four components allows for several investigations. Auditors could exploit historical data for scoping specific cases / events in the log for ad hoc audit questions or for diagnosing deadlocks or anomalies. Other purposes may be to extend the derived *de facto* model with additional data on performance indicators, or to perform a CC. The *de jure* and the *de facto* models could be juxtaposed for enabling further in-depth analysis: for instance, the *de facto* model could be not consistent with the standard pre-existing model, and it could be promoted to be the new *de jure* model if it performs better. Current data may be used for predicting specific outcomes or for evaluating the likelihood of violating some constraints.

PM has been applied to fraud detection (Yang & Hwang, 2006) and mitigation (Jans, van der Werf, Lybaert, & Vanhoof, 2011) in healthcare but, despite the above-mentioned scenarios, the number of contributions on PM in IS audit is quite limited. Accorsi & Stocker (2012) report on a case study about the application of PM for audit purposes, with an exclusive focus on security requirements. By performing a CC on a simulated log, they show the power of PM in detecting control deviations

and separations of duties. Also, they stress that the PM tools they exploited – developed in ProM, an open source framework for PM algorithms – are not suitable for industrial size analyses because of scalability and technical interoperability issues. More recently, Tawakkal et al. (2017) investigated the role of PM in data collection and validation within an IS audit limited to a single sub-process of the COBIT 5 framework. By analyzing approximately one month of data from a fashion distribution firm, they singled out some weaknesses in conformance and risk management. Interestingly, they advocate the need to combine the results from PM with other data sources – *e.g.* interviews, questionnaires, document review – because the event logs do not contain all the information required for conducting an effective IS audit.

Thus, extant efforts in PM-enabled IS audits have privileged very focused contexts through small datasets, without providing a comprehensive overview on the topic. Moreover, an operational methodology for conducting such audits has not yet been proposed.

3. The methodology

This section describes the methodology we propose for supporting IS audits through PM. The methodology draws on the PM lifecycle by van der Aalst et al. (2012), contextualizing it in the IS audit, and specifying sub-tasks, decision points / rules / criteria, inputs, outputs, and procedures. Given the audit objective, it mostly focuses on *discovery* and *conformance*: *enhancement* activities may be enabled by a sound audit, but they fall outside the auditor's tasks. Figure 2 illustrates the methodology, whose steps are detailed in the following sub-sections.

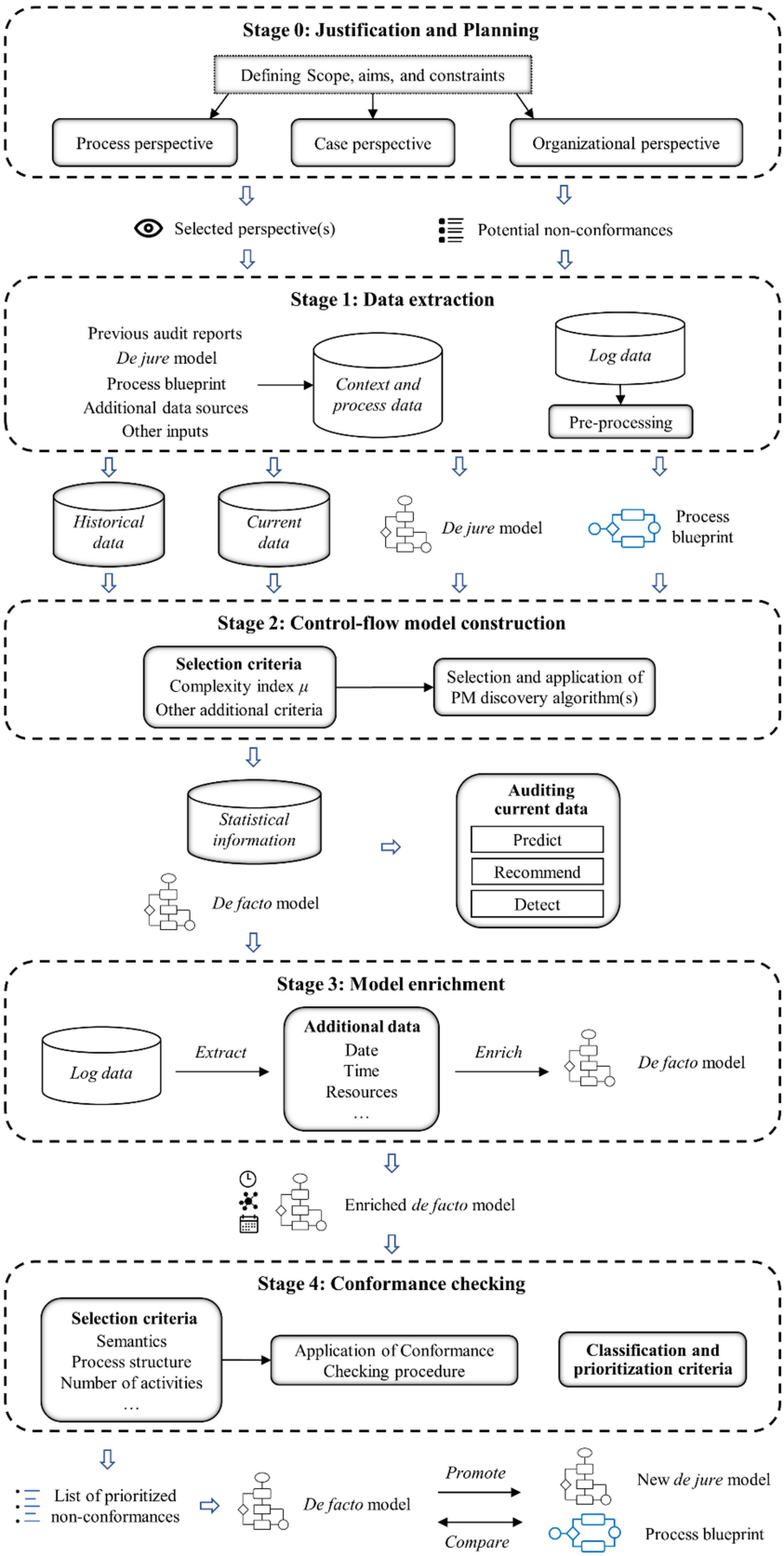

Figure 2. The proposed methodology for PM-enabled IS audit

3.1 Stage 0: Justification and planning

The PM-enabled audit process starts with a justified planning that should clarify, at least: (I) scope, (II) aims, (III) constraints of the initiative, and (IV) the PM perspective(s). Organizational and case perspectives could benefit from focusing on one or more non-conformances ex ante, *e.g.* a specific

fraud or abuse (Jans et al., 2011): in this case, a list of such criticalities must be prepared. Differently, the process perspective should be approached through an open-minded attitude, without any pre-conceived non-conformance because it can influence the auditor.

3.2 Stage 1: Data extraction

As a process log does not contain all the information required for a thorough IS audit, the data extraction should be addressed and enriched by a deep understanding of the domain which the audit should be executed in. In table 3 we propose all the data sources to consider, if possible, and the justification for including them.

Table 3. Inputs for PM-enabled IS audit

Data source	Detail
Past audit report(s)	They may be useful for collecting information about past non-conformances.
De jure model	It allows to figure out how the selected process should be executed by the ISs, according to those criteria whose compliance with will be assessed in the audit.
Business process blueprint	It differs from the *de jure* model because the former focuses on how the process should be executed, on its general flow, and on how it could relate to the other business processes, while the latter is bound to the logic of the IS. These two blueprints could differ because of a poor implementation of the IS, or due to a low fit between the tasks to execute and the technology supporting the execution (*cf.* Goodhue & Thompson, 1995). The blueprints could be depicted in BPMN or other standards.
Qualitative and quantitative data about the domain under analysis	They are fundamental for improving knowledge and awareness about all the dynamics, typical or less frequent, performed in the domain which the audit refers to. They can be collected through several methods: *e.g.* interviews, questionnaires, focus groups, analysis of documents about the IS and the processes.
Historical and / or current log data	They are the main input for PM algorithms, and they usually require pre-processing in order to make them available for and consistent with the audit purposes. Pre-processing activities can involve the selection of a particular data format, *e.g.* XML or XES, and a check for the quality of the databases (Wang, Caron, Vanthienen, Huang, & Guo, 2014). The attributes identifying the process instances should be defined according to the perspective chosen in the stage 0.
Additional inputs	These are all the further inputs that could be useful to an auditor, *e.g.* additional handmade blueprints.

If the process instances have been appropriately defined, it is possible to extract data from the log. The timespan of the extraction should be sufficient to include at least the occurrence of the main process dynamics depicted in the *de jure* model. Incomplete cases in historical data, identified by analyzing the extracted log and the other available information, should be removed.

3.3 Stage 2: Control-flow model construction

This stage consists in creating the Control Flow (CF) model – *i.e.* the ordering of the activities – from historical data by means of process discovery techniques. As regards the algorithm selection (*cf.* sub-section 2.2.1), we propose a rule for classifying the process structure by means of the *complexity index* μ *= (total of unique traces) / (total of cases)*. We contend that a process is *spaghetti-like* when $\mu \geq 0.8$, that is, when 80% or more of the process cases do not comply with the process flow described by the *de jure* model. The chosen threshold corresponds to the dual of the one suggested by van der Aalst (2016) for the *lasagna* processes.

First analyses – *e.g.* extraction of statistical information about cases, events, and different sequences – can be conducted but, given the audit aim, it is unlikely to perform relevant actions in this step. It is advisable not to rule out all the outliers a priori because they might highlight useful clues that could be exploited downstream, especially in a *case perspective*. The outputs of this stage are

statistical information and the *de facto* model, which could likely differ from both the *de jure* model and the process blueprint collected in the stage 1.

Although a decision-maker could opt for auditing current data, we concur with van der Aalst et al. (2010) about the risk of jeopardizing the independence of the audit if it is asked him/her to recommend possible actions proactively. Therefore, in this methodology, we privileged the detection of non-conformances in historical data.

3.4 Stage 3: Model enrichment

The CF model must be enriched with further perspectives from the event log, *e.g.* time and resources, for improving the diagnostic power in detecting non-conformances (van der Aalst et al., 2010). For instance, given a process case, the exploitation of the timestamps can allow to determine the idle time of the IS between two consecutive activities, pointing out a possible violation of a time constraint.

The selection of which data the CF model should be enriched with depends on the perspective(s) chosen in stage 0. For instance, from a *process perspective*, time details could witness anomalies in the ordering of the activities, or in the execution of an activity after a closing-time. From an *organizational perspective*, data about the originators of the activities could show that an activity is executed by the wrong person, who is not in charge of it.

3.5 Stage 4: Conformance checking

This stage consists in performing a CC between the *de jure* and the *de facto* models. In table 4, we suggest three criteria in support of the algorithm selection.

Table 4. Criteria supporting the selection of the CC algorithm(s)

Criterion	Justification
Algorithm logic	Most CC algorithms require a translation of process models in Petri nets, this approach may entail false negatives and additional difficulties in exploiting the calculated metrics for further analyses (Adriansyah, van Dongen, & van der Aalst, 2011). Fuzzy-based techniques may fix these problems, but their excessively relaxed semantics can lead to further issues in conducting additional analyses on the basis of the conformance value.
Process structure	In structured processes, CC is simpler, and all the techniques may be adequate. This is because structured processes are more likely to involve repeatable, potentially automatable activities, with well-defined inputs and outputs and with a lower need for human judgment. Vice versa, the less the process is structured, the more a CC may generate broad results (van der Aalst, 2016), and the more the comparison of results from different CC techniques for seeking for convergence may be essential.
Number of activities	The less activities a process encompasses, the less paths the process could show and, thus, the less potential deviations it might include. Hence, if the set of potential deviations is small, the comparison of the results stemming from the application of different CC techniques might be less critical. A lower number of activities may also foster a manual CC, supported by the evidences from the stages 2 and 3 and from the PM analytics.

PM-enabled auditing can detect an immense amount of deviations (non-conformances) because normative models are often rigid, and CC techniques single out fine-grained deviations. Thus, it is necessary to categorize the deviations for assigning the right priority to them, *e.g.* through Multi-Criteria Decision Making, or by a classification framework (*e.g.* Hosseinpour & Jans, 2016; *cf.* Adriansyah et al., 2011; De Leoni & Marrella, 2017; Garcia-Banuelos, van Beest, Dumas, La Rosa, & Mertens, 2017). We strongly recommend involving experts about the process and the IS under analysis in formulating the criteria for estimating the severity of the non-conformances. Their better knowledge and awareness about the process dynamics can help in distinguishing which process deviations can be a concrete threat to both performance and conformance.

The output of the stage 4 is a list of the prioritized non-conformances. If all the collected evidences are needed not only for expressing to which extent the audit criteria are met, but also for providing further operational support, it is possible to proceed with two additional possibilities. First, the *de facto* model might be promoted to be the new *de jure* model. Second, it could be useful to compare the *de facto* model to the business blueprint collected in the stage 1: this could highlight that some process or security requirements were not implemented correctly in the IS (*cf.* table 3).

4. Validation of the methodology

Although PM is already well-established in literature, it is necessary to verify both applicability and effectiveness of the PM-enabled methodology for IS audit in a real context. The methodology was validated through a case in a maritime context. The main reasons for this choice are that ports are increasingly relying on innovative IT (Dong, Gang, Li, Guo, & Lv, 2013) and require more transparency in the transactions through a stronger digitalization (Meersman, van de Voorde, & Vanelslander, 2016). Hence, since port processes involve strong legal requirements, mostly related to the customs, auditing their ISs for minimizing deviations and violations is fundamental.

Nowadays, port environments are affected by a collaborative trend based on information transfer among port actors, fostered by innovative Information and Communication Technologies (ICTs) (Carlan, Sys, & Vanelslander, 2016). An answer to such trend is the increasing development of PCSs – open electronic platforms that enable a secure information exchange in seaport communities for managing and optimizing port processes through a single submission of data (IPCSA, 2014).

The validation was conducted on a dataset of the export process from a PCS of a medium-sized Mediterranean port, whose container throughput in 2016 was over 800000 Twenty-foot Equivalent Units. Because of a confidentiality agreement, further details about the port area are kept private. The export process was chosen among those supported by the PCS because of its relevance to the value creation for port customers and because, according to the port authority, it is the most data intensive. The following five sub-sections describe the application of the methodology step by step.

4.1 Justification and planning

The aim of this validation was to audit how the PCS executes the export process. After a kickoff meeting with the port authority, we chose to adopt a *process perspective* because, given the ongoing investments for integrating the PCS with another platform, the detection of possible deviations in the CF of the export process was considered more relevant than deepening specific cases (*case perspective*) or the originators of the activities (*organizational perspective*). Consequently, as suggested in sub-section 3.1, no ex-ante list of possible non-conformances was prepared.

4.2 Data extraction

Three meetings with the port authority and with the main programmer of the PCS were conducted over four months for shedding light on how the export process works, how the PCS executes it, which information sharing it involves, and which stakeholders participate in it. According to table 3, the following data and documents were collected and analyzed.

- Description of the last evolutive maintenance intervention on the PCS, including information about the last technical audit.
- Export process blueprint (figure 3), with a description of the main documents exchanged in the process (table 5), and the main stakeholders involved (table 6). The blueprint should be read from the left to the right, and it shows how the information flow, represented by the arrows, develops among the stakeholders from starting an export instance until the end of the loading of the goods on the vessel. The name of some documents and actors was slightly modified due to confidentiality reasons.
- *De jure* model of the export process executed by the PCS, in accordance with all the requirements enforced by law and by the port authority itself, drawn in BPMN (figure 4) –

see Chinosi & Trombetta (2012) for methodological details about this notation. The model stemmed from recent activities of improvement and alignment.

- The database of the PCS, with detailed information about both its structure and all the main attributes.
- Documentation about the evolution of the PCS from its introduction, future improvements, and details about how the main port stakeholders can use it.

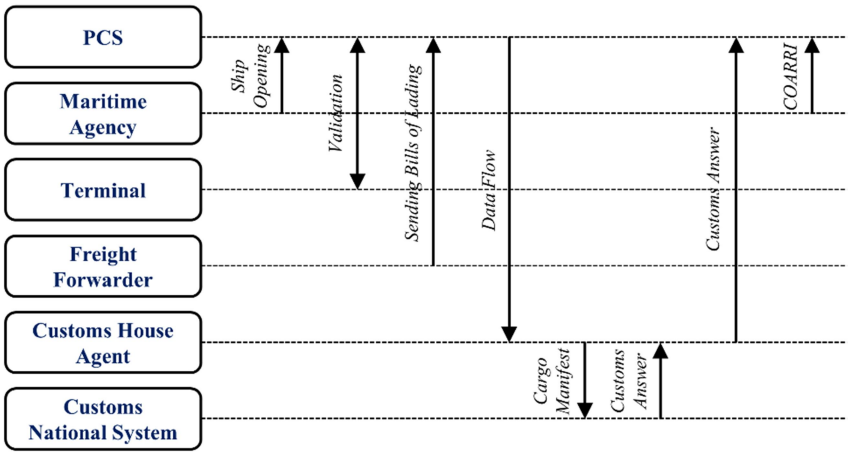

Figure 3. Blueprint of the export process

Table 5. Main documents involved in the export process

Document	Content
Bill of lading (BoL)	List of the goods that a single Freight Forwarder needs to load
Cargo Manifest (CM)	List of all the goods to be loaded on the ship
Customs Answer	Outcome of the customs check on the CM
COARRI	Message reporting which containers have been loaded

Table 6. Main stakeholders involved in the export process

Stakeholder	Description
Maritime Agencies	On behalf of the shipping companies, they manage the activities concerning Coast Guard, public security offices, and other institutional offices.
Terminals	Port logistics operators that work between the shipping companies and the carriers for material and container handling, quay management, freight storage for import and export, and other activities.
Freight Forwarders	On behalf of the customer, they organize and manage the freight forwarding, including the administrative and customs aspects.
Customs House Agents (CHAs)	Also known as Customs Forwarders, they handle the customs activities on behalf of the vessel.

According to figure 3, the export process starts with the *Ship Opening* procedure, by a Maritime Agency, that contains general data – ship name, itinerary – and other data that the Terminal should agree upon: Estimate Time to Arrival, Closing Time, and which CHA should take care of the ship operations. After the validation of *Ship Opening* by the Terminal, the Freight Forwarders send the BoLs to the PCS within the Closing Time. Once the Closing Time has come, the CHA can draw from the BoLs data for building up the CM, which is sent to the Customs National System (CNS) for a check about its completeness and correctness (*Cargo Manifest*). The CNS provides the outcome of the check – *Customs Answer* – to the CHA, which in turn sends it to the PCS: if the load

of the goods has been authorized and executed, the Terminal transmits the COARRI that ends the export process.

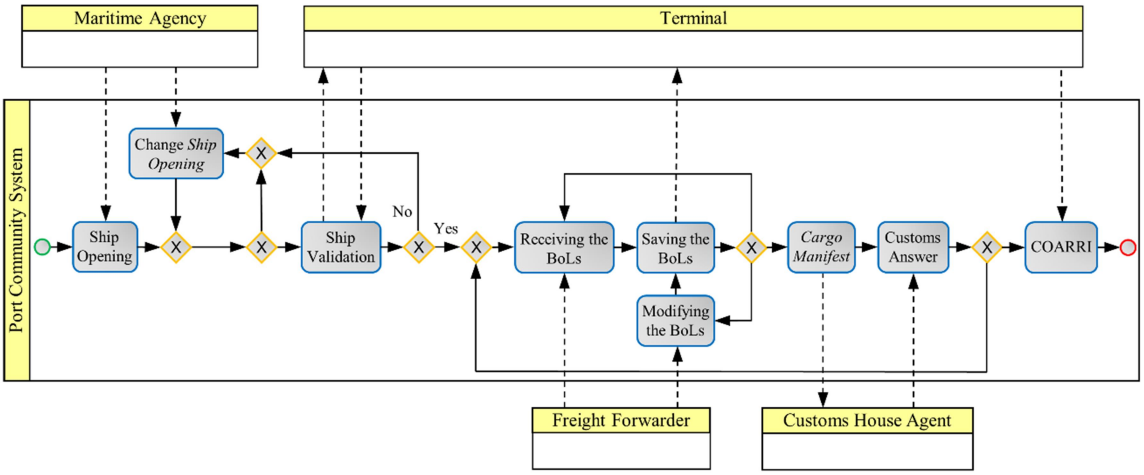

Figure 4. *De jure* model of the export process executed by the PCS

Figure 4 illustrates the *de jure* model of the export process, and it enriches figure 3 with two relevant aspects: first, data concerning *Ship Opening* can be edited before the BoLs activities; second, the elaboration of the BoLs, that we label as *BoL process*, is organized into three activities – Receiving, Saving, and Modifying the BoLs.

In accordance with the information from the port authority, the export data of the PCS were streamlined and explored critically for reconstructing the event log. Table 7 contains an anonymized fragment of the log, which is organized in five attributes: the International Maritime Organization (*IMO*) number – a unique seven-digit numerical sequence corresponding to a specific ship whose tonnage exceeds 100 tons; *Route* – a coding of the destination of the ship; the *Activity* of the export process; the *Lifecycle* of the activity – Start or Complete; and the *Timestamp*. Thus, the export process instances (the cases) were univocally identified by the couple IMO and Route, and every row of the log is an event.

Table 7. An excerpt of the event log

IMO	Route	Activity	Lifecycle	Timestamp
922xxxx	BRxxx	COARRI	Start	06/02/2016 08:49:30:513
922xxxx	BRxxx	COARRI	Complete	06/02/2016 08:49:30:810
914xxxx	ESxxx	Customs Answer	Start	06/02/2016 09:23:08:533
914xxxx	ESxxx	Customs Answer	Complete	06/02/2016 09:23:13:509
947xxxx	FRxxx	Customs Answer	Start	06/02/2016 09:38:18:919
947xxxx	FRxxx	Customs Answer	Complete	06/02/2016 09:38:35:876

4.3 Control-flow model construction

Six months of data – from 1 January 2016 to 30 June 2016 – were extracted and refined by removing incomplete events. Although the *de jure* model is not really tangled, the six-month log accounts for 843 cases that include 749 variants, resulting in a complexity index $\mu = {}^{749}/_{843} = 88.8\%$: only 11.2% of the cases follow the prescribed flow. Accordingly, as the export process is *spaghetti-like* (*cf.* sub-section 3.3), we chose the Fuzzy Miner algorithm because it "*is able to clean up a large amount of confusing behavior, and to infer and extract structure from what is chaotic*" (Günther and van der Aalst, 2007, p. 341).

The discovery activity was performed through *Disco*® 2.0.0, a PM software based on the Fuzzy Miner by Günther and van der Aalst (2007). In line with figure 4, a filter on the endpoints was

applied for considering only those cases that begin with *Ship Opening* Start and end with *COARRI* Complete, and for excluding partial cases. Hence, the final dataset consisted of 686 cases, 624 variants, 171630 events, and 18 activities, with a complexity index $\mu = {}^{624}/_{686} = 91\%$.

4.4 Model Enrichment

Disco® elaborates additional information about time and frequency stemming from the log, which is in line with the choice of a *process perspective*. Hence, the *de facto* model in output from the previous stage was an already enriched version, and the obtained details were considered as sufficient for the aim of this validation.

4.5 Conformance checking

Since the lifecycle of all the activities was found to be almost instantaneous, *Start* and *Complete* were coherently merged. Therefore, the number of the activities was reduced to 9. According to the third selection criterion proposed in table 4, we decided to perform a manual CC on the basis of the information by the analytics of Disco®. In fact, the software allows to visualize all the deviations in the model, and it details all the information concerning the variants.

All the deviations were attributable to one of more of the categories proposed by Hosseinpour and Jans (2016) – missing / repetition / replacement of a sequence; existence of an extra sequence; swapping two sequences; loop on a sequence. Two criteria were considered in prioritizing the deviations. First, we organized them into two clusters: one including the deviations whose case frequency was less than 1%, and another one for all the remaining ones. Second, we assigned a high priority to the deviations showing an inappropriate *Customs Answer* and / or *Cargo Manifest* because they might involve jurisdictional issues linked to the national customs. The deviations were critically compared and argued with the main programmer of the PCS and with a port authority employee recognized as one of the main experts in the port processes. Deviations unanimously considered as outliers, for instance those with unitary case frequency and whose impact on time performance and on the priority of some activities was considered as not relevant, were overlooked.

Six major deviations were detected. In the next section, we discuss them critically.

5. Results and discussion

Figure 5 is the *de facto* model resulting from the application of the Fuzzy Miner. The figure was adapted from the one in output from Disco® for improving its compactness and understandability. Thin arches are those paths that, according to figure 4, were already known. The six major deviations were divided in two groups: the first one – paths A and B, represented by dashed arrows – encompasses those deviations whose nature requires a further analysis for formulating a correct judgment about them; the second one – paths from 1 to 4, highlighted with bold red arches – consists of unforeseen paths that clearly are a threat to the conformance of the IS.

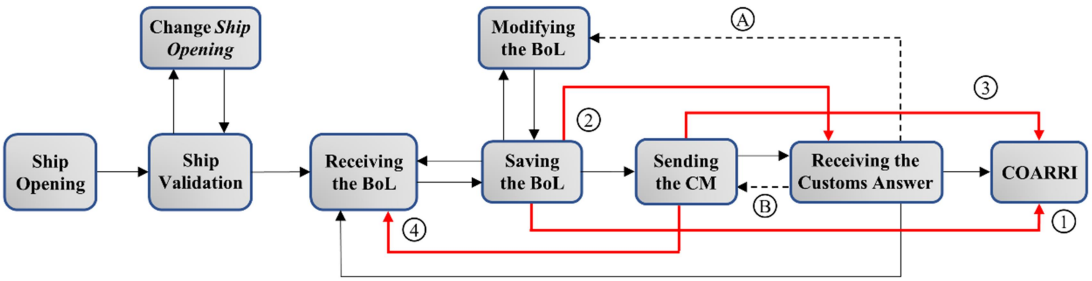

Figure 5. *De facto* model of the export process executed by the PCS

Table 8 summarizes the most relevant details of the six critical paths: their code, for fostering their identification in figure 5; start and end activities; absolute frequency; case frequency out of the 686 cases; cumulated duration; median and mean; maximum and minimum duration. In the following sub-sections, we discuss these findings.

Table 8. Details about the six major deviations

Code	From	To	Absolute Frequency	Case Frequency	Total Duration	Median	Mean	Max	Min
A	Receiving the Customs Answer	Modifying the BoL	119	104	17.8 days	15 min	3.6 hrs	48.4 hrs	3.4 s
B	Receiving the Customs Answer	Sending the CM	322	211	29.9 days	14.3 min	2.2 hrs	4 days	14.4 s
1	Saving the BoL	COARRI	54	54	17.8 wks	23.7 hrs	55.5 hrs	25.9 days	66 min
2	Saving the BoL	Receiving the Customs Answer	459	244	9 days	3.3 min	28.3 min	49.4 hrs	218 ms
3	Sending the CM	COARRI	29	29	36.1 days	25.5 hrs	29.9 hrs	4.8 days	94.7 min
4	Sending the CM	Receiving the BoL	352	214	6.2 days	2.7 min	25.3 min	45.4 hrs	796 ms

5.1 Arches A and B

As described in sub-section 4.2, if the outcome from *Receiving the Customs Answer* is negative, it is necessary to go back and fix the CM. Figure 5 points out that there are three ways of going back from *Receiving the Customs Answer* to previous activities: arch A, arch B, and an arch towards *Receiving the BoL*. A comparison among their details is in table 9.

Table 9. Details about the arches that proceed from *Receiving the Customs Answer* to previous activities

Code	To	Absolute Frequency	Case Frequency	% of case occurrence	Total Duration	Median	Mean
--	Receiving the BoL	1397	446	65%	23.2 weeks	16.2 min	2.8 hrs
A	Modifying the BoL	119	104	15.2%	17.8 days	15 min	3.6 hrs
B	Sending the CM	322	211	30.7%	29.9 days	14.3 min	2.2 hrs

According to the *de jure* model (figure 4), if the customs check on the CM is negative, it is necessary to modify it starting from *Receiving the BoL*, as the CM is built up drawing from the BoLs incrementally. This occurrence is widely acknowledged by the port authority, and table 8 shows that it is quite frequent – 65% of the cases. Instead, the port authority was not aware about the possibility to correct the CM directly (arch B), or through *Modifying the BoL* (arch A). The programmer of the PCS stated that the standard process flow requires a feedback to *Receiving the BoL* only, but that the two other alternatives were consciously enabled and kept for offering more flexibility in the procedure through the additional paths. Their occurrence is not negligible – 104 for the arch A, 211 for the arch B – and their median time is not far from that of the standard flow (table 9), and this witnesses that these two deviations are quite exploited by the pertinent stakeholders connected to the PCS. In line with such evidences, the two deviations were considered as desirable because they enhance the flexibility of the process without jeopardizing the conformance in a severe way.

5.2 Arch 1 – from Saving the BoL to COARRI

This path occurs in almost 8% of the cases and bypasses both *Sending the CM* and *Receiving the Customs Answer* (figure 5). Nine cases out of the 54 (1.3% of the total) skip the two aforementioned activities bluntly, while the remaining 45 cases (6.5% of the total) follow the standard flow until the customs answer, they go back for fixing some mistakes in the BoLs, and then they proceed from *Saving the BoL* to *COARRI* directly. In other words, there are two different process deviations: on one side, in 9 export instances, the CM was neither built and sent nor checked by the customs, and the goods were allegedly stowed in the hold of the ship. On the other side, in 45 process instances

the BoLs were modified according to the customs answer, but the CM was not updated and was not checked again and, even so, the containers were loaded on the ship.

Both the port authority and the programmer clearly stated that creation / update and sending of the CM to the customs for the check are mandatory activities, and that proceeding with *COARRI* directly should be impossible and not allowed. Therefore, it could suggest that a relevant number of events between the BoL process and the COARRI are performed outside of the PCS and that their data are not submitted to the platform.

5.3 Arch 2 – from Saving the BoL to Receiving the Customs Answer

About 36% of the cases include this straight arch, but its interpretation can be deceptive: actually, the 244 process instances do not skip *Sending the CM* totally. Differently, within the loop between *Receiving the BoL* and *Receiving the Customs Answer*, which can occur several times per case, they skip the submission of the updated CM at least one time but they send at least one CM.

This means that, sometimes, a case could proceed with the loading on the vessel on the basis of an obsolete manifest, and that the customs might not be informed about it. A possible reason for such an issue might be that the data in the PCS are correct and the process flow is fine, but the system allows the users to access the BoL process even when the final customs check is positive: consequently, after this inappropriate and unnecessary feedback to *Saving the BoL*, the users try to follow the standard process flow by skipping to the last activity before the physical loading. Accordingly, activities involved in this process deviation might be dummy, without any substantial modifications in the data.

5.4 Arch 3 – from Sending the CM to COARRI

This arch shows a case frequency (4.2%) which is lower than that of the other criticalities we singled out, but it was considered as relevant because it involves issues concerning customs activities. Similarly to arch 1, 22 process instances iterate the flow between *Receiving the Customs Answer* and *Receiving the BoL* correctly but, in the last iteration, they jump from the CM to *COARRI*. Possible reasons for such behavior may be akin to the ones provided for arch 2. The seven remaining instances do not perform any customs check: given their low case frequency, they could be considered as outliers.

5.5 Arch 4 – from Sending the CM to Receiving the BoL

Almost one case out of three encompasses this arch. From a process perspective, it points out that, in 214 process instances, the BoLs related to a CM were enriched / modified after the manifest was sent to the customs but before receiving the outcome of the customs check. Nevertheless, all the cases present a CM, even though 22 of them display a feedback to the BoL process without updating the manifest and / or checking for its correctness. Thus, the PCS allows to update the BoLs before receiving the answer from the customs about the CM.

The total duration of this arch is quite low, but its case frequency is so high (214) that its occurrence cannot be due to process instance overlays or extemporaneous errors. A plausible reason could be that, after sending the CM, the users realize that there are some mistakes in the BoLs, and they try to fix them as soon as possible. In accordance with the port authority, it should not be advisable to hasten such updates in the BoLs in this way before receiving the answer from the customs.

6. Discussion about the methodology

The findings from the validation also offer the possibility to discuss our methodology in terms of advantages and limitations (sub-section 6.1) and practical implications (sub-section 6.2).

6.1 Evaluation of the methodology

The methodology exploits PM as an ES engine for analyzing the log-based knowledge base of ISs and, in doing so, it presents some advantages when compared to extant methods for off-line IS audits. Table 10 highlights the main strengths of the methodology compared to the approaches we

reported in tables 1 and 2. Such strengths are mostly linked to the possibility to analyze whole IS logs without any sampling need, in a quasi-automatic, low-invasive way.

Table 10. The advantages of our PM-enabled methodology over other approaches to IS off-line audit

Compared to		Advantages
CAATTs	ITF	No need for advanced programming skills; no dummy transactions; possibility to simulate a process instance;
	Parallel simulation	No need to duplicate the IS;
	EAM	Less invasive;
	GAS	Greater depth of detection of the IS non-conformances;
ESs	Tsudik & Summers (1990)	Wider and deeper overview of the IS actual behavior;
	Sanchez and Rodriguez (1994)	Combination of both qualitative and quantitative approaches;
	Akoka & Comyn-Wattiau (1996); Comyn-Wattiau & Akoka (1996);	More objective, quantitative evidences about the IS instances;
	Atymtayeva, Bortsova, Inoue, & Kozhakhmet (2012); Kanatov, Atymtayeva, & Yagaliyeva (2014); Piech & Grodzki (2017);	Wider scope; greater level of detail;

Our methodology presents some drawbacks too. First, the *complexity index* equalizes all the process variants without evaluating their severity. Some process deviations can be desirable for enhancing the flexibility of a process model in case of unforeseen situations (van der Aalst, 2016), but the complexity index evens all the deviations out, without discerning the desirable from the undesirable ones. Thus, although the PM algorithm selection according to the complexity index led to a clear process map, it is advisable to have a sound knowledge about the process and IS dynamics before analyzing the log.

Second, the exclusion of potential outliers from the log (see sub-section 3.3) should be conducted with a high level of awareness. The PM-enabled deepening of wide event logs yields fast, fine-grained results, and this could imply a higher possibility that an outlier is a major non-conformance to tackle, and not to overlook.

Third, the capability to conduct on-line IS audits through PM is not mature yet. Although this scenario is out of the scope of our methodology, EAMs are still one of the best solutions for near-real-time auditing.

Fourth, a certain, reasonably-low amount of manual work is still needed because of the specificities that the context of analysis could present. Even though this consideration is in line with Tawakkal et al. (2017) about the fundamental, manual work in PM-enabled IS audits, it questions the possibility (or the usefulness) of developing a PM-based ES whose rules are general for all the IS audits.

6.2 Implications based on the experimental outcomes

The PM-enabled methodology showed its usefulness in conducting both substantive and control tests. In particular, it directly tackles the IT application controls, with a specific focus on the processing controls, which address completeness and accuracy of the processing, and on the output controls, which compare the actual results of a process to the intended ones.

In the case we analyzed, most issues in the *Sending the CM* or in the *Customs Answer* activities may be related to a cumbersome BoL process. Moreover, according to the port authority, a great part of the deviations might be due to human errors in filling the BoLs, which entail revising the CM – more than one time, in some cases. The issues in sending and receiving information from the Customs may imply a huge gap between which goods are declared before the loading and which

goods are actually exported: this criticality is the final result which all the four major non-conformances contribute to.

On the basis of our findings, the port authority could intervene by introducing some boundaries in the PCS, reducing the potential occurrence of the detected non-conformances without jeopardizing the needed flexibility. First, in a given process instance the PCS should not allow any modification to the BoL activities if the Customs Check is positive. A positive outcome authorizes the physical loading of the goods, and any subsequent modification in the CM would necessarily require a new check. This solution would reduce the probability of a gap between what is actually loaded on the vessel and what the Customs Check authorizes to load. In addition, it could be cheaper than introducing an additional internal control.

Second, the PCS should not allow any feedback to the BoL process after the submission of the CM to the Customs, but before receiving the output of the Customs Check. In this case, any update in the BoLs would entail the potential creation of additional CMs that would now be checked and that could create overlaps among different process cases.

Third, the PCS should force the user to update the CM, before its submission, after any change to the pertinent BoLs, in order to align the document flow to the physical, subsequent one. The port authority should consider the decision to develop an application control for this purpose.

Fourth, it could be advisable to reduce the data entry error, *e.g.* by replacing the visual checking with a more effective double data entry (*cf.* Barchard & Pace, 2011).

Thus, the methodology is a flexible tool that enables further kinds of IS operational support, such as paving the way for a possible integration of the methodology within the IT general controls and standard practices of a firm. In this way, decision makers could systematically audit their ISs, streamlining the workload of the auditors, making more effective decisions concerning the management of both business processes and the ISs supporting them, and underpinning the IS risk management process with advanced mining approaches.

7. Conclusions and future work

Normative advancements and the digitalization trend have called for more transparency, enhanced IS risk management approaches, and more advanced IS audit systems. Yet, current ESs and IT solutions for IS audits suffer from some limitations – such as sampling or qualitative approach, lack of automatic tools, narrow scope – that could hinder a sharper identification of non-conformances. Scientific literature has suggested that PM could cope with such weaknesses, but contributions about PM-based auditing are almost absent. Thus, by framing PM as an ES engine, we developed a five-step PM-enabled methodology for IS audit, and we validated it in a freight export port process supported by a PCS. The methodology showed some advantages, compared to other IT audit approaches, in terms of level of detail, higher and easier possibility to automate, wider scope, and lesser invasiveness. It allowed to single out six major process deviations. Two of them were classified as desirable because they enhance the flexibility of the export process without compromising both its conformance and its operational execution. The other four were categorized as a relevant threat from a managerial, operational, and legal perspective. The audit outcome paved the way to four solutions for revising the export process within the PCS in order to limit the occurrence of the identified issues.

Our work opens interesting avenues for future research. First, the role of PM as an ES engine could be deepened by expanding the methodology to on-line IS audit – the most desirable audit form – comparing our PM-based approach to the EAM one in on-line settings. Second, further quantitative metrics of process structuredness (*cf.* Laue & Mendling, 2010) in the PM context could be useful to a more accurate selection of the best PM algorithm for process discovery and CC – a research topic that, currently, has not been tackled enough. Third, to contextualize and to improve the ranking of the non-conformances by considering the specificities of the context which the audit is conducted in, the IS audits could benefit from linking the definition of the classification criteria to the different

PM perspective(s) chosen in stage 0 of the methodology. Different PM perspective(s) imply different nuances in the purposes of the PM application to IS audits. This relationship could be automatized within a specific plug-in, *e.g.* in a ProM environment.

References

Accorsi, R., & Stocker, T. (2012). On the Exploitation of Process Mining for Security Audits: The Conformance Checking Case. *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, 1709–1716. https://doi.org/10.1145/2480362.2480634

ACFE. (2010). *Report to the Nations on Occupational Fraud and Abuse*. Retrieved from http://www.acfe.com/uploadedFiles/ACFE_Website/Content/documents/rttn-2010.pdf

Adriansyah, A., van Dongen, B. F., & van der Aalst, W. M. P. (2011). Towards robust conformance checking. In *Lecture Notes in Business Information Processing* (Vol. 66 LNBIP, pp. 122–133). https://doi.org/10.1007/978-3-642-20511-8_11

Ahmi, A., & Kent, S. (2012). The utilisation of generalized audit software (GAS) by external auditors. *Managerial Auditing Journal*, *28*(2), 88–113. https://doi.org/10.1108/02686901311284522

Akoka, J., & Comyn-Wattiau, I. (1996). A Knowledge-Based System for Auditing Computer and Management Information Systems. *Expert Systems with Applications*, *11*(3), 361–375. https://doi.org/10.1016/S0957-4174(96)00051-6

Atymtayeva, L. B., Bortsova, G. K., Inoue, A., & Kozhakhmet, K. T. (2012). Methodology and ontology of expert system for information security audit. In *6th International Conference on Soft Computing and Intelligent Systems, and 13th International Symposium on Advanced Intelligence Systems, SCIS/ISIS 2012* (pp. 238–243). https://doi.org/10.1109/SCIS-ISIS.2012.6505287

Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. In *Computers in Human Behavior* (Vol. 27, pp. 1834–1839). https://doi.org/10.1016/j.chb.2011.04.004

Bellino, C., & Hunt, S. (2007). Global Technology Audit Guide (GTAG) 8: Auditing Application Controls. The Institute of Internal Auditors.

Braun, R. L., & Davis, H. E. (2003). Computer-assisted audit tools and techniques: analysis and perspectives. *Managerial Auditing Journal*, *18*(9), 725–731. https://doi.org/10.1108/02686900310500488

Burattin, A., Maggi, F. M., & Sperduti, A. (2016). Conformance checking based on multi-perspective declarative process models. *Expert Systems with Applications*, *65*, 194–211. https://doi.org/10.1016/j.eswa.2016.08.040

Carlan, V., Sys, C., & Vanelslander, T. (2016). How port community systems can contribute to port competitiveness: Developing a cost-benefit framework. *Research in Transportation Business and Management*, *19*, 51–64. https://doi.org/10.1016/j.rtbm.2016.03.009

Carlin, A., & Gallegos, F. (2007). IT audit: A critical business process. *Computer*, *40*(7), 87–89. https://doi.org/10.1109/MC.2007.246

Chinosi, M., & Trombetta, A. (2012). BPMN: An introduction to the standard. *Computer Standards and Interfaces*, *34*(1), 124–134. https://doi.org/10.1016/j.csi.2011.06.002

Coderre, D. (2015). *Internal Audit: Efficiency Through Automation*. *Internal Audit: Efficiency Through Automation*. Wiley Blackwell. https://doi.org/10.1002/9781119203544

Comyn-Wattiau, I., & Akoka, J. (1996). Logistics information system auditing using expert system technology. *Expert Systems with Applications*, *11*(4), 463–473. https://doi.org/10.1016/S0957-4174(96)00062-0

De Leoni, M., & Marrella, A. (2017). Aligning Real Process Executions and Prescriptive Process Models through Automated Planning. *Expert Systems with Applications*, *82*, 162–183. https://doi.org/10.1016/j.eswa.2017.03.047

De Medeiros, A. K. A., Weijters, A. J. M. M., & van der Aalst, W. M. P. (2005). Genetic process mining: A basic approach and its challenges. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3812 LNCS*, 203–215.

De Weerdt, J., De Backer, M., Vanthienen, J., & Baesens, B. (2012). A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Information Systems*, *37*(7), 654–676. https://doi.org/10.1016/j.is.2012.02.004

Debreceny, R., Lee, S.-L., Neo, W., & Shuling Toh, J. (2005). Employing generalized audit software in the financial services sector: Challenges and opportunities. *Managerial Auditing Journal*, *20*(6), 605–618. https://doi.org/10.1108/02686900510606092

Dong, X., Gang, X., Li, Y., Guo, X., & Lv, Y. (2013). Intelligent ports based on Internet of Things. In *Proceedings of 2013 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2013* (pp. 292–296). https://doi.org/10.1109/SOLI.2013.6611428

Dyer, J. S. (1990). Remarks on the Analytic Hierarchy Process. *Management Science*, *36*(3), 249–258. https://doi.org/10.1287/mnsc.36.3.249

Erasmus, L., & Coetzee, P. (2018). Drivers of stakeholders' view of internal audit effectiveness: Management versus audit committee. *Managerial Auditing Journal*, *33*(1), 90–114. https://doi.org/10.1108/MAJ-05-2017-1558

Garcia-Banuelos, L., van Beest, N., Dumas, M., La Rosa, M., & Mertens, W. (2017). Complete and Interpretable Conformance Checking of Business Processes. *IEEE Transactions on Software Engineering*, (99), 1–28. https://doi.org/10.1109/TSE.2017.2668418

Giarratano, J. C., & Riley, G. D. (2005). *Expert Systems: Principles and Programming* (4th ed.). Thomson Course Technology.

Gibson, D. (2014). *Managing risk in information systems*. Jones and Bartlett.

Goodhue, D. L., & Thompson, R. L. (1995). Task-Technology Fit and Individual Performance. *MIS Quarterly*, *19*(2), 213–236. https://doi.org/10.2307/249689

Günther, C. W., & van der Aalst, W. M. P. (2007). Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. In *Business Process Management - Lecture Notes in Computer Science* (Vol. 4714, pp. 328–343). https://doi.org/10.1007/978-3-540-75183-0

Hosseinpour, M., & Jans, M. (2016). Categorizing identified deviations for auditing. In *CEUR Workshop Proceedings* (Vol. 1757, pp. 125–129).

Huang, S.-M., Yen, D. C., Hung, Y.-C., Zhou, Y.-J., & Hua, J.-S. (2009). A business process gap detecting mechanism between information system process flow and internal control flow. *Decision Support Systems*, *47*(4), 436–454. https://doi.org/10.1016/j.dss.2009.04.011

Institute of Internal Auditors. (2014). *Enhancing value through collaboration: a call to action*. Retrieved from https://dl.theiia.org/AECPublic/2014-Global-Pulse-of-the-ProfessionReport-Enhancing-Value-Through-Collaboration-A-Call-to-Action.pdf

IPCSA. (2014). Port Community Systems. Retrieved April 20, 2017, from http://www.ipcsa.international/pcs

Issa, H., Sun, T., & Vasarhelyi, M. A. (2016). Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation. *Journal of Emerging Technologies in Accounting*, *13*(2), 1–20. https://doi.org/10.2308/jeta-10511

Jans, M., Alles, M. G., & Vasarhelyi, M. A. (2014). A field study on the use of process mining of event logs as an analytical procedure in auditing. *Accounting Review*, *89*(5), 1751–1773. https://doi.org/10.2308/accr-50807

Jans, M., van der Werf, J. M., Lybaert, N., & Vanhoof, K. (2011). A business process mining application for internal transaction fraud mitigation. *Expert Systems with Applications*, *38*(10), 13351–13359. https://doi.org/10.1016/j.eswa.2011.04.159

Kamsu-Foguem, B., Rigal, F., & Mauget, F. (2013). Mining association rules for the quality improvement of the production process. *Expert Systems with Applications, 40*(4), 1034-1045. https://doi.org/10.1016/j.eswa.2012.08.039

Kanatov, M., Atymtayeva, L., & Yagaliyeva, B. (2014). Expert systems for information security management and audit. Implementation phase issues. In *2014 Joint 7th International Conference on Soft Computing and Intelligent Systems, SCIS 2014 and 15th International Symposium on Advanced Intelligent Systems, ISIS 2014* (pp. 896–900). https://doi.org/10.1109/SCIS-ISIS.2014.7044702

Lang, M., Bürkle, T., Laumann, S., & Prokosch, H.-U. (2008). Process mining for clinical workflows: challenges and current limitations. *Studies in Health Technology and Informatics*, *136*, 229–234. https://doi.org/10.1007/978-3-642-19345-3

Laue, R., & Mendling, J. (2010). Structuredness and its significance for correctness of process models. *Information Systems and E-Business Management*, *8*(3), 287–307. https://doi.org/10.1007/s10257-009-0120-x

Lee, J. K., & Jeong, M. W. (1995). Intelligent audit planning system for multiple auditors: IAPS. *Expert Systems With Applications*, *9*(4), 579–589. https://doi.org/10.1016/0957-4174(95)00026-7

Liao, S. H. (2005). Expert system methodologies and applications-a decade review from 1995 to 2004. *Expert Systems with Applications*, *28*(1), 93–103. https://doi.org/10.1016/j.eswa.2004.08.003

Lombardi, D. R., & Dull, R. B. (2016). The Development of AudEx: An Audit Data Assessment System. *Journal of Emerging Technologies in Accounting*, *13*(1), 37–52. https://doi.org/10.2308/jeta-51445

Meersman, H., van de Voorde, E., & Vanelslander, T. (2016). Port competitiveness now and in the future: What are the issues and challenges? *Research in Transportation Business and Management*, *19*, 1–3. https://doi.org/10.1016/j.rtbm.2016.05.005

Mock, R., & Corvo, M. (2005). Risk analysis of information systems by event process chains. *International Journal of Critical Infrastructures*, *1*(2/3), 247–257. https://doi.org/10.1504/IJCIS.2005.006121

Omoteso, K. (2012). The application of artificial intelligence in auditing: Looking back to the future. *Expert Systems with Applications*, *39*(9), 8490–8495. https://doi.org/10.1016/j.eswa.2012.01.098

Pedrosa, I., & Costa, C. J. (2012). Financial auditing and surveys: how are financial auditors using information technology? An approach using Expert Interviews. In *Proceedings of the Workshop on Information Systems and Design of Communication - ISDOC '12* (pp. 37–43). Lisbon, Portugal. https://doi.org/10.1145/2311917.2311925

Pedrosa, I., & Costa, C. J. (2014). New trends on CAATTs: what are the Chartered Accountants' new challenges? In *ISDOC '14 Proceedings of the International Conference on Information Systems and Design of Communication, May 16–17* (pp. 138–142). Lisbon, Portugal. https://doi.org/10.1145/2618168.2618190

Piech, H., & Grodzki, G. (2017). Audit expert system of communication security assessment. *Procedia Computer Science*, *112*, 147–156. https://doi.org/10.1016/j.procs.2017.08.188

Rezaee, Z., Elam, R., & Sharbatoghlie, A. (2001). Continuous auditing: the audit of the future. *Managerial Auditing Journal*, *16*(3), 150–158. https://doi.org/10.1108/02686900110385605

Richard Ye, L. (1995). The Value of Explanation in Expert Systems for Auditing: An Experimental

Investigation. *Expert Systems with Applications*, *9*(4), 543–556. https://doi.org/10.1016/0957-4174(95)00023-2

Rozinat, A., & van der Aalst, W. M. P. (2008). Conformance checking of processes based on monitoring real behavior. *Information Systems*, *33*(1), 64–95. https://doi.org/10.1016/j.is.2007.07.001

Ruiz, P. P., Kamsu-Foguem, B., & Grabot, B. (2014). Generating knowledge in maintenance from Experience Feedback. *Knowledge-Based Systems, 68*, 4-20. http://doi.org/10.1016/j.knosys.2014.02.002

Sanchez, A., & Rodriguez, P. (1994). Zyanya: EDP Auditing and Expert Systems. In J. Liebowitz (Ed.), *Moving Toward Expert System Globally in the 21st Century*. New York: Cognizant Communication Corporation.

Sayana, A. S. (2003). Using CAATs tu Support IS Audit. Information Systems Audit and Control Association. Retrieved from https://csbweb01.uncw.edu/people/ivancevichd/classes/MSA 516/Extra Readings on Topics/CAATS/Using CAATTS to Support IT Audit.pdf

Sutton, S. G., Holt, M., & Arnold, V. (2016). "The reports of my death are greatly exaggerated"—Artificial intelligence research in accounting. *International Journal of Accounting Information Systems*, *22*, 60–73. https://doi.org/10.1016/j.accinf.2016.07.005

Tawakkal, I., Kurniati, A. P., & Wisudiawan, G. A. A. (2017). Implementing heuristic miner for information system audit based on DSS01 COBIT5 (Case study: CV Narnia distribution). In *Proceeding - 2016 International Conference on Computer, Control, Informatics and its Applications: Recent Progress in Computer, Control, and Informatics for Data Science, IC3INA 2016* (pp. 197–202). https://doi.org/10.1109/IC3INA.2016.7863049

Tsudik, G., & Summers, R. (1990). AudES-An Expert System for Security Auditing. In *Proceedings of the AAAI Conference on Innovative Application in Artificial Intelligence* (pp. 221–232). Retrieved from http://www.aaai.org/Papers/IAAI/1990/IAAI90-017.pdf

van der Aalst, W. M. P. (2016). *Process mining: Data science in action*. *Process Mining: Data Science in Action* (2nd ed.). Springer. https://doi.org/10.1007/978-3-662-49851-4

van der Aalst, W. M. P., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., … Wynn, M. (2012). Process mining manifesto. In *Lecture Notes in Business Information Processing* (Vol. 99 LNBIP, pp. 169–194). https://doi.org/10.1007/978-3-642-28108-2_19

van der Aalst, W. M. P., van Hee, K. M., van der Werf, J. M., & Verdonk, M. (2010). Auditing 2.0: Using process mining to support tomorrow's auditor. *Computer*, *43*(3), 90–93. https://doi.org/10.1109/MC.2010.61

Wang, Y., Caron, F., Vanthienen, J., Huang, L., & Guo, Y. (2014). Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port. *Expert Systems with Applications*, *41*(1), 195–209. https://doi.org/10.1016/j.eswa.2013.07.021

Weber, B., Reichert, M., & Rinderle-Ma, S. (2008). Change patterns and change support features - Enhancing flexibility in process-aware information systems. *Data and Knowledge Engineering*, *66*(3), 438–466. https://doi.org/10.1016/j.datak.2008.05.001

Weijters, A. J. M. M., van der Aalst, W. M. P., & De Medeiros, A. K. A. (2006). Process Mining with the Heuristics Miner Algorithm. *Technische Universiteit Eindhoven, Tech. Rep. WP*, *166*, 1–34. https://doi.org/10.1.1.118.8288

Werner, M. (2017). Financial process mining - Accounting data structure dependent control flow inference. *International Journal of Accounting Information Systems*, *25*, 57–80. https://doi.org/10.1016/j.accinf.2017.03.004

Werner, M., & Gehrke, N. (2015). Multilevel Process Mining for Financial Audits. *IEEE Transactions on Services Computing*, *8*(6), 820–832. https://doi.org/10.1109/TSC.2015.2457907

Wright, C. (2008). *The IT Regulatory and Standards Compliance Handbook: How to Survive Information Systems Audit and Assessments*. https://doi.org/10.1016/B978-1-59749-266-9.X0001-X

Yang, W. S., & Hwang, S. Y. (2006). A process-mining framework for the detection of healthcare fraud and abuse. *Expert Systems with Applications*, *31*(1), 56–68. https://doi.org/10.1016/j.eswa.2005.09.003

Zhao, N., Yen, D. C., & Chang, I. (2004). Auditing in the e-commerce era. *Information Management & Computer Security*, *12*(5), 389–400. https://doi.org/10.1108/09685220410563360