# Materiality Maps – Process Mining Data Visualization for Financial Audits

Michael Werner
Copenhagen Business School
mwe.acc@cbs.dk

## Abstract

*Financial audits are a safeguard to prevent the distribution of false information which could detrimentally influence stakeholder decisions. The increasing integration of computer technology for the processing of business transactions create new challenges for auditors who have to deal with increasingly large and complex data. Process mining can be used as a novel Big Data analysis technique to support auditors in this context. A challenge for using this type of technique is the representation of analyzed data. Process mining algorithms usually discover large sets of mined process variants. This study introduces a new approach to visualize process mining results specifically for financial audits in an aggregate manner as materiality maps. Such maps provide an overview about the processes identified in an organization and indicate which business processes should be considered for audit purposes. They reduce an auditor's information overload and help to improve decision making in the audit process.*

## 1. Introduction

Financial audits are an important control mechanism to ensure the reliability of published financial information. Companies use information systems to improve and automate their business operations. These systems create large amounts of data which also serve as an important source to prepare an entity's financial statements. This data shows characteristics of Big Data due to its volume and speed of accumulation.

Auditors are in charge to assure that a company's financial statements provide a fair and true view of its financial situation and performance. Auditing standards issued by national and international bodies such as the International Auditing and Assurance Standards Board (IAASB), the Public Company Accounting Oversight Board (PCAOB) or the Institut der Wirtschaftsprüfer (IDW) define how financial audits have to be conducted. International and national auditing standards require the application of a risk based audit approach that encompasses the inspection of relevant business processes that affect the financial accounts, related internal controls and information systems [1].

Auditors struggle to carry out financial statements audits because of the ongoing integration of information systems for the operation of business processes and the growing amount of processed data. Process mining is a relatively new data analysis technique which uses event data as input that is recorded in operational information systems to create business process models as output [2]. Process mining can be used in financial statement audits to create reliable business process models very effectively and efficiently. This automates audit procedures which have traditional been carried out manually. Process Mining is able to deal with large input data sets from various source systems [3] which exhibit characteristics of Big Data [4]. However, empirical studies show that applying process mining techniques lead to hundreds of different process variants that have to be taken into account by an auditor [5], [6]. It therefore remains unclear if auditors will be able to use the techniques in real world scenarios due to the volume and complexity of the mining results.

This study deals with the question of how data generated by using process mining techniques can be aggregated to provide a comprehensive overview of an audited entity and how it can be used to guide further audit procedures.

Dealing with complex process mining results is not idiosyncratic to the application domain of financial audits. van der Aalst introduced the idea of process maps [2] that aggregate mined data and enable the data analyst to display different information and different levels of detail. He refers to cartography and navigation software such as Google Maps that provide such functionality. This study follows a similar idea and introduces materiality[1] process maps and materiality heat maps as novel data visualization approaches in the context of process mining for financial audits. The corresponding models provide a comprehensive overview about audit relevant information. They show in aggregate which business

---

[1] Materiality is a key concept in financial audits. It is defined in ISA 320 as follows: "Misstatements, including omissions, are considered to be material if they, individually or in the aggregate, could reasonably be expected to influence the economic decisions of users taken on the basis of the financial statements." [7, Para. 2].

HȈCSS

processes do exist in an organization, how they relate to the financial accounts and which of the analyzed processes should be considered for further inspection from a materiality perspective.

The research presented in this study follows a design science research approach (DSR) [8]–[10]. The main artefacts presented in this study are two novel types of data visualization for process mining results in the context of financial audits. The evaluation aspect is incorporated by using examples to illustrate the different model types. Data from a SAP IDES training system was used to mine and create the different models. The models were generated by using an implementation of the Multilevel Process Mining (MLPM) algorithm [6].

The next section describes the study's methodology. It follows a description of the analysis results in section three which provide the foundation for the development of the main artefacts that are presented in section four. The study closes with a conclusion and outlook to future research.

## 2. Methodology

DSR is commonly carried out in iterative cycles [9] covering different phases [11], [12]. This study is exploratory in nature and primarily addresses the analysis and design phase of a specific research iteration with using test data for evaluation purposes as a proof of concept. It complements published research work from prior iterations of related research on process mining for financial audits [13]–[15].

The analysis was carried out by reviewing related scientific publications and relevant International Standards on Auditing (ISA). The primary research methods for the design of the presented solutions were method engineering, prototyping and computational experiment. Method engineering [16] is commonly used in information systems science for the systematic design of methods [17]. A method in this context consists of different parts (method fragments) that can be combined and reused [18]. A new method can be engineered by recombining existing method fragments or by developing completely new method fragments. Several method fragments served as input for this study. It relies on the MLPM algorithm [6]. It was chosen because it is able to deal with specific characteristics observed in the source data, maintains the data perspective and is capable to analyze the totality of recorded events in an ERP system. The results of the MLPM for a given event log is a set of process models. The mined models are represented on the technical level as Coloured Petri Nets (CPN) [19] due to the mathematical robustness of this modelling

language and the ability to simulate the execution of mined models. For graphical representation in this study the Petri Net models were transformed into Financial Business Process Modelling Notation (FBPMN) [20]. It makes it possible to explicitly represent data objects that are relevant for accounting and auditing.
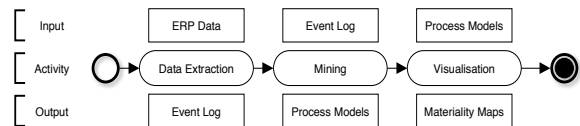


**Figure 1 Experimental Setup**

The different segments were implemented in a software prototype. Prototyping is traditionally used in software engineering [21]. It focuses on the development of a software artefact that implements the intended core functionality in iterative cycles. Prototyping was used as a research method to develop a software artefact that can be used to assess the effectiveness of the designed methods. The evaluation was carried out by conducting computational experiments. The setup of the experiment is shown in Figure 1. The data was extracted from a SAP IDES training system[2] and transformed into an event log. The extracted data included the tables BKPF and BSEG. BKPF stores the data of recorded journal entries, BSEG of recorded journal entry items. The source data consists of 115,060 journal entries and 419,106 journal entry items. The event log served as input for the MLPM algorithm that produced a set of process models. The MLPM mined 81,418 process instances that belong to 362 different business processes. These 362 served as the input for the computation of materiality process and heat maps. The mining of the process models from the SAP IDES test data set and the calculation of the materiality maps took two and a half minutes on a MacBook Pro with 16 GB of RAM and 2,8 GHz Intel Core i5 CPU. The yEd graph editor [23] was used to visualize process models whereas the gnuplot software [24] was used as a plotting software to produce the different maps.

The next section continues with the description of relevant literature related to this study as well as necessary background information on selected fundamental aspects related to financial audits, business processes and process mining which serve as a foundation for the subsequent sections.

## 3. Background

### 3.1. Related Literature

IDES test data set was used to enable other researchers to reproduce the presented results.

Of particular interest for this study is the literature related to process mining and financial audits. Process mining is a specific data analysis technique that uses recorded event log data to provide information about business processes. Process mining algorithms produce models by analyzing the available source event logs. Related research first emerged in the context of software engineering. It has matured significantly in the past decades with the development of powerful heuristic, fuzzy and genetic mining algorithms. The Process Mining Manifesto [3] provides a comprehensive overview of contemporary challenges in process mining. An overview of basic and advanced concepts on process mining can be found in [2].

Process mining can potentially be used in the context of financial audits to create reliable process models for audit purposes effectively and efficiently. Process mining is primarily used for the purpose of process discovery, conformance checking and process enhancement [3]. It has been applied in scientific studies to a variety of different application domains. The application domains range from the healthcare and IT sectors to the textile, shipping and manufacturing industry. Scholars have also applied process mining for the purpose of internal auditing [25] and general data mining techniques on journal entry data for auditing [26] and internal fraud detection purposes [27].

A fundamental challenge in process mining is the creation of useful process models. Mined process models are often too complex for simple interpretation and analysis. Extremely complex process models are referred to as *spaghetti* processes in the process mining community due to their graphical characteristics of intertwined arcs in the respective process models [28]. Complexity reduction of process models has been addressed by scholars following different approaches. Advanced mining algorithms such as the Fuzzy Miner [29] provide functionality to abstract from infrequent events and execution paths and are able to deliver less complex process models. However, infrequent behavior might indeed be relevant for compliance and conformance checking purposes [30]. Their omission in the process model might lead to process models that are of little use, especially in the context of financial audits. Other scholars suggest abstraction methods such as aggregation and reduction to decrease the complexity in process models [31].

This study deals with the complexity reduction for a set of different process models. Several publications deal with process repositories [32] and different related aspects such as similarity search [33] and clone detection [34] or clustering [35]. Melcher and Sees [36] discuss the concept of using heat maps for representing a process repository, but we are not aware of any publications that deal with process repository visualization for the purpose of financial audits.

## 3.2. Financial Audits and Business Processes

Business processes and related internal controls play an essential part in contemporary risk-based approaches to audit financial statements [1]. A business process is a set of connected activities that in combination realize a specific business goal [37]. Auditing business processes requires that the auditor has a sufficient understanding of the relationship between business activities, financial accounts and internal controls (Müller-Wickop et al. 2013). External auditors use process models to gain an understanding of the audited business processes, to identify incomplete, inaccurate, invalid or unauthorized transaction processing and to assess the design effectiveness of internal controls. These process models are usually created using traditional audit procedures like interviews and inspections of available documents. Process models and textual descriptions in current audit practices are prepared manually by using simple general purpose modelling tools such as Microsoft Visio, PowerPoint or Word. These procedures are highly time-consuming and error-prone. They become inefficient or even ineffective when the level of process automation and the volume of processed data increase. Contact persons from the audited entity traditionally serve as an important information source to gain information via interviews about relevant business processes. If these are operated automatically without any or with just limited human interaction the contact persons may not possess the necessary process knowledge anymore and performed interviews might provide little if any reliable information. The high volume of transactions in modern companies makes manual investigations of randomly selected samples very inefficient.

## 3.2. Financial Audits and Business Processes

Process mining techniques can be used to automatically create models in different stages of the audit process.

Figure 2 illustrates the main phases of a financial audit. The auditor first has to gain an understanding of the audited entity and its environment in order to identify risks of material misstatements at the financial statement and assertion level [1]. It follows the actual test of controls and substantive audit procedures as responses to identified risks [39]. An audit finishes with forming and communicating an audit opinion [40]. The audit planning itself is perceived as an ongoing iterative activity which continues throughout the complete audit [41].
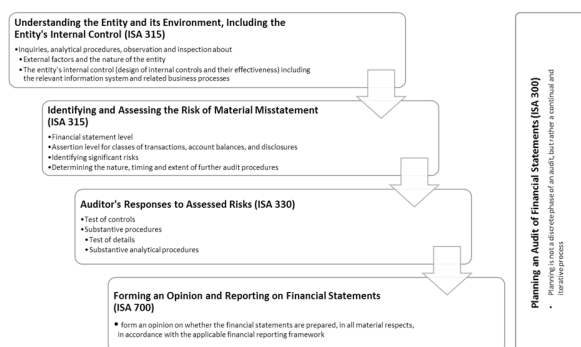
**Figure 2 Financial Statement Audit Process**

Process models are particularly relevant in the phases of understanding the entity, risk assessment and responses to assessed risks. In order to understand an entity the auditor has to gain information about the entity's internal control, relevant information systems and business processes. Process mining enables the auditor to receive information about which business processes do exist in an entity and how these relate to the financial accounts. During the risk assessment phase this information helps the auditor to decide the nature, timing and extent of testing. If transactions are processed, for example, by well-structured processes, test of controls are likely to be effective, whereas the opposite is the case if these transactions are processed by poorly structured and fragmented processes. During the response phase to assessed risks mined process models provide reliable information about the structure of audited business processes and potentially embedded internal controls. This can serve as reliable information to identify which internal controls have to be tested to gain the required audit evidence. It also helps to guide and determine the nature, timing and extent of further substantive audit procedures. Process mining makes it possible to identify potential high risk transactions that have caused high posting volumes and to detect outliers which represent deviations from standard procedures and therefore usually exhibit a higher audit risk. These can be inspected via targeted tests of details.

This study focuses on supporting an auditor during the collection of relevant data to understand the audited entity and its environment as well as the identification and assessing the risk of material misstatement. The following section shows how results from applied process mining techniques can be aggregated to provide the auditor an overview of relevant business processes, how they relate to financial accounts and if they are signifi-

cant from a materiality perspective. This allows the auditors to identify high risk processes and to allocate audit resources effectively. The next section describes the necessary procedures to produce materiality process and heat maps with an illustration of how these models can be integrated into the overall audit process.

## 4. Data Visualization

Process mining algorithms are able to produce process models based on a given event log. The MLPM maintains the data perspective when mining process models. This means that the data values related to recorded events are maintained. In the context of financial audits relevant data objects are the postings in the financial accounts. Figure 3 shows a typical process model mined by the MLPM for the SAP IDES test data set represented as a FBPMN model.[3] The FBPMN notation extension [42] uses specific symbols for modelling financial accounts and account entries. It provides information on the activities that were executed, involved financial accounts and posted values. The grey rectangles represent activities that were executed in the process. The BPMN group symbols (dotted colorless rectangles) represent the involved financial accounts. They consist of the account name and number at the top, the account symbol and credit or debit postings. These posting are modelled as BPMN data objects (paper symbols). The color of each of these objects signifies if it is a debit or credit posting on a balance sheet (blue and yellow) or profit and loss account (red and green). The dotted single-headed arrows leading from an activity to a posting denote that the corresponding activity posted a journal entry item to the connected account. The dotted double-headed arrows denote that an entry item was cleared in the respective account by the connected activity. The value of the posted or cleared item is displayed as an inscription for the corresponding data object.

The model represents a simple purchase process that consists of three activities *Post Received Goods*, *Post Received Invoices* and *Post Payment*. It starts with activity *Post Received Goods*. This executed activity created debit postings in the account *Raw Materials* with the aggregated amount of *289,911.68 €* and debit postings on the *Goods Received / Invoices Received* accounts with the same amount. It follows the activity *Post Received Invoices*. The activity *Post Payment* is the last activity which is carried out in this process. The model represents 24 process instances, hence 24 identical executions of the same business process.

---

[3] The MLPM produces Coloured Petri Nets (CPN) models [19]. The CPN models were manually transformed into

FBPMN models for illustration purposes in this study. The original CPN model is available in Appendix A
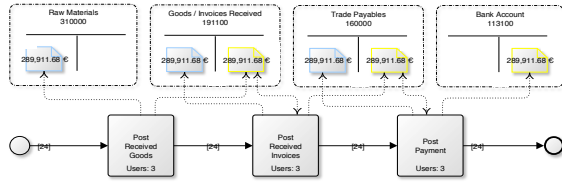
**Figure 3 Example Process Model**

The MLPM algorithm produced 362 process models of the type shown in Figure 3 for the complete SAP IDES test data set. The question now arises of how an auditor can use the mined models for a financial audit without suffering from information overflow. Inspecting each individual process model seems to be time-consuming and inefficient.

During a financial audit an auditor intends to assess if the financial statements are free of material misstatements. Auditors only need to consider those business processes that have a material effect on the financial accounts. All processes that are not material can be neglected. If the mined process models provide information about the effect on the financial accounts this information can be used to visualize those that are material.

This prerequisite is given for the model shown in Figure 3. It contains information about the aggregated values that have been posted to the financial accounts by the different process activities. These can be aggregated and analyzed for the whole process model repository. The MLPM algorithm produces models as Coloured Petri Nets (CPN). A Coloured Petri Net is formally expressed by the tuple CPN = $(T, P, A, \Sigma, V, C, G, E, I)$ [19, p. 87] as explained in Listing 1.

1) $T$ is a finite set of transitions
2) $P$ is a finite set of places
3) $A \in P \times T \cup T \times P$ is a set of directed arcs
4) $\Sigma$ is a set of non-empty color sets
5) $V$ is a finite set of typed variables such that Type$[v] \in \Sigma$ for all variables $v \in V$
6) $C: P \rightarrow \Sigma$ is a color set function that assigns a color set to each place
7) $G: T \rightarrow EXPR_V$ is a guard function that assigns a guard to each transition $t$ such that $Type\ [G(t)] = boolean$
8) $E: A \rightarrow EXPR_V$ is an arc expression function that assigns an arc expression to each arc $a$ such that $Type\ [E(a)] = C(p)_{MS}$ , where $p$ is the place connected to the arc $a$.
9) $I: P \rightarrow EXPR_\emptyset$ is an initialization function that assigns an initialization expression to each place $p$ such that $Type\ [I(p)] = C(p)_{MS}$

**Listing 1 Coloured Petri Net Specification**

Places can be modelled as control places (CP) that determine the control flow in a model or they can represent financial accounts (AP). Arcs can be modelled as control arcs (CA), posting arcs (PA) or clearing arcs (CLA). Arc inscriptions are modelled as constants with $V = \{\}$. The color set function assigns different color sets to places depending if they belong to the group of control or account places. Guards are not necessary as these types of model do not model dynamic behavior of transitions that depends on specific input but illustrate the processing of already executed processes. Hence $G(t) = true\ \forall\ t\ \in T$. The arc expression function assigns to each posting and clearing arc a set of constants that denote the posted or cleared value, the account number, account type and an indicator if it is a credit or debit posting. For each control flow arc the number of execution times is assigned indicating how often this path was discovered in the process model. The process model shown in Figure 3 can be expressed using the above CPN specification as shown in Listing 2.

*Transitions*: $T = \{A, B, C\}$

*Places*: $P = \{CP, AP\}$ *with*
  $CP = \{Source, S1\ , S2, Sink\}$
  $AP = \{310000\_D, 191100\_C, 191100\_D,$
    $160000\_C, 160000\_D, 113100\_C\}$

*Arcs*: $A = \{CA, PA, CLA\}$ *with*
  $CA = \{(Source, A), (A, S1), (S1,B), (B, S2),$
    $(S2, C)\ (C, Sink)\}$
  $PA = \{(A, 310000\_D), (A, 191100\_C),$
    $(B, 191100\_D), (B, 160000\_C),$
    $(C, 160000\_D), (C, 113100\_C)\}$
  $CLA = \{(B, 191100\_C), (C, 160000\_C)\}$

*Color sets:* $\Sigma = \{VALUES, ACCOUNTS, AC-$
    $COUNTTYPE, CREDorDEB,$
    $EXECUTIONS\ , ACCOUNT-$
    $PLACES\}\ with$
  $VALUES = \{289,911.68\}$
  $ACOUNTTYPE = \{0,1\}$
  $CREDorDEB = \{0,1\}$
  $EXECUTIONS = \{24\}$
  $ACCOUNTS = \{310000, 1910000, 160000,$
    $113000\}$
  $ACCOUNTPLACES = VALUES * ACCOUNTS$
    $* ACCOUNTTYPE$
    $* CREDorDEB\}$

*Initialization*: $I(p) = \{1'24\ for\ p$
  $= Source\ and\ \emptyset_{MS}\ \ for\ p \neq Source$

**Listing 2 CPN Notation for Example Process**

In order to model the relationship between a set of business processes, financial accounts and posted volumes it is necessary to summarize the postings created by different activities in a process for a particular account. The pseudocode shown in Listing 3 describes the necessary procedures.

1) $PM$ = Set of all mined process models
2) For each process model $pm_i \in PM$ select all account places $ap_j \in AP_i$
3) For each account place $ap_j$ select those posting arcs $pa_{kj}$ where $pa_{kj} \in PA_i = (t_k, ac_j)$
4) For each $ap_j$ calculate the sum of all arc inscriptions $\sum_j E(pa_{kj})$

**Listing 3 Procedures for Calculating Posting Volumes per Account**

The result of this procedure for the example process shown in Figure 3 is illustrated in Table 1.

**Table 1 Summarized Postings per Account**

| $ap_j$ | 310000_D | 191100_C | 191100_D | 160000_C | 160000_D | 113100_C |
|---|---|---|---|---|---|---|
| $\sum_j E(pa_{kj})$ | 289,911.68 | 289,911.68 | 289,911.68 | 289,911.68 | 289,911.68 | 289,911.68 |

This data can now be used to create a matrix which contains the accounts and posting volumes of all mined processes. The pseudocode shown in Listing 4 describes the necessary procedures.

The result of the operations described in Listing 3 and Listing 4 is a matrix where each row represents a particular process and each column a separate financial account.

1) Let M be an initially empty matrix
2) Set the number of rows $M_{|r|} = |PM|$
3) Set the number of columns $M_{|c|} = |AP|$ where $AP = ap_i \cup ap_j \, \forall \, pm_i \lor \, pm_j \in PM$
4) For each $pm_i \in PM$ set $\sum E(pa_{kj})$ at $M_{i,j}$

**Listing 4 Procedures for Calculating a Materiality Matrix**

Such a matrix can be graphically represented by using suitable plotting programs. Figure 4 shows an example of a visual representation by using the plotting program gnuplot. It refers to a different set of five simple process models that have been mined from the SAP IDES test data. The script for executing the plotting is available in Appendix B.
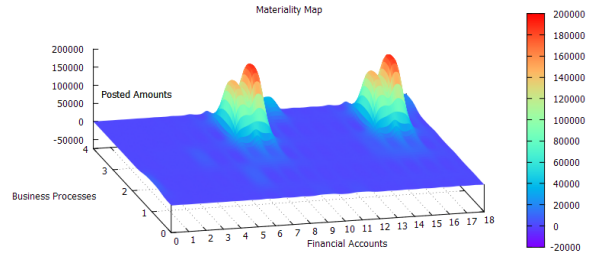


**Figure 4 Example Materiality Process Map**

Figure 4 shows a materiality process map. It is called a materiality process map because it shows the different financial processes and indicates if these have created material postings in related financial accounts. The different financial accounts are shown on the abscissa, the different process models on the applicate and the posted amount on the ordinate. The model shows the five different processes and 19 different financial accounts. This type of diagram visualizes the aggregate effect of the different processes on the financial accounts. The color scheme visualizes the magnitude of recorded postings. Figure 5 shows the result for the same set of processes if the three-dimensional materiality process map is transformed into a two-dimensional materiality heat map. It illustrates that a single process (#3) in the analyzed process repository created the most significant postings on four different accounts (#7, #8, #15, and #16).
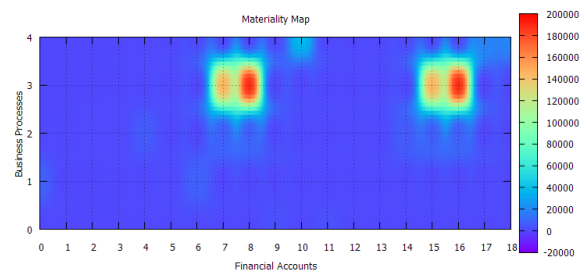


**Figure 5 Example Materiality Heat Map**

The different diagrams visualize the combined effect of five different processes on the related financial accounts. The implications of the suggested visualizations become more insightful if the procedures are applied to a larger data set. Figure 6 shows the materiality process map for a subset of processes from the SAP IDES System. The map represents those processes that were not

considered as being trivial[4]. For illustration purposes just those processes were chosen where the aggregated posting volume was above 1 million Euro.[5]

The materiality process map shows 82 processes that met these specifications. They caused postings in 188 different accounts. It can be seen that just a few pro-

cesses created very large postings in relatively few accounts which are visualized as steep peaks. Figure 7 shows the corresponding materiality heat map. It illustrates that accounts 0 to 68 are comparatively hot whereas accounts 69 to 187 are cool with the exception of accounts 160 and 162.
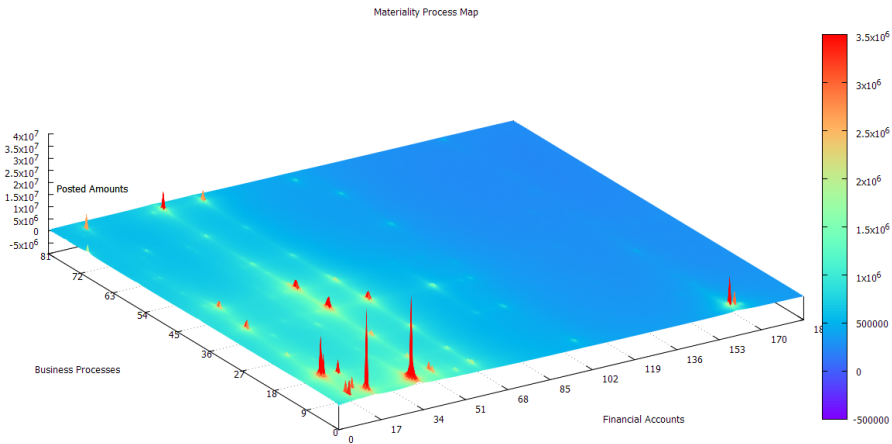


**Figure 6 Materiality Process Map for IDES Test Data Set**

The gained information displayed in Figure 6 and Figure 7 helps the auditor to understand the entity and its environment, to identify and assess the risks of material misstatements as well as to determine the nature, timing and extent of further audit procedures.
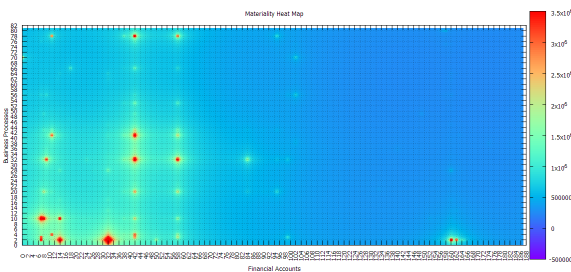


**Figure 7 Materiality Heat Map for IDES Test Data Set**

Figure 6 shows the auditor the overall number of identified processes and accounts. It reveals that the processes in the audited entity are very heterogeneous. A few create extremely large amounts of postings to the financial accounts whereas others create just relatively

small amounts. The same is the case for the financial accounts, just a few show very high overall posting volumes. From an audit perspective it can be assumed that especially those processes contribute to a risk of material misstatement that have created large posting amounts.[6] These are those processes that are shown as hot spots in Figure 7 (processes #2, #3, #4, #10, #20, #32, #41, and #78). As a response an auditor could dedicate audit resources to test the related internal controls which govern the identified material processes. Transactions belonging to other processes could be tested via using analytical procedures or sampled test of details.

Currently the information shown in Figure 6 and Figure 7 is not available to auditors at all. Auditors usually have no sophisticated knowledge about which processes do exist in an audited entity and how they relate to the financial accounts. They have to rely on their experience or potentially incomplete and unreliable information gathered though time-intensive interviews. The procedures described in this study allows the auditor to get insights into information which can be gained by automated analysis procedures and that enables them to

---

[4] Process models are considered as being trivial if they just consist of a single process activity. Trivial processes do not represent the execution of business processes but the execution of atomic activities. Although important from an overall audit perspective such transactions can be neglected from a process perspective.

[5] The overall posting volume of mined processes is comparatively high because all recorded debit and credit postings are considered and not just balances at year-end.

[6] It should be taken into account that the materiality maps illustrate the aggregate impact on the financial accounts, hence posting volumes can be the results of just a few large scale but also of many small scale postings in a process.

plan adequate audit procedures and to allocate audit resources effectively. It can be assumed that the gained information greatly increases the auditor's ability to direct audit resources to those processes that exhibit the highest audit risk.

This study focusses on the results from empirical studies which showed that process mining algorithms used for external auditing purposes produced comparatively large sets of mined process models [5], [6]. This is the case because auditors are interested in deviations from standard procedures to identify potential compliance violations. Other mining algorithms abstract from individual process variants (e.g. [29]) and produce models that represent several variances in a single model but at the cost of potentially imprecise and unfitting models [44]. Future research might look into alternatives of aggregating process models into less variants using procedures which are suitable for the application domain. Other aspects could include investigations of how temporal aspects can be included and how to annotate CPNs accordingly during the modelling process.

## 5. Conclusion

Financial audits are an important safeguard to prevent the distribution of false or misleading information which could detrimentally influence the decisions made by stakeholders in a society. The increasing integration of computer technology for the processing of business transactions and the growing amount of financially relevant data in organizations create new challenges for external auditors. Traditional manual audit procedures like inquiries and the inspection of sampled evidence documents become inefficient and eventually ineffective with the increase of automated transaction processing.

The availability of digital data opens up new opportunities for innovative audit procedures. Process mining is a novel Big Data analysis technique that potentially can help auditors in this context. Process mining algorithms produce process models as graphical representations by analyzing potentially extreme large event log data extracted from the source information systems used by the audited entity which commonly exhibit characteristics of Big Data due to its volume and speed of accumulation.

A challenge for using process mining is the representation of analyzed data. Empirical studies indicate that applying process mining techniques lead up to hundreds of different process models which cannot be inspected efficiently by manual observation. This study deals with the question of how data generated by using process mining techniques can be aggregated to provide a comprehensive overview of an audited entity and how it can be used to guide further audit procedures. It intro-

duces a new approach to visualize process mining results for financial audits in an aggregate manner as materiality process and heat maps. Materiality is a fundamental concept in financial audits. Just those processes have to be considered by the auditor which have a material effect on the financial accounts. Materiality process and heat maps provide an overview about the processes identified in an organization and indicate which business processes should be considered for further audit purposes due to their impact on the financial accounts. They reduce an auditor's information overload and help to improve decisions making in the audit process. Material processes can be easily identified and audit resources allocated accordingly. It can be assumed that the application of suggested solutions significantly improves decision making processes in financial audits related to the extent, nature and timing of further audit procedures.

This study focuses on the technical aspects to create the introduced models and relies on a comparatively small test data event log to demonstrate its applicability. The extension to real data sets, integration of additional examples and evaluation of the desired effects in real world audits is intended to be covered by future research.

A limitation of this study is the fact that the presented procedures rely on a specific mining algorithm which uses CPN as the output modelling language. The majority of mining algorithms use Petri Nets as the representational modelling language [45], but it is unclear if other mining algorithms also provide the necessary data relationships between process activities and financial accounts. To date it remains unknown if the presented procedures also work for models produced by other mining algorithms.

## 6. References

[1]  IFAC, ISA 315 (Revised), Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment. 2012.

[2]  W. M. P. van der Aalst, Process Mining: Data Science in Action. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016.

[3]  W. M. P. van der Aalst et al., "Process Mining Manifesto," in BPM 2011 Workshops Proceedings, 2012, pp. 169–194.

[4]  H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS Quarterly, vol. 36, no. 4, pp. 1165–1188, Dec. 2012.

[5]  M. Werner, "Financial Process Mining - Accounting Data Structure Dependent Control

Flow Inference," International Journal of Accounting Information Systems, vol. 25, pp. 57–80, May 2017.

[6] M. Werner and N. Gehrke, "Multilevel Process Mining for Financial Audits," IEEE Transactions on Services Computing, vol. 8, no. 6, pp. 820–832, Nov. 2015.

[7] IFAC, ISA 320 Materiality in Planning and Performing an Audit. 2009.

[8] S. T. March and G. F. Smith, "Design and Natural Science Research on Information Technology," Decision Support Systems, vol. 15, no. 4, pp. 251–266, 1995.

[9] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research," MIS Quarterly, vol. 28, no. 1, pp. 75–105, Mar. 2004.

[10] V. K. Vaishnavi and W. Kuechler, Design science research methods and patterns: innovating information and communication technology. CRC Press, 2015.

[11] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," Journal of Management Information Systems, vol. 24, no. 3, pp. 45–77, Dec. 2007.

[12] H. Österle et al., "Memorandum on Design-oriented Information Systems Research," European Journal of Information Systems, vol. 20, no. 1, pp. 7–10, 2010.

[13] M. Werner, N. Gehrke, and M. Nuttgens, "Business Process Mining and Reconstruction for Financial Audits," in Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS 2012), Maui, 2012, pp. 5350–5359.

[14] M. Werner and M. Nuttgens, "Improving Structure: Logical Sequencing of Mined Process Models," in Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS 2014), Big Island, 2014, pp. 3888–3897.

[15] M. Werner, "Process Model Representation Layers for Financial Audits," in Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS 2016), Kauai, 2016, pp. 5338–5347.

[16] S. Brinkkemper, "Method Engineering: Engineering of Information Systems Development Methods and Tools," Information and Software Technology, vol. 38, no. 4, pp. 275–280, 1996.

[17] T. Wilde and T. Hess, "Forschungsmethoden der Wirtschaftsinformatik," Wirtschaftsinformatik, vol. 49, no. 4, pp. 280–287, 2007.

[18] A. F. Harmsen, J. N. Brinkkemper, and H. Oei, Situational Method Engineering for Information System Project Approaches. University of Twente, Department of Computer Science, 1994.

[19] K. Jensen and L. M. Kristensen, Coloured Petri Nets. Springer, 2009.

[20] N. Müller-Wickop, "Integration von Wertflüssen in Geschäftsprozessmodellierungssprachen: Ein gestaltungsorientierter Ansatz zur Unterstützung von Revisoren bei der wertflussorientierten Planung und Durchführung von Prozessprüfungen," Hamburg, 2014.

[21] J. D. Naumann and A. M. Jenkins, "Prototyping: the New Paradigm for Systems Development," MIS Quarterly, pp. 29–44, 1982.

[22] SAP, "SAP-UCC," 2017. [Online]. Available: http://www.sap-ucc.com/.

[23] yWorks GmbH, "yEd - Graph Editor," 2017. [Online]. Available: http://www.yworks.com/products/yed.

[24] gnuplot, "gnuplot homepage," 2017. [Online]. Available: http://www.gnuplot.info/.

[25] M. Jans, M. G. Alles, and M. A. Vasarhelyi, "A Field Study on the Use of Process Mining of Event Logs as an Analytical Procedure in Auditing," The Accounting Review, vol. 89, no. 5, pp. 1751–1773, Sep. 2014.

[26] R. S. Debreceny and G. L. Gray, "Data Mining of Electronic Mail and Auditing: A Research Agenda," Journal of Information Systems, vol. 25, no. 2, pp. 195–226, Nov. 2011.

[27] R. S. Debreceny and G. L. Gray, "Data Mining Journal Entries for Fraud Detection: An Exploratory Study," International Journal of Accounting Information Systems, vol. 11, no. 3, pp. 157–181, Sep. 2010.

[28] W. M. P. van der Aalst, "Process Mining: Discovering and Improving Spaghetti and Lasagna Processes," in IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 2011, pp. 1–7.

[29] C. Günther and W. M. P. van der Aalst, "Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics," Business Process Management, pp. 328–343, 2007.

[30] W. M. P. van der Aalst and A. K. A. de Medeiros, "Process Mining and Security: Detecting Anomalous Process Executions and Checking Process Conformance," Electronic Notes in Theoretical Computer Science, vol. 121, pp. 3–21, 2005.

[31] M. Reichert, "Visualizing Large Business Process Models: Challenges, Techniques, Applications," in 1st Int'l Workshop on Theory and Applications of Process Visualization, Tallin, 2012.

[32] M. La Rosa et al., "APROMORE: An advanced process model repository," Expert Systems with

Applications, vol. 38, no. 6, pp. 7029–7040, Jun. 2011.

[33] R. Dijkman, M. Dumas, B. van Dongen, R. Käärik, and J. Mendling, "Similarity of business process models: Metrics and evaluation," Information Systems, vol. 36, no. 2, pp. 498–516, Apr. 2011.

[34] C. C. Ekanayake, M. Dumas, L. García-Bañuelos, M. L. Rosa, and A. H. M. ter Hofstede, "Approximate Clone Detection in Repositories of Business Process Models," in Business Process Management, 2012, vol. 7481, pp. 302–318.

[35] C. Li, M. Reichert, and A. Wombacher, "The minadept clustering approach for discovering reference process models out of process variants," International Journal of Cooperative Information Systems, vol. 19, no. 03n04, pp. 159–203, Sep. 2010.

[36] J. Melcher and D. Seese, "Visualization and Clustering of Business Process Collections Based on Process Metric Values," 2008, pp. 572–575.

[37] M. Reichert and B. Weber, Enabling Flexibility in Process-aware Information Systems Challenges, Methods. Berlin; New York: Springer, 2012.

[38] N. Müller-Wickop, M. Schultz, and M. Peris, "Towards Key Concepts for Process Audits – A Multi-Method Research Approach," in Proceedings of the 10th International Conference on Enterprise Systems, Accounting and Logistics, Utrecht, 2013.

[39] IFAC, ISA 330 The Auditor's Responses to Assessed Risks. 2009.

[40] IFAC, ISA 700 Forming an Opinion and Reporting on Financial Statements. 2009.

[41] IFAC, ISA 300 Planning an Audit of Financial Statements. 2009.

[42] N. Mueller-Wickop and M. Nuettgens, "Conceptual Model of Accounts - Closing the Gap between Financial Statements and Business Process Modeling," in Proceedings of the Modellierung 2014, Wien, 2014.

[43] H. Krcmar, Informationsmanagement. Berlin; Heidelberg: Springer, 2010.

[44] A. Rozinat, A. K. A. de Medeiros, C. W. Günther, A. Weijters, and W. M. P. van der Aalst, "The Need for a Process Mining Evaluation Framework in Research and Practice," in Business Process Management Workshops, 2008, pp. 84–89.

[45] A. Tiwari, C. J. Turner, and B. Majeed, "A Review of Business Process Mining: State-of-the-art and Future Trends," Business Process Management Journal, vol. 14, no. 1, pp. 5–22, 2008.

## Appendix A CPN Process Model

Figure 8 shows the original CPN model discovered by the MLPM algorithm. The corresponding FBPMN model is shown in Figure 3.
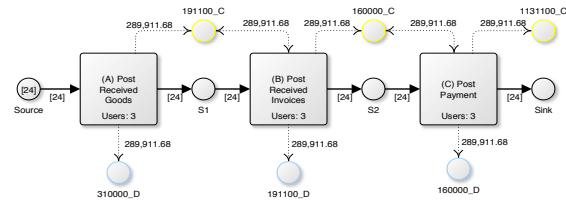


**Figure 8 CPN Example Process Model**

## Appendix B gnuplot Script

The script described in Listing 5 creates the type of model shown in Figure 4, Figure 5, Figure 6, and Figure 7 respectively. It requires a text file as input where data values are grouped in three columns. The first column represents the numbers for the financial accounts, the second the process model number, and the third the aggregated posted values for each financial account.

***#Plot Materiality Process Map***
```
unset key;
set title 'Materiality Process Map';
set yrange [*:*];
set ytics 1;
set xrange [*:*];
set xtics 1;
set zrange [*<-1:200000];
set cbrange [*<-1:*];
set xlabel 'Financial Accounts';
set ylabel 'Business Processes';
set zlabel 'Posted Amounts' offset 15;
set palette rgb 33,13,10;
set grid;
set dgrid3d 48, 216, 3;
set pm3d;
set xyplane 0.1;
set view 49, 346;
set colorbox user origin 0.9,0.1 size 0.03,0.8;
splot 'ExampleFile.dat' using 1:2:3 with pm3d;
```

***#Plot Materiality Heat Map***
```
set title 'Materiality Heat Map';
set view map;
replot;
```

**Listing 5 gnuplot Script for Creating Materiality Process and Heat Maps**