Preparation Instructions

Case „*Big Data and Audit*“

1. Signup

Please sign up for the academic version of Celonis using the following link:

https://www.celonis.com/academic-signup

You will receive access to your own Celonis Intelligent Business Cloud (IBC) workspace, yet, in order to gain access to the data for our case study, it is necessary for you to be added to our workspace. Therefore, please follow steps two and three.

2. Invitation

We kindly ask you to send an e-mail containing your Celonis e-mail with the reference “Celonis STAMP-Online Big Data and Audit” to be provided access to the Celonis workspace to the following e-mail-address: maico.schoene@fau.de

You will receive an invitation to the Celonis team “maico-schoene”.

3. Login

Please log in at https://academic-maico-schoene-fau-de.eu-2.celonis.cloud/ using the credentials you signed up with. Further instructions on how to use Celonis will be provided on Thursday, November 25.

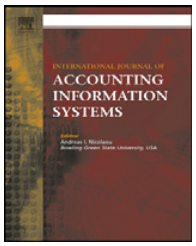
4. Prepare (Optional)

To prepare for the case study you can read the paper “*The case for process mining in auditing: Sources of value added and areas of application*” by Mieke Jan, Michael Alles, and Miklos Vasarhelyi published in the International Journal of Accounting Information Systems.

International Journal of Accounting Information Systems 14 (2013) 1–20

Contents lists available at SciVerse ScienceDirect

International Journal of Accounting Information Systems



The case for process mining in auditing: Sources of value added and areas of application

Mieke Jans [a], Michael Alles [b,*], Miklos Vasarhelyi [b]

[a] Hasselt University, Belgium
[b] Rutgers Business School, Newark, NJ, USA

ARTICLE INFO

Article history:
Received 19 September 2010
Received in revised form 12 June 2012
Accepted 26 June 2012

Keywords:
Event logs
Process mining
Internal auditing
Continuous auditing

ABSTRACT

Process mining aims to extract knowledge from the event logs maintained by a company's ERP system. The objective of this paper is to make the case for why internal and external auditors should leverage the capabilities process mining offers to rethink how auditing is carried out. We do so by identifying the sources of value added of process mining when applied to auditing, which are as follows: 1. process mining analyzes the entire population of data and not just a sample; 2. critically that data consists of meta-data—data entered independently of the actions of auditee—and not just data entered by the auditee; 3. process mining allows the auditor to have a more effective way of implementing the audit risk model by providing effective ways of conducting the required walkthroughs of processes and conducting analytic procedures; 4. process mining allows the auditor to conduct analyses not possible with existing audit tools, such as discovering the ways in which business processes are actually being carried out in practice, and to identify social relationships between individuals. It is our argument that these sources of value have not been fully understood in the process mining literature, which has focused on developing it as a statistical methodology rather than on applying it to audit practice. Only when auditors and audit researchers appreciate what is new and unique about process mining will its acceptance in auditing practice become feasible.

* Corresponding author. Tel.: +1 9733535352.
E-mail addresses: Mieke.jans@uhasselt.be (M. Jans), alles@business.rutgers.edu (M. Alles), miklosv@andromeda.rutgers.edu (M. Vasarhelyi)

1467-0895/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.accinf.2012.06.015

1. Introduction

Process mining is a new and highly promising means of systematically analyzing data recorded by a business's Enterprise Resource Planning (ERP) system. With ERPs now pervasive in large as well as many midsize businesses, process mining offers a way of exploiting the vast amount of data they routinely gather and store in ways that lead to unique insights into how processes are being undertaken in those businesses.

The underlying focus of process mining is a business process, which is a "defined set of business activities that represent the steps required to achieve a business objective".[1] The identification and analysis of processes are central to such modern business practices as activity based management, business process reengineering and business intelligence. Since actual business processes are extremely complex, with numerous interactions across different processes taking place either concurrently or with various lags, they often bear no relation to the ideal as envisaged by the process designer (Kogan et al., 2010; van der Aalst, 2010b). Process mining enables a comparison to be made between how processes take place in practice versus the way they are meant to operate as designed, which is what makes process mining potentially of such value in auditing.

In particular, process mining enables an internal or external auditor to understand the unintended consequences of relaxing ERP control settings in order to accommodate such unforeseen eventualities as expediting orders for valued customers or dealing with rush orders. Allowing process owners to have the flexibility to alter control settings is unavoidable if the business is to run effectively in the face of uncertainty, but it means that auditors cannot rely solely on the integrity of those controls and must also conduct tests of detail. Process mining gives auditors a new and more comprehensive way of conducting those tests of detail and of understanding the state of the control environment than the procedures that they rely on today.

To illustrate, consider the example—drawn from the personal experience of the authors—of a telecommunications company that was forced to lay off a large number of its managers whose responsibilities include authorization of transactions. To satisfy segregation of duty controls, these responsibilities are routinely spread among various managers, but after the redundancies took place the absence of designated signees meant that those who remained instituted ad-hoc work-around arrangements without adequate documentation. The company's internal auditors faced great difficulty in reestablishing adequate controls as a result. In this situation process mining could have been used to determine what the new arrangements were after the layoffs by discovering how transactions are actually now being authorized and processed. It could potentially also have been used before the event to determine which managers were most critical in the authorization process to better prioritize the redundancies. The key both ex-ante and ex-post is to determine how authorizations are actually being carried out as opposed to relying on the theoretical SOD process which typically has been deviated from since its inception to take into account changes in personnel, the addition of new procedures and vendors and so forth.

Process mining was originally developed by computer scientists and there is a large body of literature in that field as well as in engineering and management (Schimm, 2003; van der Aalst and Weijters, 2004; Rozinat et al., 2007; Lijie et al., 2009; van der Aalst, 2010a). Process mining has been used, for example, to understand the way in which patients are treated in hospitals (Mans et al., 2010), while Wynn et al. (2009) use process mining to develop a method to efficiently and thoroughly recall unsafe products from a supply chain. Van der Aalst and de Medeiros (2005) develop process mining algorithms for network security and demonstrate how their algorithm identifies anomalous trails and enables the detection of the point where the security breach took place. van der Aalst et al. (2007) analyze the process of handling invoices in a provincial office of the Dutch national public works department. A listing of over one hundred process mining papers written over the last decade is maintained at the website of the Business Process Mining Center.[2]

[1] www.modernanalyst.com/Resources/Articles/tabid/115/articleType/ArticleView/articleId/936/More-Confusing-SOA-Terms.aspx.

[2] http://bpmcenter.org/reports.

In contrast to this intensive activity in other academic disciplines, there have been only a handful of papers in accounting that have discussed process mining, and moreover, they have been essentially technical in character, focusing more on the methodology of process mining than on its specific application to accounting. Thus, Jans et al. (2009) explored using process mining for fraud detection, while Gehrke and Mueller-Wickop (2010) developed an algorithm for creating and analyzing event logs in SAP™. The only more specific call to use process mining in auditing was provided by van der Aalst et al. (2010) in which the authors argue that "Auditing 2.0—a more rigorous form of auditing that couples detailed event logs with process mining techniques—will dramatically change the auditing profession." However, this paper—only three pages long, and published in a trade journal—is more focused on discussing what is new in process mining as an analytic tool rather than what is new in process mining as an audit practice.[3]

It is our contention, though, if auditors are to seriously consider adopting process mining that is precisely what will be needed to make a compelling case that process mining represents a new way of doing auditing and not just a new way of doing statistical analysis. Simply offering it as yet another option in the audit toolkit will not by itself persuade auditors to adopt process mining given that there are very real barriers to its use, with most auditors today not possessing the skill sets necessary to implement process mining. Moreover, to enable process mining to move from beyond an academic curiosity to actual application in internal and external auditing it has to be placed within the context of contemporary audit practice so that auditors have the assurance that its use meets professional standards. In other words, given that using process mining is costly in terms of effort and skill acquisition it has to be seen as reducing the auditor workload elsewhere. This requires process mining to replace existing audit practice rather than adding to it—and doing so while still keeping the engagement compliant with audit standards.

The object of this paper is to make the case for why auditors should leverage the capabilities process mining offers to rethink how auditing is carried out. We do so by identifying the sources of value added of process mining when applied to auditing, which are:

1. Process mining analyzes the entire population of data and not just a sample.
2. Critically that data consists of meta-data—data entered independently of the actions of auditee—and not just data entered by the auditee.
3. Process mining allows the auditor to have a more effective way of implementing the audit risk model by providing effective ways of conducting the required walkthroughs of processes and conducting analytic procedures.
4. Process mining allows the auditor to conduct analyses not possible with existing audit tools, such as discovering the ways in which business processes are actually being carried out in practice, and to identify social relationships between individuals.

It is our argument that these sources of value have not been fully understood in the process mining literature, which has focused on developing it as a statistical methodology rather than on applying it to audit practice. Only when auditors and audit researchers appreciate what is new and unique about process mining will its acceptance in auditing practice become feasible.

In the next section of the paper we discuss the first two sources of the value added that process mining brings to auditing. In particular, we discuss the use of the population of data and the role of meta-data in the event log and provide examples of how process mining of meta-data can be used by auditors to achieve better process understanding and control. Section 3 of the paper examines the integration of

[3] Another related paper is Alles et al. (2004). They advocated the creation of "a 'black box (BB) log file' that is a read-only, third-party-controlled record of the actions of auditors, especially in regard to their interactions with management and choice of audit metric and models. Comprehensive and secure in a way that the current system of working papers is not, and accompanied by sophisticated search and analytic algorithms, the log files will serve as an 'audit trail of an audit', thus enabling an efficient and effective tertiary assurance system." The inspiration for a BB log came from the black boxes carried on all passenger aircrafts and which are intended to assist investigations if the plane would crash, by recording the last 30 min of cockpit conversation and instrument settings. In the wake of the failings of Arthur Andersen at Enron and WorldCom, Alles et al. (2004) felt that a BB log could perform a similar function for auditing, albeit with records that were longer in duration and far more comprehensive, since the storage of such logs would not be constrained by size and survivability issues as they are on an aircraft. However, while the BB log was focused on the working of the audit and would require developing a way of automatically recording audit work papers, process mining analyses data created by the auditee firm itself and which is often already being recorded and stored by its ERP systems. Hence, this paper focuses on process mining which provides the "sophisticated search and analytic algorithms" of event logs, rather than the technology needed to create logs in the first place.

process mining into audit practice and proposes it as a superior way of conducting the required walkthroughs and analytic procedures. Section 4 lists the numerous ways in which event log data can be process mined, and the different perspectives and opportunities that analysis offers to auditors. Section 5 offers concluding comments. An appendix provides an overview of how an event log is created using information drawn from an ERP system.

2. The sources of value added of process mining when applied to auditing

2.1. Event logs and meta-data

The Business Process Mining Center describes *process mining* in the following terms[4]:

> The basic idea of process mining is to extract knowledge from event logs recorded by an information system. Until recently, the information in these event logs was rarely used to analyze the underlying processes. Process mining aims at improving this by providing techniques and tools for discovering process, control, data, organizational, and social structures from event logs. Fuelled by the omnipresence of event logs in transactional information systems… process mining has become a vivid research area.[5]

As the quote indicates, the source of data for process mining is an "event log", also called an "audit trail", which is defined as "a chronological record of computer system activities which are saved to a file on the system. The file can later be reviewed by the system administrator to identify users' actions on the system or processes which occurred on the system."[6] Despite the presence of the word "audit" in the term audit trail, there has been little use made by financial auditors of process mining to examine the data contained in event logs, and the term audit trail is mostly confined to IT and cyber-security circles. Hence, to avoid confusion, we avoid the term "audit trail" in this paper and confine ourselves to "event log".[7]

As the definition of the event log indicates, one of the sources of value added that process mining brings to auditing is that its use will force the auditor away from standard audit practice with its reliance on samples, towards a consideration of the population of data that is available in the firm's ERP system. Without sufficiently detailed information it becomes impossible to recreate all the steps in a business process and the more data that is fed into the process mining tool the better the understanding of underlying business processes. Of course, considerations of data complexity may induce limits on the dimensionality of the event log, but the fact remains that today's ERP systems routinely record far more data than is actually used by standard audit practice and process mining provides auditors with the means of accessing the information content of that unused data.

This is not, though, the most important source of value added that process mining brings to auditing. To understand what is, we first have to recall the way in which accounting used to be done. Until only a few decades ago (and still in many business in the developing world), transactions were manually entering into paper ledgers in an environment where accounting could accurately be described as "bookkeeping". While ledgers are now electronic datasets in an ERP system, one essential aspect of bookkeeping has not changed: that the data that the auditor has to rely on when checking what transactions the client firm has undertaken and how it has accounted for those transactions comes entirely from entries made by employees of the auditee, and, moreover, the auditor typically has no way of independently verifying who made those ledger entries and when they did so.

This is a fundamental dilemma in auditing—one that has been considered as so unavoidable that it has rarely been questioned even as the underlying technology has changed beyond recognition—that while the auditor is hired to verify the activities of the auditee, audit procedures are almost entirely dependent

[4] The BPM Center is a collaboration between the Information Systems groups (IS@CS and IS@IEIS) at Eindhoven University of Technology) and the Faculty of Information Technology of Queensland University of Technology. See http://is.tm.tue.nl/staff/wvdaalst/BPMcenter/index.htm.

[5] http://is.tm.tue.nl/staff/wvdaalst/BPMcenter/process%20mining.htm.

[6] http://www.fas.org/irp/congress/1996_hr/s960605a.htm.

[7] Other terms also used more or less synonymously for event logs are "transaction logs" and "history logs".

upon data entered by the auditee. Process mining provides the means by which the auditor can finally escape the constraints of this dilemma. As the definition of process mining indicates, it is the systematic analysis of an event log created from data obtained from the firm's ERP system. But that basic description fails to fully capture what is truly unprecedented about an event log, for it is far more than simply a "chronological record of computer system activities" as its definition states. That alone would make it not much more than a paper ledger since that is also a chronological record. Rather, thanks to the digital capabilities of the ERP system from which it is drawn, an event log also includes data recorded automatically and independently of the person whose behavior is the subject of the audit. With access to an event log the auditor is thus no longer restricted to data entered by the auditee, but also possesses an independent set of what we describe as contextual "meta-data" about the circumstances under which the auditee made those entries.

At a minimum, that meta-data encompasses a time stamp for transactions and an identifier for the person making those entries, though it can potentially go much further, including tracking change entries or even all keystrokes. The scope of the event log is constrained only by the choices made by IT personnel as to how much information is automatically recorded when an entry is made, but it certainly surpasses the null set of meta-data that accompanies entries in a paper ledger. The more complete is the information contained in event logs and the more access the auditor has to the totality of logs maintained by all relevant devices (the ERP system, the logs of the handheld devices of the users, their web histories etc.) the more closely the discovered process will match the underlying actual business process. However, using even the only most basic meta-data of time and ID process mining enables the auditor to reproduce much of the history of any given transaction and to trace the relationship of that particular entry and its author to all prior recorded transactions by that or related parties.

It is this inherent relationship-creating aspect of event logs that gives process mining its great power and its name: the ability through analysis of event logs to recreate the business processes of the firm. For example, through process mining the auditor has the ability to compare how processes such as purchase to pay were actually conducted as opposed to how they are supposed to be, or to ascertain, as in the example above, how the layoff of key workers impact segregation of duty controls. Such process views of the business are much more difficult, if not impossible, to obtain from transactional data alone, but feasible when that transactional data is supplemented by the meta-data and history contained in the event lots and made visible by the techniques of process mining.

2.2. Obtaining audit relevant information from an event log

While ERP systems automatically record meta-information about data entries, that information is not stored in any systematic or easily accessible fashion. Moreover, most IT systems, including the native auditing feature in many AIS systems, can record more meta-information than they actually do in practice, with the data capture feature not fully turned on in the absence of any demand for that information. The more information that is recorded in the IT system in addition to the actual data being entered, the slower the system tends to get. For privacy reasons, some data is either never stored, or deleted a short time after capture, such as those that track an individual's GPS location via their cellular phone (Prosch et al., 2010). In short, those seeking to do process mining have to first construct the event log, ideally determining in advance which information is to be recorded, but more often the case, making use of whatever meta-information already exists within the ERP system.

The starting point of event log creation is taking advantage of the fact that at the very minimum virtually all ERP systems will at least independently date stamp transactions (i.e. rather than rely only on the date entered by the user) and require users entering data to enter their login information. This date and originator information is by itself sufficient for a large amount of process mining analysis to be undertaken, but obviously more can be done if other meta-information is gathered, such as initial and corrected data entries, fingerprint or other biometric information to preclude use of stolen login passwords, internet search histories, communication logs or even all keystrokes.

The meta-information that is captured by the ERP system is located across numerous tables, whose logic schema depends on the characteristics of each ERP system as well as individual company settings, facts which increase the challenge in bringing this data together into a structured and usable event log. The scope and power of process mining is dependent on how comprehensive the event log is in including data on all activities relevant to the process being analyzed. Thus, when creating the event log it is

essential to first develop a holistic understanding of the activities that constitute the process of interest to the researcher. Jans (2009) and Gehrke and Mueller-Wickop (2010) both develop methodologies for extracting data from ERP systems and organizing it systematically into an event log, but each is forced to do this step from first principles and adopt somewhat different procedures in each case. The challenge facing process mining users is that there is as yet to no established or best practices in event log creation.

A full description of how an event log is created is beyond the scope of this paper, but an overview is provided in Appendix A. Our focus is this section is on the insights that the auditor can obtain from the event log.

Using an invoice as an example, Fig. 1 illustrates information that is entered into a firm's ERP system, and the way in which an event log would describe that transaction. The left hand side shows the data about the invoice that is entered by the person making the data entry, which is the auditee in the case of auditing. We will call this 'input data' as it is characterized by the controllable act of a person entering the data, and contrast it with the meta-data that is recorded by the ERP system independent of the auditee. This input data is the type of data available at the moment this data is stored, such as the invoice number, the posting date, the supplier, etc. This is also the type of data that is predominantly used by auditors.

The right hand side of Fig. 1 shows the data that is stored in the event log for that same invoice. By the definition of an event log drawn from its computer science roots, all input data is also part of the event log, but clearly it is the entries that are unique to the event log that are of particular interest to an auditor

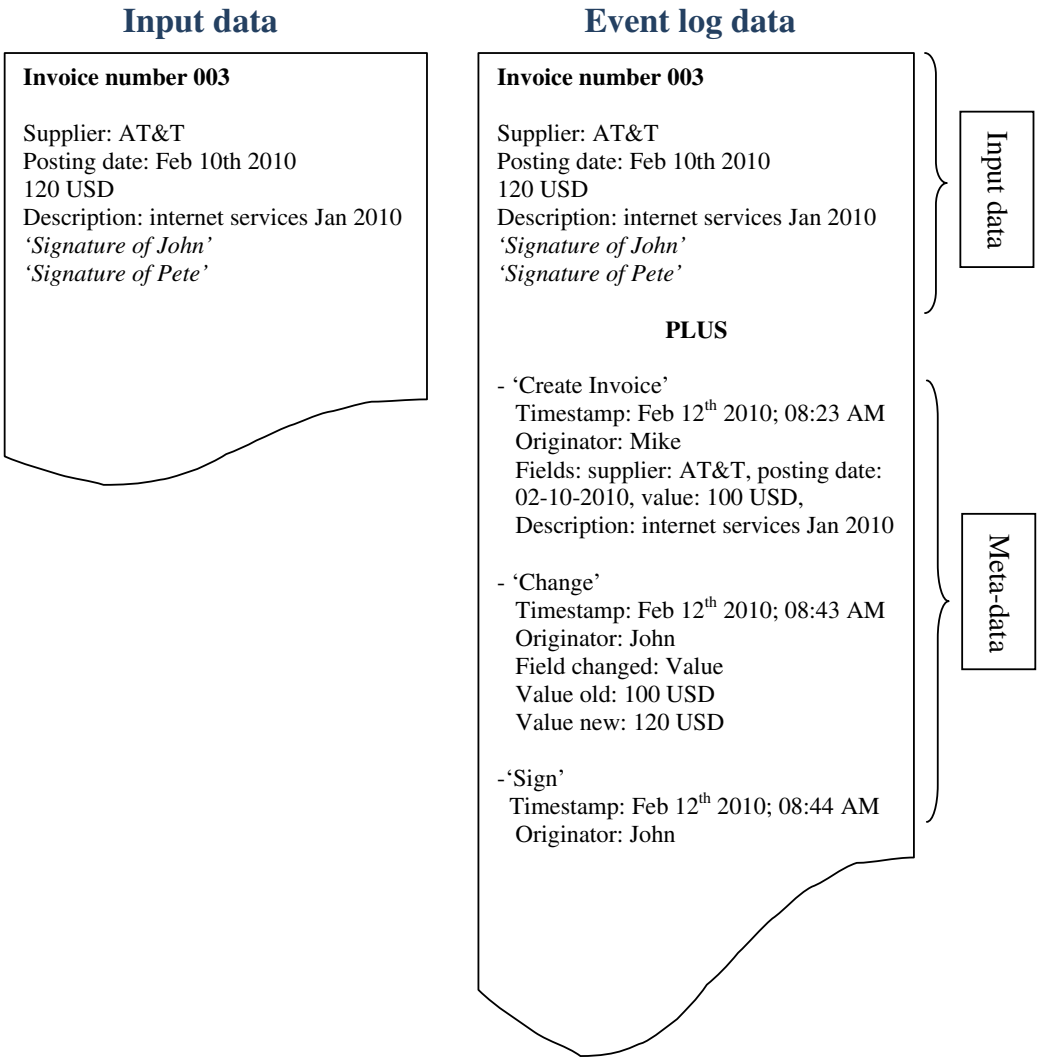


Fig. 1. Visualization of input data and event log data of an invoice.

because that data is recorded automatically by the system and not entered by the auditee.[8] It is this meta-data which makes an event log of larger dimension than the set of input data.

Contextual meta-data enables the auditor to reconstruct the history of a transaction by identifying relationships between that transaction and all other transactions in the database that share parameters with it. These include changes to the invoice and the identities of other individuals who interacted with the invoice in any way during its progress through the business process.

In the example illustrated in Fig. 1, input data would only show a transaction with a value of $120 with a posting date of Feb. 10, and that is what the auditor would see too if this was a paper based ledger system, or even an ERP based system in which the auditor chose to not construct an event log. By contrast the meta-data in that event log makes it clear that two separate individuals were involved with entering data on this transaction and that the dates they did so do not coincide with the entered date, and, further, not only was the entry amount changed, but that change was authorized by the very same person making the change.

Hence, by analyzing the information contained in the event log of the transactions shown in Fig. 1, the auditor can reconstruct the following sequence of events:

1. on Feb 12, 8:23 AM: Mike entered invoice no. 3 in system, filling out the supplier (AT&T), posting date (02-10-2010), invoice value (100 USD) and description (internet services Jan 2010)
2. on Feb 12, 8:43 AM: John changed 'Value' from '100 USD' to '120 USD'
3. on Feb 12, 8:44 AM: John signed invoice no. 3

While there may well be a perfectly acceptable reason for this sequence of events, this is information which would cearly make an auditor want to further investigate this transaction.

Two points should be kept in mind, however, with this example. First, as discussed above, whether all the meta-information shown in Fig. 1 is actually recorded and stored depends on whether a choice had been made earlier by the IT administrators to keep track of this data—in particular, in this example, of change entries, for otherwise only the last entered data would be available. Secondly, the history of this transaction is apparent from Fig. 1 because all relevant data had already been extracted and arranged into an easy to read narrative format. In reality, those various pieces of information would be stored at numerous locations in the IT system, and the auditor seeking to construct an event log would have to aggregate and assemble it before being able to obtain such insights so readily. Another factor that facilitated determining what had really happened with this transaction was that the event log extract shown in Fig. 1 consisted of only those entries relevant to it alone. In practice, even the best constructed event log would consist of a few anomalous transactions among a mass of routine ones and it requires the systematic procedures of process mining to extract the former from the latter.

2.3. Illustrating the value of meta-data in auditing

We close this section by giving some illustrations of the value added that the process mining of meta-data contained in event logs can bring to auditing. We begin with an example related by Mr. Vivek Kundra, the Chief Information Officer of the US federal government:

> "Another measure Mr. Kundra wants the public to see on these dashboards is a bit of bureaucratic sleight of hand known as 'rebaselining.' That's when the start data of a project is retroactively moved forward to make it look less late. 'They would reset the clock and pretend like they were in a year one, even though they lost a year,' he said."[9]

What Mr. Kundra is calling for is access to the meta-data of when the baseline of a project was entered into the system, rather having to rely solely on the data as entered by the project owners. Assuming the

[8] In other words, it would have been more consistent if the event log only consisted of meta-data, but it in fact contains both meta-data and input-data and they have to be distinguished when conducting process mining analyses.

[9] New York Times, June 15, 2009, "The Nation's C.I.O.: Government Needs a Dashboard".

baseline information is stored in an electronic database as opposed to a paper report, it is precisely such information that is potentially available in an event log: a record of all entered baseline dates, any changes made to them, when those changes were made and the identifier of the parties making those changes. With that information, the CIO or an auditor can detect the "bureaucratic sleight of hand" and hold those responsible to account.

A second example arises from forensic accounting where it is a routine assumption that fraud typically takes place at times where there are fewer other employees around to ask questions, such as at lunchtime. Some forensic accountants thus monitor the firm's ledgers at lunchtime to see who else is on the system, whereas the person committing the fraud might wish to cover their tracks by entering a different time for the fraudulent transaction they are inputting, even assuming that the system requires a time to be entered in the first place. But an event log will store the actual time of this transaction, regardless of what the person fills out, and that information is available to be detected by the auditor without recourse to monitoring at precisely the same time as the data is being entered.

Mining the event log data is also useful for detecting violations of segregation of duty controls. Blatant SOD violations, such as an invoice with no signature or approval, or an invoice signed only by one person instead of having the required dual signatures, are probably detectable using existing audit practices that utilize only input data. More subtle control violations such as employees not following procedures in the required order—for example, first getting approval and only then ordering the goods—would not be apparent by only looking at transactional data, but systematically analyzing the timestamps in the event log would reveal this circumvention of procedures to the auditor. Of course, such a violation of procedures could be prevented if the firm's ERP system is configured in such a way that transactions are prohibited from taking place out of order. But even if that were the case the auditor would still wish to ensure that such control settings have not been altered since they were set and process mining provides a way to check through a comprehensive test of details whether the control settings have consistently been in place since the last audit took place.

Another situation when process mining can aid the auditor is by providing a means to check whether the flexibility in the ERP system has been abused for personal gain. Consider, for instance, a collaborative fraud conducted between a supplier and an employee who changes a purchase order subsequent to its last approval. The supplier gets paid more than was agreed upon and provides a kickback to the involved employee. This type of abuse may be discovered by analyzing event log data since it captures changes to invoices, and the social network analysis tool of process mining can be used to trace anomalous relationships between employees and suppliers.

A final illustration of what process mining entails comes from an analogy to the tagging financial statements using XBRL. XBRL tags are often described as providing meta-information about the accounting information that is being tagged. For example, a sales figure drawn from the face of a company's income statement can be tagged with information about the accounting period that it refers to, the accounting standard used, the monetary unit and data format it is measured in, even whether it has been audited, all of which greatly increases the insights that the user obtains relative to just seeing the sales number alone. Moreover, once accounting statements are tagged, researchers can then analyze those tags to better understand the information they convey, for example, what the implications are of the company choosing to use an extension tag rather than a standard one (Debreceny et al., 2010).

Modern IT systems, particularly enterprise resource planning (ERP) systems, can be thought of as doing the equivalent of "tagging" the data that is entered into it, by independently recording such information as the time when the data entry was made, which authorized user made it and whether any corrections were subsequently made to that initial entry. ERPs tend to use relational databases and along these systems logs, data dictionaries, and information fields to track information about information (meta-information). It is this meta-information about the data in the ERP system, along with the transaction entry itself, which is extracted into an event log in a systematic fashion that facilitates its analysis. Process mining is the methodology for analyzing that meta-information contained in the event log, the equivalent of the accounting researcher analyzing not just the data on the face of the accounting statement, but also the data within its XBRL tags.

There is, however, a fundamental difference between process mining and XBRL tagging. In the case of XBRL, it is the individual in the company or the corporate publisher preparing the accounting statements for submittal to the SEC, who is responsible for tagging the data. By contrast, IT systems record

meta-information about data entries automatically and without the ability of the user to prevent or alter the recording of that information. It is that which makes an event log such a uniquely valuable resource for auditors and one we argue that they need to take advantage of to a much greater extent than they have done up to now.

Of course, for this meta-information to be of value it has to be assumed that the typical user entering data into a company's IT systems lacks the super-user privileges that would enable them to subvert the controls that lead to the automatic recording of that meta-information and which safeguards it from manipulation. But that constraint is not particular to process mining, since no type of data analytics, including those in utilized in standard audit practice, can be undertaken in the absence of a basic guarantee of data integrity.

3. Process mining in audit practice

Commenting on the reluctance of the audit profession to adopt such new technologies as continuous auditing, Mr. Bill Titera, a senior partner at Ernst & Young, stated that one cause was that the technology was "not sufficiently grounded" in auditing standards, meaning that it was not "baked into audit practice" in a way that would enable auditors to understand how the technology helped them do their jobs.[10] A similar concern applies to process mining, which is why in this section we place process mining within the context of the audit practices specified by audit standards. Our objective is to give auditors, both internal and external, a new and better way of meeting those standards, not just another way of doing so. To put it another way, realistically if process mining is to be adopted by auditors they have to be convinced that it is a replacement for what they are already doing, not an addition to their workload.

Our case for the adoption of process mining in auditing is built upon the Audit Risk Model since that framework determines the parameters of the audit. As described by Auditing Standard No. 8, audit risk is the risk that the auditor expresses an inappropriate audit opinion when the financial statements are materially misstated.[11] The audit risk model acts as a guide for carrying out audits of financial statements with the objective of reducing audit risk to an appropriately low level. Audit risk is a function of two components: detection risk and the risk of material misstatement. Detection risk is the risk that the procedures performed by the auditor will not detect a material misstatement.

Auditing Standard No. 12 indicates that the auditor should assess the risks of material misstatement at two levels: the inherent risk and the control risk.[12] The inherent risk refers to the susceptibility of an assertion to a misstatement, before any control is exercised. The control risk expresses the risk that a misstatement will not be prevented or detected by the company's internal control. AS 12 establishes requirements regarding the process of identifying and assessing the risk of material misstatement. Paragraph 5 of AS 12 provides in six risk assessment procedures (a. through f.). We see two procedures in particular that would significantly benefit from using process mining:

b. Obtaining an understanding of internal control over financial reporting
d. Performing analytical procedures

Concerning the first procedure, paragraph 18 of AS 12 explains further: "The auditor should obtain a sufficient understanding of each component of internal control over financial reporting ('understanding of internal control') to (a) identify the types of potential misstatements, (b) assess the factors that affect the risks of material misstatement, and (c) design further audit procedures." In particular, as paragraph 20 elaborates: "Obtaining an understanding of internal control includes evaluating the design of controls that are relevant to the audit and determining whether the controls have been implemented. Procedures the auditor performs to obtain evidence about design effectiveness include inquiry of appropriate personnel, observation of the company's operations, and inspection of relevant documentation. Walkthroughs… that include these procedures ordinarily are sufficient to evaluate design effectiveness."[13]

[10] Speech given at the 22nd World Continuous Auditing and Reporting Symposium, Sao Paolo, Brazil, June 2011.
[11] http://pcaobus.org/Standards/Auditing/Pages/Auditing_Standard_8.aspx.
[12] http://pcaobus.org/Standards/Auditing/Pages/Auditing_Standard_12.aspx.
[13] All emphasis added.

In paragraphs 20 and 37 of AS 12, walkthroughs are presented as a way of obtaining an understanding of internal control over financial reporting: "In performing a walkthrough, the auditor follows a transaction from origination through the company's processes, including information systems, until it is reflected in the company's financial records, using the same documents and IT that company personnel use. Walkthrough procedures usually include a combination of inquiry, observation, inspection of relevant documentation, and re-performance of controls."

We have quoted these standards at length in order to make the argument about the centrality of the walkthrough procedure as central to the implementation of the audit risk model. Moreover, walkthroughs are also an essential element in Audit Standard No. 5, which deals with the auditor assessment of internal controls under section 404 of the Sarbanes–Oxley Act. This also implies that walkthroughs are then also important to internal auditors, whose work the external auditor may wish to rely on under AU Section 302.[14]

Walkthroughs are an obvious place to introduce process mining into auditing. For the part of the process that is supported by an information system (which will be the largest part in any ERP-enabled business), process mining offers a superior alternative to the traditional walkthrough. By using process mining, more insights will be revealed than performing a regular walkthrough for the two reasons given in the prior sections: 1. the auditor can examine the population of transactions and not just a sample, and, 2. because the auditor can validate transaction entries with meta-data rather than relying only on data entered by the auditee.

The second procedure presented in AS 12 Paragraph 5 for assessing the risk on material misstatements is the use of analytical procedures. Paragraph 46 of AS 12 states that these analytical procedures should be designed to:

a. Enhance the auditor's understanding of the client's business and the significant transactions and events that have occurred since the prior year end; and
b. Identify areas that might represent specific risks relevant to the audit, including the existence of unusual transactions and events, and amounts, ratios, and trends that warrant investigation.

Both goals can better be attained by employing process mining techniques as an improvement over current analytical procedures, as we shall show in greater detail below. Turning to the audit standard that discusses analytical procedures (AU Section 329), we identify several points that could be enhanced by applying process mining (emphasis added)[15]:

AU Section 329 Paragraph .02 "A basic premise underlying the application of analytical procedures is that plausible relationships among data may reasonably be expected to exist and continue in the absence of known conditions to the contrary. Particular conditions that can cause variations in these relationships include, for example, specific unusual transactions or events, accounting changes, business changes, random fluctuations, or misstatements."

AU Section 329 Paragraph .04 "Analytical procedures are used as a substantive test to obtain evidential matter about particular assertions related to account balances or classes of transaction. In some cases, analytical procedures can be more effective or efficient than tests of details for achieving particular substantive testing objectives."

AU Section 329 Paragraph .05 "Analytical procedures involve comparisons of recorded amounts, or ratios developed from recorded amounts, to expectations developed by the auditor. The auditor develops such expectations by identifying and using plausible relationships that are reasonably expected to exist based on the auditor's understanding of the client and of the industry in which the client operates."

Process mining expands the evidential domain for analytic procedures by allowing the auditor to examine the business process that gives rise to transactions and not just the outcome of those processes. Using process mining embeds tests of details in the analytical procedures, complementing tests of controls. As with walkthroughs, the potential of process mining to be superior to existing analytic methods arises from its use of a larger set of data, including meta-data.

Of course, whether that potential will indeed result in the detection of audit relevant information not otherwise obtainable using only standard audit practices is an empirical question that is the subject of ongoing research. As far as the acceptance of process mining by auditors is concerned, though, the key is

[14] http://pcaobus.org/Standards/Auditing/Pages/AU322.aspx.
[15] http://pcaobus.org/Standards/Auditing/Pages/AU329.aspx.

not just its possible greater efficacy, but the fact that it fits into established audit standards: that process mining can indeed be "baked into" audit practice. Hence, auditors can be assured that using process mining will replace some of their existing tasks, not simply add to their workload.

4. Process mining as an audit tool

With current audit practice the information that is analyzed to issue an audit opinion is essentially the same as when the audit was performed in a paper-based setting: it relies almost entirely upon input data, and the same is true with internal audit practice. This data may well now be analyzed in a more sophisticated way using computerized tools, such as by using search queries in contrast to manual examination, but that is simply automating an existing manual procedure not reengineering audit practice to take full advantage of the capabilities inherent in ERP-enabled businesses.

The power of process mining of event logs comes not just from gaining meta-information about individual transaction data entries, but the ability that provides to detect patterns across transactions and the users entering that data, such as whether certain transactions are always associated with a certain supplier or employee, at a certain time, or in a certain order.

Given the wealth of information potentially contained in an event log, methodologies continue to be developed to mine them. In this section we briefly discuss the range of different ways of analyzing the information in event logs. Event log data are so rich that there are numerous lenses through which the information can be viewed, yielding many different types of insights into how underlying business processes operate. It is beyond the scope of this paper to examine all these methodologies in detail and research is needed on which ones are best suited to the particular needs of auditors. What we focus on in this section is to indicate the options available to auditors when analyzing event logs using process mining techniques.

At the most generic level, there are three fundamental process mining perspectives: the *process perspective*, the *organizational perspective* and the *case perspective*, which correspond to analyzing the event log to determine "How was the process undertaken?", "Who was involved in the process?" and "What happened with this particular transaction?" respectively.

The process perspective can be used by researchers to compare the process as it is meant to be performed against how it actually is and thus identify control failures and weaknesses. Adopting the organizational perspective enables underlying relations between those entering data or between those individuals and specific tasks to be made visible. The obvious use of this perspective is in checking segregation of duty controls. The case perspective focuses on a single process instance, tracing back its history and relationships of users that are involved in that history. This is particularly useful as a tool to follow up and investigate the details of anomalous transaction identified using the other perspectives.

The methodologies of process mining can be further classified by the approach used to implement these three perspectives. There are at least five different such approaches in process mining: 1. process discovery, 2. conformance check, 3. performance analysis, 4. social networks analysis, 5. decision mining and verification.

4.1. Process discovery

The most fundamental use of process mining is to analyze the event log in order to discover how the business process is actually carried out as contrasted with the ideal designed process model from which deviations have taken place in practice. Process discovery is carried out by examining timestamps to systematically establish the flow of activities through the process from beginning to end. For example, in the purchasing process, the object would be for the auditor to trace a paid invoice back to its approved purchase order and to understand the steps in-between. In theory all POs should follow the steps as specified by the system designer (for example, Create PO → Sign → Release → Receive goods → Receive invoice → Pay) but in practice there may be variations due to change orders, partial shipments, multiple shipments and/or payments and so on. Indeed, in practice, a process with a single designed procedure might be fulfilled in practice in hundreds of different ways, a fact which would clearly be relevant to an auditor concerned with understanding the extent to which controls have been relaxed.

Process discovery is unique to process mining, since it utilizes the meta-data on activities and timestamps. Using traditional analysis techniques would not yield these insights. Using the process discovery approach,

we argue, would give the auditor a far more comprehensive and thorough walkthrough of business processes than existing audit practices.

For example, the added value of this task is that it can both assure that processes take place as they are meant to and on the other hand reveal processes which are not supposed to take place. When employees circumvent procedures by not following the preferred following order of activities, this will become apparent in the process discovery output, which can also be demonstrated visually to the auditor. An example of a circumvented procedure could be placing an order with a supplier before getting an approval. It is important for auditors to know whether procedures are being followed for when employees know that controls are capable of being circumvented, they may perceive that as a window of opportunity for committing fraud.

Apart from a graphical output, the process discovery can also result in a summary of the various sequences by which processes have been carried out. A sequence is a log trace a process instance follows, such as, for example Create PO → Release → Receive goods → Receive invoice → Pay, in which the "Sign" step is absent. Sequences that are infrequently present in the event log provide the auditor with outliers to subject to follow up investigation, building upon the assumption of AU Section 329 that transactions that behave differently have a greater likelihood of representing fraud than transactions that occur frequently.

4.2. Conformance check

A conformance check does, as its name suggests, a confirmation as to whether the process reality matches the expectation or a standard. The expectation model can be either descriptive or prescriptive, in much the same way that standards in costing can be attainable or ideal. The point of comparing against a prescriptive model is often to see how employees have had to deviate from established procedures because of an unexpected constraint, such as the lack of key personnel or the need to expedite an order to please an important customer. Comparing practice with expectation is the essence of auditing in general and of analytic procedures in particular. Process mining offers a way of conducting far more sophisticated analytic procedures than such standard practices as ratio analysis and the like which fail to take advantage of the information contained in the entire business process.

4.3. Performance analysis

Performance analysis techniques such as Process Performance Analysis, Business Performance Analysis and Business Activity Monitoring, focus on the measurement of business process' performances. There are numerous commercial tools available to perform performance analysis of event logs (for example, Aris PPM, Business Objects, HP BPC), along with academic tools like EMiT, developed at Eindhoven University of Technology. Typically, performance analysis creates reports on Key Performance Indicators (KPI) for a business, such as throughput time of a process (providing the minimum, the maximum and average throughput time).

While performance analysis is not a new methodology, extending the techniques to take advantage of the meta-data in event logs is still a work in progress. This technique is perhaps of greater relevance to management than to auditors, but it can assist internal auditors to improve business process efficiency, as well as providing early warning of process failure.

An example of how to use this type of analysis in the monitoring role of an auditor can be found in identifying cases that go extremely quickly or slowly through the process. Further analyzing these cases with regard to the involved persons may reveal potential malpractices or failures in controls.

4.4. Social network analysis

Social network analysis utilizes the information contained in the event log about which authorized user entered each transaction. Not only does this allow the behavior of that individual employee to be tracked, but also the social networks that they are part of in the workplace and beyond to be ascertained. Auditors have long admitted that collusive fraud is the most difficult to detect. Research needs to be undertaken whether by using social network analysis it is possible to ascertain anomalous relationships, such as a pattern of invoicing and authorizations between the same set of people, an unexpected and repeated set of

transactions between employees in different functional areas, or the frequent interaction between the same employee and supplier. None of these are proof of fraud, but they are also the kind of outliers which auditors are required to identify and examine in detail under AU Section 329.

4.5. Decision mining and verification

Decision mining focuses on decision points in a discovered process model and is used to test assertions on a case by case basis. For example, this technique can examine whether after a user changes an invoice that auditee's next step is to seek a new authorization, or alternatively, to input receipt of the good. Acceptable variations from standard practice can be built into the analysis to detect material deviations.

In this section we have discussed three different perspectives when doing process mining and five different approaches, which yields a possible fifteen different combinations of possible analytic paradigms. Some of these possibilities will presumably prove to be more useful for auditors than others, but it will take a sustained investigative effort before it can be determined which is which. Moreover, the best way of utilizing process mining is not to see these techniques in isolation, or as substitutes, but as working in a complementary fashion as enhanced analytic procedures.

Researchers have explored many different analytic procedure tools, ranging from simple ratio analysis to continuity equations and cluster analysis (Hirst and Koonce, 1996; Kogan et al., 2010; Thiprungsri, 2010). Process mining does not replace these techniques, but rather, potentially provides way of refining their results. One of the major problems with any analytic procedures technique is the number of false positives, caused not by material anomalous transactions, but rather, as AU Section 329 Paragraph .02 states, by "routine business changes or random fluctuations". One possible way of separating the audit relevant from the immaterial is to examine the circumstances which gave rise to transactions flagged as potentially suspicious.

Thus, Kogan et al. (2010) use continuity equations as a means of modeling business processes to use as a benchmark in audit tests. With access to the universe of data provided by a continuous auditing system, they have the ability to provide modeling of complex business processes to an unprecedented level of detail. However, continuity equations are based only on transactional data, which means that while Kogan et al. (2010) can detect anomalies, they cannot be sure if the cause is an unusual but acceptable business event—such as a need to expedite an order for a valued customer—or caused by fraud. It is at this point that process mining may prove of value to explore in depth the circumstances which gave rise to that anomaly and to either identify a control failure, or alternatively, to refine the continuity equation benchmark to reduce future instances of false positives.

This procedure uses process mining as a follow up to a first step in which transactional data are analyzed and filtered to a set of potentially suspicious events which require further examination. Of course process mining can itself be used as the primary or even only analytic procedure, and it is an open research question whether a combination of data and process mining yields more efficient results, similar insights or different ones, meaning that the two techniques are either complements or substitutes. But given the greater research effort in auditing into transaction based analytic procedures, there is some advantage in using process mining as a validation check and follow up to those more familiar methodologies, rather than trying to convince auditors to abandon those established procedures altogether in its favor.

Hence, consider the tests that are used in a large Latin American Bank to examine anomalous transitory accounts. A transitory account is a holding account used to store funds until a more appropriate destination account for them can be determined. Not all of these accounts can be examine in detail without prohibitive cost, hence Kim et al. (2009) used continuous auditing techniques to identify transitory accounts that are particularly anomalous. The problem is that any transitory account is anomalous to some extent, which makes it more difficult to identify ones that are particularly suspicious. Process mining offers promise here to better understand the process by which a transitory account is created and to detect patterns in such behavior across branches or individual tellers or customers. For example, if incorrect entry of account numbers is a source of transitory accounts, process mining can inform the bank when such errors are more likely to occur, for instance, late in the day, or with newer tellers.

Another application of process mining to help refine the results of transactional analytic procedures arises in the work undertaken by Thiprungsri (2010) to apply the statistical technique of cluster analysis to

auditing. Again, the intention with that work is to follow the spirit of AU Section 329 by developing new methods of identifying outliers in transactional data, but as with all such techniques the issue is separating suspicious outliers from merely unusual ones. Applying process mining to the entirety of a firm's data warehouse may be prohibitively costly and difficult to do in a timely fashion, but once a cluster is identified, which by definition is small and anomalous, then process mining can be applied to the event logs of just those transactions to see what the commonalities are that made them cluster together.

The actual tools used in implementing process mining methodologies consist of various types of software that systematically analyzes the data contained in event logs. Gehrke and Mueller-Wickop (2010) develop their own algorithm using Java and the SAP™ proprietary Remote Function Call, since their focus is entirely on that ERP system. By contrast, the open source site www.processmining.org which is a collaboration between academics and such leading IT firms as Phillips and IBM, is devoted to developing generic tools for process mining. Their products include ProM, a generic open-source software for implementing process mining tools in a standard environment and ProMimport, a framework for the extraction of MXML-formatted logs from various information systems. The full discussion of these and other tools of process analysis, such as YAWL ("Yet Another Workflow Language") and the role played in process mining by Petri Nets and other workflow analysis and visualization tools, is beyond the scope of this paper (see ter Hofstede et al., 2010).

5. Conclusion

An analogy that can be applied to event logs is that of video surveillance used in businesses to safeguard assets and deter crime. While, as with all analogies, the parallels are not perfect, pursuing this line of thought enables us to begin to appreciate the new capabilities that process mining of event logs can provide to auditing—capabilities that are difficult to reproduce through auditing of input data alone.

The major difference between event logs and video surveillance recordings is that storing transactional data in an ERP system is cheap and that it is time and location stamped so that it is feasible for an auditor to search for anomalies and track back history in the event of a detected problem, such as theft or fraud. By contrast, many surveillance cameras in large business facilities, such as museums, are actually non-functioning replicas since there is no possibility of cost effectively monitoring their feeds. Of course, even searching an indexed event log is non-trivial given the magnitude of data they are likely to contain and particularly when the auditor does not have something very specific to search for, and that is why we call for further research into the application of process mining techniques to auditing. But the potential of process mining of event logs to add value to auditing is a very real one.

And it is that potential that gives rise to perhaps the most important benefit of creating event logs and process mining them by internal and external auditors: the resulting deterrence effect. That is the same reason why many businesses purchase replica surveillance cameras and feel that it is worthwhile to install them. Just the chance that someone might be watching a person's behavior can serve to constrain that behavior. How much more effective this deterrence effect would be then if an auditee knew that event logs were indeed being automatically and continuously monitored by auditors for anomalies and subject to tests of analytic procedures? What the protection surveillance cameras offer, after all, is trivial compared to that promised by process mining of event logs. The latter is potentially the equivalent of a situation when all the installed surveillance cameras are real and their feeds are actually being monitored all the time.

But like video cameras in a store, monitoring comes with a cost. Just as a store may not record every single rack of the store because the store owner is not able to watch all videos, the company needs to take a position on the extent of logging—for instance, while the employee is required to login with a unique identifying password before entering data into the firm's ERP system, typically they would not be required to do so before each and every subsequent keystroke, though in theory their place could have been taken by some other party.

Indeed, a company may turn off the creation of logs on some points in order to keep track of everything in especially critical areas. Returning to the surveillance camera analogy, the store owner could hire someone to continuously watch all their tapes in order to deter pilfering, for example, but this would be prohibitively costly. So the store owner has to decide on which products he wants to have a video camera on, and use it only as a means to provide evidence in case theft occurs, rather than as preventive control

designed to catch thieves in the act. Similarly, similar decisions made by the company about the extent of logging clearly impact the use that auditors can make of those event logs. As process mining becomes more mainstream in auditing, though, the greater the likely ability of auditors to work with the company's IT department to increase the scope of meta-data tracking and recording.

When considering the value added that process mining of event logs can provide to auditing, an important caveat to be kept in mind is the possibility of types of frauds that stay undetected even when processes are monitored and analyzed using process mining techniques. For example, obviously frauds that leave no electronic trace will not be detectable by process mining, or indeed, with any other information system based detection method. Also, despite the deterrence effect of having and mining the event log, the strength of that deterrence effect depends on the personal incentives and opportunities of those parties susceptible to committing fraud.

While in this paper we have made the case that process mining of event logs can add value to auditing and provided examples of what that tool can do, we cannot claim to have exhausted all possibilities for how it may do so. The fundamental reason for this is that many of the examples of process mining as applied to auditing that we present in this paper take as their starting point familiar manual audit procedures. In doing so, we are following the standard route in technology adoption, which is to first automate manual processes and only then, once a level of comfort is attained, to reengineer those processes to take full advantage of the capabilities of the technology.

Audit practice as we know has evolved in a world where the auditor only had access to input data. Those of us who grew up in that world can only imagine how different auditing would have been if the starting point was the meta-data in the event log and auditors were as familiar with the tools of process mining to exploit that data, as they are with such data mining techniques as regression analysis used with input data. Technology moves on, and so, we hope, will auditing.

Acknowledgments

We thank seminar participants at the 1st International Symposium on Accounting Information Systems in Orlando, Florida, Wuhan University in China, the International Symposium on Auditing Research in Singapore, the SET/AIS Midyear meeting in Florida and the European Accounting Association in Rome for their helpful comments. Further comments are welcome and may be addressed to alles@business.rutgers.edu.

Appendix A. Event log creation

The event log data that is captured by the ERP system is potentially vast in magnitude and dispersed over numerous tables (with a certain logic schema depending on the ERP system and company settings). In order to mine the event log and, hence, the process, a rigorous and defensible method of structuring the data needs to be developed. This configuration should enhance and facilitate the analysis of the event log. In this section the structure of the event log as input for process mining is analyzed and a case study using data extracted from the ERP system of a financial services firm in Europe is used to show how an event log is created.

The aim in creating an event log is to enable process mining of that data. Hence, the scope and power of that process mining are dependent on how comprehensive that event log is in including data on all activities relevant to the process being analyzed. Thus the two critical steps when creating the event log are the identification of activities and the selection of a process instance.

The first preparatory step is for the auditor to develop a holistic understanding of the activities that constitute the process being audited. For example, when the log data is about the process preceding a paid invoice, the underlying activities can be ‘create purchase order’, ‘sign purchase order’, ‘release purchase order’, ‘enter Goods Receipt’, ‘enter Invoice Receipt’, and ‘pay invoice’. Extra activities could be ‘alter purchase order’, ‘send goods back to supplier’, and so forth. Identifying these underlying activities is the first preparation step. Clearly, to some extent activity identification is a judgment call by the auditor, trading off the comprehensiveness of the process understanding versus the desire to reduce the dimensionality of the accompanying event log.

The second step in event log creation is the selection of a *process instance*, or a *case*. A process instance is the subject that undergoes the identified activities. In the case of the paid invoice, the process instance can be the invoice itself, an item line of the invoice, the corresponding purchase order or maybe the item line of the purchase order that triggers the invoice. In order to make the event log, a decision has to be made on the selection of the process instance.

Once the process activities are identified and the process instance is selected, the manipulation of log data into the required structure can be affected.[16] The suggested event log structure is based on four related tables. Fig. 2 visualizes the event log structure, which is a relational database with process instances being the cases that undergo activities. Each process instance (PI) has a unique ID, the PI-ID, which is recorded in the PI-ID table. Each activity a process instance is subjected to is called a *unique event entry* (UEE). The UEE table stores all these activities. The activity itself and a unique identifier (UEE-ID) are stored, along with the instance they are belonging to (PI-ID) and two extra fields of meta-data: the timestamp when this activity occurred and by which originator. Meta-data is shown as green in Fig. 2.

The activity is performed by the auditee and hence intuitively may be seen as input data (for example the activity of changing a PO results indeed in newly entered information), but the storage of the act itself is beyond the control of the auditee and hence should be seen as meta-data. All unique event entries belonging to one process instance constitute the *log trace* of this process instance. In Fig. 2, PI-ID 1 shows a log trace *Activity A–Activity B–Activity A*, being UEE-ID 1, UEE-ID 2, and UEE-ID 3.

In two separate tables, extra information on both the process instances and the unique event entries is captured. This is done by means of *attributes*, which is a term for the variables that describe the process instances and unique event entries. For each process instance, the same list of attributes is stored. In this example the supplier and the value of each purchase are attributes and examples of input data. Input data is shown as yellow in Fig. 2. However, attributes can also have a meta-data nature. With regard to the unique event entries, the attributes that are stored depends on the activity itself. For example: with activity A (Change PO) the attributes 'field changed' and 'old value' (of the field changed) are stored, both meta-data; with activity B (Enter Goods Receipt) attributes 'Goods Receipt number' and 'reference to invoice' are stored. The Goods Receipts number and the reference to invoice are usually meta-data since these linkages are mostly automatically in current ERP systems. Following this assumption, the list of UEE attributes in this example is purely meta-data based. However, a mixture of input data and meta-data could also occur. As is visually clear by looking at the colors in Fig. 2, by supplementing the input data (yellow-colored) with meta-data (green-colored), the auditor has far more data to monitor the auditee when using an event log than when analysis is restricted to input data alone.

The decision on which activities and attributes will be captured in the event log is determined by what attributes are available to be logged by the system and the judgment of the auditor as to the scope of the event log. It also has to be kept in mind that while ERP systems can automatically log a very large set of variables, for example, the table settings that control the system, comprehensive logging is computationally demanding and often the set of variables that are logged is constrained by having some of the logging capability turned off.

To illustrate the processes and judgments required to create an event log, we turn to a case study using data provided by a multinational financial services firm in Europe. The firm agreed to cooperate with this research and provided extracts of their ERP system, which we use to demonstrate how the event log was composed, using its procurement business process as the subject of analysis.

As discussed above, in order to create an event log two important preparatory steps are necessary: the identification of activities and the selection of a process instance. Based on interviews with the domain experts on the procurement process, the following activities were identified as constituting the procurement process: 'create a purchase order (PO)', a possible 'change of a line item', 'sign', 'release', 'input of the Goods Receipt (GR)', 'input of the Invoice Receipt (IR)', and 'pay'. The process instance we selected to undergo these activities is a line item of a PO. This choice of process instance is based on the typical structure the data is stored in within SAP, namely: the 'sign' and 'release' activities are linked to a complete PO instead of to a single

[16] The term manipulation refers to the restructuring of the data and should not be confused with altering the data itself.

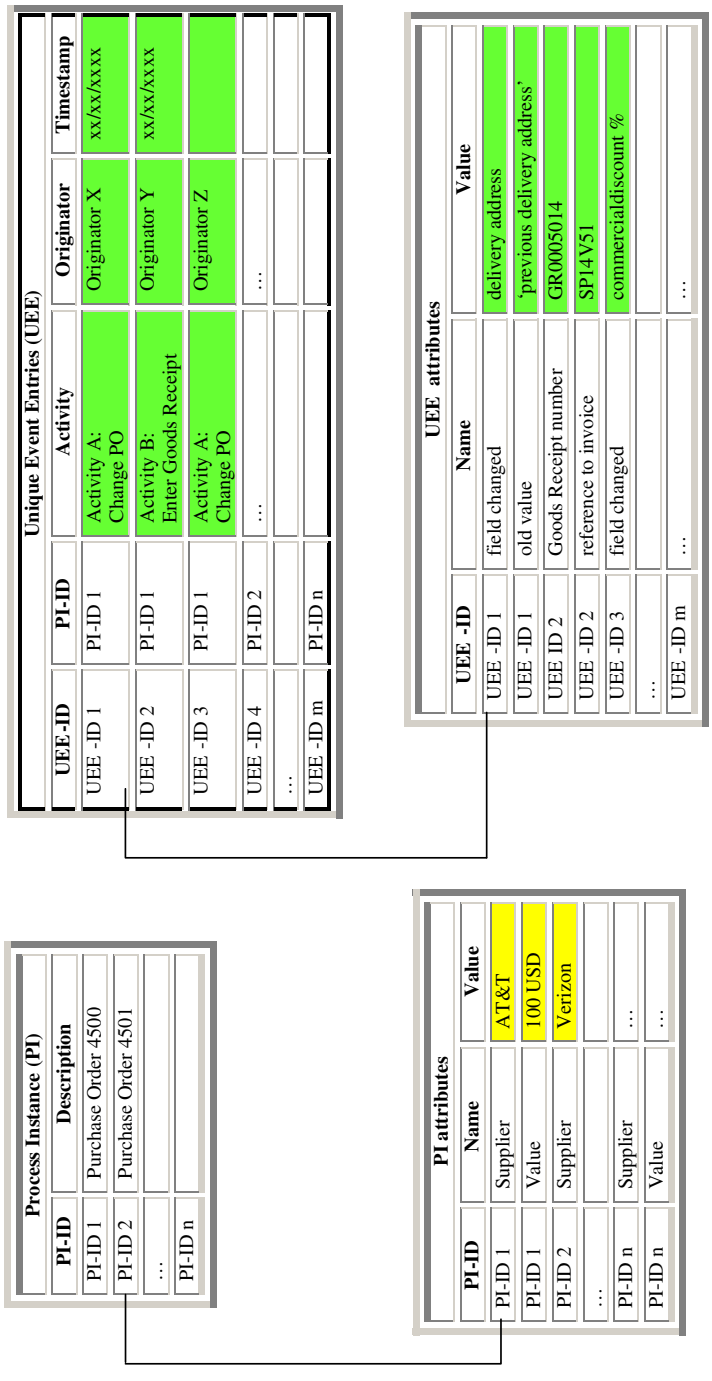


Fig. 2. Event log structure.

Table 1
Exemplary process instance table of case company.

Process instance (PI)	
PI-ID	Description
450000000110	Purchase order 4500000001, line 10
450000000120	Purchase order 4500000001, line 20
…	
450000025110	Purchase order 4500000251, line 10

line item, but the ultimate payment refers to a PO line item. That is why we chose a line item of a PO as process instance.

Table 1 shows how the table with process instances looks like in this case study. This table contains all the identifiers of the process instances under examination with a brief description.

In Table 2, an example of how the unique event entries look like with the selected activities is shown. As explained before, the data in this table is all meta-data.

Aside from the information in the table with the unique event entries, an additional table is created with data attributes of each process instance. This table contains information about the parent PO and about the item line itself, due to the double dimensionality of data in SAP. The following information was recorded about the parent PO: the document type, the purchasing group that creates this PO and the supplier involved. The selected information about the process instances concerns the value (in EUR) of the item line (Net value), the unit in which the quantity is expressed (Unit), the amount of ordered units on this line (Quantity PO) and whether the GR indicator was flagged. If this indicator is flagged, a choice which in this case is controlled by the employee, the input of a GR is mandatory for the payment of the invoice. If GR is turned off, an invoice can be paid without a receipt. Next to this PO related information, the total quantity and total value of all Goods Receipts that are linked to this PO item line are stored at the attributes table. The same is done for the related Invoice Receipts and the total value of all payments that are associated with this process instance. Notice that in this case study all process instance attributes are input-data. In Table 3 an example of the recorded data attributes of a process instance is shown.

The data attributes in Table 3 concern attributes of the process instances, but also a table with extra attributes of the unique event entries is created. In particular four activities are enriched with additional information: 'Change Line', 'IR', 'GR', and 'Pay'. If the event concerns a 'Change Line', the following information about the change is stored: when it was a change of the net value, what was the absolute size of this modification? If not the net value was changed but another field, for example the delivery address, this attribute stores a modification of zero. The second stored attribute of a 'Change Line' gives us, in case of a change in net value, the relative size of the modification (hence a percentage). If the event concerns an

Table 2
Exemplary unique event entry table of case company.

Unique event entries (UEE)				
UEE-ID	PI-ID	Activity	Originator	Timestamp
1	450000000120	Create PO	John	March 21, 2009
2	450000000120	Change Line	John	March 21, 2009
3	450000000120	Sign	Katy	March 30, 2009
4	450000000120	Release	Paul	April 1, 2009
5	450000000120	GR	Sarah	May 4, 2009
6	450000000120	IR	Mike	May 25, 2009
7	450000000120	Pay	Peter	May 30, 2009
8	450000000130	Create PO	John	…
9	…			
…	…			

Table 3
Exemplary PI attributes table of case company.

PI attributes		
PI-ID	Name	Value
450000000120	Document type	DI
450000000120	Purchasing group	B01
450000000120	Supplier	45781
450000000120	Net value	10,025
450000000120	Unit	EA
450000000120	Quantity PO	1
450000000120	GR indicator	X
450000000120	GR total quantity	1
450000000120	GR total value	10,000
450000000120	IR total quantity	1
450000000120	IR total value	10,000
450000000120	Pay total value	10,000
450000000130	Document type	FO

'IR', four attributes are stored: the references that contain the (possible) link to the 'GR' and 'Pay', the quantity of the units invoiced, and the credited amount, called the value. Notice that these quantities and values only concern this specific Invoice Receipt, as opposed to the Invoice Receipt related attributes of the Process Instance which were overall sums. Also beware that this information is not collected from an entire invoice, but only from the specific line that refers to the PO item line of this process instance. Similar to the 'IR', three attributes are stored when the activity concerns a 'GR': the reference to possibly link this Goods Receipt to the associated 'IR' (this is not always possible, only in a specific number of cases), the quantity of goods received and the resulting value that is assigned to this Goods Receipt. This value is the result of multiplying the Goods Receipt quantity with the price per unit agreed upon in the PO. The last activity that is provided with attributes is 'Pay'. The value of this payment is captured, as well as the key to create a link to an associated 'IR'. Table 4 is an example of how the table with UEE attributes could look like, based on the exemplary unique event entry table in Table 2. In Table 2, only the unique event entries with UEE-ID 2, 5, 6 and 7, representing a 'Change Line', a 'GR', an 'IR', and a 'Pay', would trigger the storage of extra attributes in the UEE attributes table.

Table 4
Exemplary UEE attributes table of case company.

UEE attributes		
UEE-ID	Name	Value
2	Modification	100
2	Relative modification	0.01
5	Reference IR	41358
5	Quantity GR	1
5	Value GR	10,000
6	Reference GR	41358
6	Reference pay	5100000832
6	Quantity IR	10,000
6	Value IR	10,000
7	Reference IR	5100000832
7	Value pay	10,000
…		

References

Alles M, Kogan A, Vasarhelyi M. Restoring auditor credibility: tertiary monitoring and logging of continuous assurance systems. Int J Account Inf Syst 2004;5(2):183–202.

Debreceny RS, Farewell C, Felden A, Graening M, Piechock. Causes and consequences of XBRL extensions: evidence from the SEC interactive data mandate. Working paper, University of Hawaii; 2010.

Gehrke N, Mueller-Wickop N. Basic principles of financial process mining: a journey through financial data in accounting information systems. Proceedings of the sixteenth Americas conference on information systems. Lima, Peru; 2010.

Hirst ED, Koonce L. Audit analytical procedures: a field investigation. Contemp Account Res 1996:457–86. [Fall].

Jans M. internal fraud risk reduction by data mining and process mining: framework and case study (PhD Thesis). Diepenbeek: Hasselt University; 2009.

Jans M, Lybaert N, Vanhoof K. A framework for internal fraud risk reduction at IT integrating business processes: the IFR² framework. Int J Digit Account Res 2009;9:1-29.

Kogan A, Alles M, Vasarhelyi M, Wu J. Analytical procedures for continuous data level auditing: continuity equations. Unpublished working paper, Rutgers Business School; 2010.

Kim, Yongbum, Miklos A, Vasarhelyi, Alex Kogan, Nilton Sigolo. Can a rule-based screening system adequately filter out the abnormal transactions? A two-stage model for Itau Unibanco, transitory accounts. Presented at ISAR Maastrich; 2009.

Lijie Wen L, Wang J, van der Aalst WMP, Huang B, Sun J. A novel approach for process mining based on event types. J Intell Inf Syst 2009;32:163–90.

Mans RS, Russell NC, van der Aalst WM, Bakker PJ, Moleman AJ, Jaspers MW. Proclets in healthcare. Journal of Biomedical Informatics 2010;43(4):632–49. [Aug] [Electronic publication ahead of print 30 Mar 2010] [PubMed PMID: 20359548].

Prosch M, Cavoukian A, David J. Privacy by design. Unpublished working paper. The University of Arizona; 2010.

Rozinat A, Alves de Medeiros AK, Günther CW, Weijters AJMM, van der Aalst WMP. The need for a process mining evaluation framework in research and practice. Lecture Notes in Computer Science, vol. 4928/2008. Springer; 2008. p. 83–9. http://dx.doi.org/10.1007/978-3-540-78238-4_10.

Schimm G. Mining most specific workflow models from event-based data. Proceedings of the International Conference on Business Process Management; 2003. p. 25–40.

ter Hofstede AHM, van der Aalst WMP, Adams M. Russell N. Modern business process automation: YAWL and its support environment. Springer; 2010 [ISBN: 978-3-642-03120-5].

Thiprungsri Sutapat. Cluster Analysis for Anomaly Detection in Accounting Data, dissertation proposal, Rutgers Business School; 2010.

van der Aalst WMP. Challenges in business process mining. Eindhoven University of Technology: Unpublished working paper; 2010a.

van der Aalst WMP. Process discovery: capturing the invisible. IEEE Comput Int Mag 2010b;5(1):28–41. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5386178&isnumber=5386087 [Feb].

van der Aalst WMP, de Medeiros AKA. Process mining and security: detecting anomalous process executions and checking process conformance. Electron Notes Theor Comput Sci 2005;121:3-21.

van der Aalst WMP, Weijters AJMM. Process mining: a research agenda. Comput Ind 2004;53:231–44.

van der Aalst WMP, Reijers HA, Weijters AJMM, van Dongen BF, Alves de Medeiros AK, Song M, et al. Business process mining: an industrial application. Inf Syst 2007;32(5):713–32. [July].

van der Aalst WMP, van Hee KM, van Werf JM, Verdonk M. Auditing 2.0: using process mining to support tomorrow's auditor. Computer 2010;43(3):90–3. March.

Wynn MT, Ouyang C, ter Hofstede AHM, Fidge CJ. Work flow support for product recall coordination. Unpublished working paper. Queensland University of Technology: Business Process Management Group; 2009.