CONTINUOUS AUDIT DATA ANALYTICS AND INTERNAL CONTROL INTELLIGENCE

BY JIAHUA ZHOU

A dissertation submitted to the

Graduate School- Newark

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

Written under the direction of

Dr. Miklos A. Vasarhelyi and Dr. Helen Brown-Liburd

and approved by

____

____

____

____

Newark, New Jersey

October 2020

ABSTRACT OF THE DISSERTATION

Continuous Audit Data Analytics and Internal Control Intelligence

By JIAHUA ZHOU

Dissertation Director:

Dr. Miklos A. Vasarhelyi and Dr. Helen Brown-Liburd

Emerging data and technologies have fundamentally changed how humans and technologies work together and created new benefits and conflicts. The PCAOB calls for studies to obtain insights into data analytics and related emerging technologies in auditing (PCAOB, 2019). In response to the PCAOB, this dissertation explores how to develop internal control intelligence with Continuous Audit Data Analytics (CADA).

The first essay develops a framework, Social Construction of Technology (SCOT), useful for adopting emerging data and technologies in developing internal controls. This paradigm can explain and also guide how to adopt emerging technologies more coherently. The SCOT describes the institutionalization in four phases: problem definition, interpretative flexibility management, stabilization, and social construction. We derived six propositions from a participatory case study that we developed a rule-based Continuous Monitoring System (CMS) for a Procure-to-Payment process at a state university.

The second essay tests a prototype of the proposed analytics. The case study validates that CADA can enhance the rule compliance investigation and discover potential control risks that do not appear in the current internal control system. A shareable data platform, which digitizes the entire internal control system, acts as the function to fit CADA to the Committee of Sponsoring Organization (COSO) framework. The proposed analytics holds its advantage after embedding Audit Data Analytics in the CMS. Theoretically, this arrangement provides flexible analytical techniques to handle domain constraints to extract relevant intelligence from other sources around this domain. Specifically, the study demonstrated two combining analytics. First, if management already has clear business rules to manage risks, the combination of prescriptive analytics and diagnostic analytics can ensure exhaustive monitoring in these areas. Second, suppose management

has uncertainties about control risks. In that case, the analytics discovers potential risky controls by combining a risk test with the result from the diagnostic analytics.

The third essay intends to manage ethical issues for algorithmic audit analytics. The interaction between the emerging technology and audit context intensifies the uncertainty of the consequences of CADA. The study explores the root cause of potential ethical issues based on the unique feature of algorithmic analytics components. We theorize eleven ethical dimensions from the technology level, artifact level, and application level. Then we construct a four-phase ethics assessment to evaluate and deal with these moral problems. The four stages include the series assessment about the data source, the training data selection, the algorithm traceability, and the output interpretation.

The dissertation contributes to the literature from the following four points. First, we develop a theoretical framework to avoid the technology usage trap and promote CADA in auditing. Second, this study guides the development of flexible combining analytical schema. It exemplifies how to embed Audit Data Analytics schema in the CMS to provide audit evidence for assessing internal controls. Third, the proposed CADA schema provides enhanced auditing performance. Auditors can use the enhanced analytic schema to implement SOX404. Last but not least, the dissertation offers a practical roadmap to solve ethical issues in data-driven CADA.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help of dissertation committee members, and the professors who offered many bits of help for sharping my critical thinking and training my serious research attitude.

Above all, I would like to give my special thanks to my dissertation advisors, Dr. Miklos A. Vasarhelyi and Dr. Helen Brown-Liburd, for your continuous encouragement, trust, and patience. I learn how to articulate research ideas in a simple but clear way. Your academic spirit inspired me and helped me to distill research ideas. You have always been there to provide your help and support.

I would like to express my gratitude to Dr. Hussein Issa for your advice, caring, and patience. You spent a lot of time on the project and the development of the dissertation. Your always-helpful enthusiasm trained me on my career attitude. The appreciation also goes to Dr. Andrea Rozario. Your help gave me encouragement and confidence.

I also would like to express my gratitude to Dr. Alexander Kogan and Dr. Michael Alles for teaching me to have a critical mindset for research, and your advice and helpful suggestions.

Finally, I would like to thank my wife, Daisy, and my daughter, Candy, for your endless love and support throughout my doctoral studies. Your encouragement and patience are the vital energy to support me in going forward for my academic dream.

Content

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

The PCAOB calls for studies to obtain insights into the application of data analytics and related emerging technologies in auditing (PCAOB, 2019). In response to the PCAOB's call for research, this dissertation explores how to enhance audit data analytics in assessing internal controls. The AICPA (2017) defines audit data analytics as "the techniques that can be used to perform a number of audit procedures (i.e., risk assessment, the test of details, substantial analytical procedures, and informing an overall conclusion) to gather [audit] evidence." Thus, Continuous Audit Data Analytics (CADA) is coined as the techniques that apply Continuous Audit methodology (Vasarhelyi and Halper, 1991), which is based on "audit by exception" at the transactional data level on an ongoing basis, to conduct auditing and collect audit evidence. In other words, CADA is a method to embed separated audit data analytics throughout the internal control system to enhance its overall performance. The beauty of this embedding mechanism is to make each analytics artifact work with its advantage in its most suitable area to improve performance.

According to the definition, CADA is a set of dynamic artifacts with three essential components: emerging technologies, auditing, and audited organizations. Thus, there exists a potential theoretical area concerning the integration of the three parts. This integration can enhance the CADA's overall performance.

This research's subject is the Internal Control and IT-supported internal control system (IC, or ICs, means specific internal control activity or activities. The paper will specify the "internal control system" if we mean the whole system.). The nature of internal control is a complex process (COSO, 2013). This process is defined and affected by an entity's board of directors, management, and other personnel. Moreover, it is designed to help management achieve the following objectives: effectiveness and efficiency of operations, reliability of financial information, and compliance with the applicable laws and regulations (COSO, 2013, AU Section 319, 2010). Capable ICs reduce the risk of assets loss and help ensure high-quality decision-making information and reliable financial statements. IC is essential to the organization's operation. Effectively implemented ICs can help streamline processes and increase operational efficiency. Besides, IC can mitigate potential risks and prevent fraud. Both external regulators and the accounting profession emphasize the importance of auditors to gain a broader understanding of an organization's operations to perform IC and related risk assessment (SOX 404, PCAOB/AS2201, SAS99, SAS128, COSO, ISA 315). Also, research questions in management control and internal control are beginning to overlap with an expanding audit scope (Cohen et al., 2017). Both auditors and managers have pressures and motivations to explore practical approaches to gain a broader understanding of an organization's operations and improve internal control quality.

CADA's objective is to provide data-driven solutions to improve auditing performance, which may require a massive amount of workloads via traditional audit methods. Auditors can spend less time gathering, correlating, formatting, and summarizing information (AICPA, 2020). CADA is far beyond a simple tool. It includes a set of analytical artifacts that have interactions with the audited institution and the related stakeholders. A series of constraints need to be managed for a data-driven auditing solution (Cao et al., 2010). CADA cannot work independently, and it needs managers' guidance because it cannot see the big picture and assess moral problems. The data-driven solution needs to handle the corresponding foundations, context, algorithms, and business hierarchy.

The internal control system, technically, is a set of rules, policies, and procedures that an institution wants to increase efficiency and strengthen adherence to policies (AICPA, 2020). IT-supported internal control system requires using data to delegate these rules and policies and explore insights via emerging data and analytics. More business operations and accounting work have been digitized. These processes can feed the necessary data to make analytics work. The IC's cross-departmental feature offers opportunities to utilize data from different sources to support enhanced analytics. Furthermore, emerging technologies can already deposit human expertise via suitable database techniques. They can integrate new findings with existing human intelligence and experience. Thus, the internal control system's complexity simultaneously also offers auditors extensive opportunities to explore ubiquitous data in other areas to explore audit evidence.

We define Internal Control Intelligence as an IT-supported internal control system with three features. First, the system has a closed-loop mechanism to accumulate domain knowledge (A closed-loop system is designed to automatically adjust the desired benchmark condition by comparing it with the actual condition. It achieves the adjustment by generating an error signal that is the difference between the output and the reference input. In other words, a "closed-loop system" is a fully automatic system in which its control action being dependent on the difference between the reality and the benchmark). Moreover, this mechanism can utilize audit data analytics to handle emerging and inherently risks. Second, the system is modular-based, and each module can learn with each other. Last but not least, the system has an interaction mechanism to integrate human expertise and artificial intelligence.

From a design perspective, an ideal IT-supported internal control system affords the management understanding of all the controls in every business working flow. It also sets up effective rules in each control with suitable emerging technologies. To achieve this goal, we need a better understanding of (1) how to determine all of the necessary controls and related control rules, (2) how to efficiently adopt suitable innovative data and technologies to monitor each control, and (3) how to manage ethical issues for data-driven solutions. This paper explores how CADA can develop internal control intelligence by integrating and improving the three components.

The exploration of suitable emerging technologies in the internal control system needs to fit appropriate technologies with specific domain context. A framework is necessary to handle the domain constraints and harness emerging data and technologies. Social Construction of Technology (SCOT) can act as the theoretical basis of this framework because this paradigm can explain and also guide how to adopt emerging technologies more coherently. This study develops an adoption framework following the four-phase institutionalization procedure: problem demanding solution, the solution's introduction, the stabilization of the solution, and the solution to social construction. With the frame's guidance, we developed an integrated modular-based schema to embed separated analytics artifacts in the entire system to develop internal control intelligence (Cao et al., 2008, 2010). Thus, we consider CADA to evaluate internal controls as an actionable knowledge discovery (Cao et al., 2008). It can help gather audit evidence or actionable business policy using ubiquitous data sources. CADA has "social" and "technical" sides. It requires the involvement and confirmation from auditors and managers to add information accountability and reliability.

Also, technically, we need a link to bridge many analytical artifacts to integrate these objects to achieve CADA's enhanced performance. In data science, the data value chain is used as a conceptual framework to describe where data is identified, acquired, processed, stored, analyzed, and finally utilized by decision-makers to add value (GSMA, 2018). CADA needs to manage the entire chain from data generation through data-driven audit evidence. A shareable data platform can bridge the Continuous Audit methodology with the COSO framework. The goal is to digitize the internal control system as a database and transform data with the structure of ICs to develop internal control intelligence.

From a system perspective, the business process is a container with a control function for the process steps that the control rules regulation. A shareable agile database can feed data to all controls to exert continuous monitoring. The database requires representing the COSO framework if management sets necessary controls in the internal control system. Thus, following this logic, the study breaks down seamlessly the system into processes and then sub-processes. Within each end sub-process, we express all of the control functions as a series of “IF-THEN” rules. The center of the rule formalization includes the object, users, and control activities of each control rule. The three elements, theoretically, can express all of the internal control activities.

The first essay develops a framework to harness innovative data and technologies to enhance CADA's overall performance. As a highly regulated area, auditing requires accountability and reliability of emerging data and technologies. Auditing needs to find suitable technologies to establish ICs. The SCOT four mechanisms provide a coherent strategy to adopt emerging technologies. The strategy seeks to harmonize stakeholder interest claims with the emerging technologies' peculiarities.

Specifically, the study examines the institutional function of IT-supported ICs under the SCOT. CADA needs to customize suitable analytics artifacts in the internal control system to maximize the organizations' ICs by integrating its social part and technical part. This process is a co-construction process; the CADA project requires a data-driven monitoring system to accommodate itself to the managers' decision-making behavior. We initiate a pilot study to examine the institutional values of a rule-based Continuous Monitoring System (CMS) for a Procure-to-Payment (P2P) process at a state university. Following the four mechanisms, we document this data-driven audit analytics project's design and implementation process. We verify the strategic roadmap using the data from the P2P project, including the P2P transactional data, weekly WebEx meeting memos, emails among the team, and other relevant documents.

The second essay tests a prototype of the proposed analytics schema. We validate that CADA can enhance the rule compliance investigation and discover potential control risks that do not exist in the current internal control system. A shareable data platform logically fit CADA to the COSO framework. Thus, we propose an agile data platform that digitizes the entire internal control system by the users and activities of all control actions. The ITGI (IT Governance Institute) and the ISACA (Information Systems Audit and Control Association) developed the CobiT (Control Objectives for Information and Related Technologies) framework to fuel the implementation of SOX with IT technology. They developed the fifth version of the CobiT framework since 2016. Guided by CobiT 5, we break down the internal control system into nine processes, and 164 sub-processes arranged by their control-objectives. We apply the business rule approach to transform all 164 control-objectives into three types of data attributes (object, users, actions). The final stage of data preparation is transforming raw data into actionable attributes, which can directly act as the conditions in the "IF-THEN" rule expression. Then, this domain-based database includes all characteristics of the users and related control actions. The data structure can exactly meet the requirement of internal control intelligence exploration.

From a solution perspective, the proposed analytics, theoretically, provides flexible analytical techniques to handle domain constraints and extract relevant intelligence from other sources around this domain (Cao et al., 2008). Specifically, we demonstrated two combining analytics. The first is the combination of prescriptive analytics and diagnostic analytics (Tschaskert et al., 2016). The rule-based scoring system is diagnostic analytics by using the "IF-THEN" logic to investigate the compliance of each transaction. If the management already has a clear understanding of risks and sets up clear rules to manage risks, this analysis ensures exhaustive and continuous monitoring in these areas. Data visualization is descriptive analytics, which provides a straightforward understanding of data. Thus, this combining artifact acts as an intelligent decision-supporting component in the Continuous Monitoring and Control System (Allies et al., 2006).

The other tests are to discover potential risky controls by combining a comparison with the result from the rule-based-scoring-system. The initial idea is to extract actionable insights by using the knowledge learned from the adjacent areas of the internal control system. We selected control attributes that are tightly related to the transactions with high violation scores from the rule-based-scoring-system. Then we apply Multiple

Correspondence Analysis to reduce the attributes dimensionality and develop control rules with the final attributes. The analytics recommend the regulations as necessary when the new-rule-related transactions have a higher violation score than the benchmark group's average score.

The third essay intends to manage potential ethical issues for algorithmic audit analytics. The ethical issue is defined as the technically common and legal problems but potentially damage the algorithmic function and the related stakeholders' interest. This problem is easy to be ignored in practice but has potential negative impacts for CADA. We adopt a futuristic approach to take proactive actions to manage the ethical issues at the beginning of the design. Moreover, it also concerns the consequential ethics that are highly related to the business environment. The interaction between emerging technology and the audit context intensify the uncertainty of the consequences of CADA. The study theorizes eleven potential ethical dimensions from the technology level, artifact level, and application-level based on algorithmic analytics' unique feature. Then we construct a four-phase ethics assessment to evaluate and deal with these moral problems. The four stages include the assessment of the data source, training data selection, the algorithm's traceability, and the output's interpretation. The assessment guides a detailed discussion about handling "ethical laziness" to decrease the evidence's entropy.

The study contributes to the literature from the following four points. First, it theorizes an adoption framework to promote the use of emerging data and technologies in auditing. Second, this study guides the process of developing an agile analytical schema. It exemplifies how to embed Audit Data Analytics in the Continuous Monitoring and Control System to provide evidence for the assessment of ICs. Third, the study demonstrates how CADA can provide qualified enhanced auditing performance. Auditors can use the enhanced analytic schema to implement SOX404. Last but not least, the dissertation offers a practical roadmap to solve the ethical issue in algorithmic audit analytics.

Chapter 2. Harnessing Emerging Data and Technologies to Develop Internal Control Intelligence

2.1. Introduction

This essay explores how to fit emerging data and technologies with auditing tasks to improve internal control effectiveness and efficiency. Specifically, we want to examine how communication between IT, auditors, and management impacts the emerging data and technologies adoption. With digitalization, robotics and data engineering are everyday gaining more momentum in organizations. These emerging technologies provide organizations opportunities and challenges to develop effective and efficient internal controls. Leveraging analytics and robotics are front-burner priorities for internal controls and related business intelligence. Auditors are increasingly aware that businesses are becoming more data-driven. Not utilizing this data can be detrimental to the proper evaluation of risks and controls and, more importantly, meeting stakeholder expectations (Teammate, 2012; KPMG, 2015). The prevalent emerging data and technologies provide the potential to enhance the internal control system's quality by embedding controls in the automated systems. However, several surveys about the importance of technologies show that emerging data and analytics are not entirely accepted by most organizations (AuditNet, 2012; EY, 2014; KPMG, 2015). Emerging technologies and automatic monitoring still have many hindrances and therefore stay a starting stage. Hence, we need a close investigation about how to harness the power of emerging technologies.

A simple answer to data analytics sparse application is that the emerging data and technology are far from a simple tool. In other words, the emerging technology is of no particular value to the firm until its function is to be utilized appropriately. The technology must-have functionality coherence with other resources, including employees, culture, and infrastructure; otherwise, it is waste. Moreover, it even has negative impacts on organizational hierarchy if the institutions cannot harness the power. Organizations tend to display institutional inertia toward introducing emerging technologies if they perceive the emerging functionality is hard to manage. Thus, organizations need theoretical guidance to, in a safe way, integrate technologies into the internal controls. “[T]he key challenge that organizations must overcome arises not from the emerging technology, but [from] its adoption.” (PwC, 2019)

Emerging data and technologies have fundamentally changed the ways that humans and technologies work together, creating both new benefits and conflicts. It also can lead to new risks that may not be adequately addressed by the current internal controls. Based on this background, the PCAOB (2019) calls for studies on observing the utilization of emerging technologies in auditing. Managers need to adopt a mentality oriented toward accessing and analyzing “data in the business.” Many managers currently do not understand what to do with analytics, especially regarding a long-term strategy to transform internal controls using a data-driven function. One of the biggest challenges for organizations that do not start the journey toward using analytics is to understand where to begin.

Academics developed viewpoints to harness the power of emerging data and technologies in the auditing and accounting area. However, the prior literature has paid enough attention to the “what” question: what the impact of emerging technologies on organizations is. However, the literature pays little attention to observing how the interaction of IT, auditors, and audited institutions can impact auditing performance. The study needs to consider practices correspond to “the way we do things around here” or “shared routines or behaviors including traditions, norms, and procedures for thinking, acting, enduring things” (Whittington, 2006). Social Construction of Technology (SCOT) is one pioneering work of technology adoption (Pinch and Bijker, 1984). It has introduced four mechanisms to understand the social foundations of innovative technology design and development. This paradigm can explain how emerging technology operates and is operated in a complex system. It also helps to observe how the interaction of technologies and humans impacts the institutionalization process. In other words, The SCOT offers a frame to observe how the institutions manage the perceptions of the emerging functionality. It also helps to explain how stakeholders articulate their interest claim on the changes of emerging functionality. Thus, the SCOT can guide how to adopt emerging technologies more coherently by finding the fitness of the technologies with the institution.

In the case of IT-supported internal controls, the SCOT helps articulate the relationship among agents (including auditors, vendors, managers) and emerging technologies. With this guidance, we can examine practical activities (including institutionalized analytics schema) and their interpretations (including how to define exceptions and institutional rules) by using the four mechanisms. In the prior auditing and AIS literature, the processes of adopting emerging technologies were often considered separately, rather than jointly. Research has mostly focused on technology adoption and the implementation of its mechanic functions. However, there is less work that has attempted to study the integrative processes. There is a need for a better conceptual understanding of the connections among the technology, humans, and the social workforce in the evolution of emerging technology adoption.

Based on the motivations described above, we initiated a pilot project to design a rule-based Continuous Monitoring System (Allies et al., 2006) for a state university's procure-to-payment process. This pilot study's team has four different groups: the procurement department, the internal audit department, the human resource department, and the data analysts. As the principal designers in this pilot study, we can observe the entire procedure from project initiation through project adoption. We notably observed how different stakeholders claim their interests by articulating the technical requirement in the algorithm selection, data preparation, and analytical results deployment. The data was collected from four sources: institutional documents, transactional data, the memoranda of weekly design WebEx meetings, and daily conversations with relevant stakeholders. We collected data from a simultaneous process of taking actions to complete a specific task and doing research. These two things are linked together by critical reflection. Thus, the evidence may have its strength in the field studies because it can avoid response biases from self-reported interviews (Keeney et al., 2010).

This essay develops as follows. Following the literature review, we develop the assumptions of the argument. The case study with derived institutional values is introduced next. The fifth section develops the six principles of the SCOT framework. We conclude the essay with confirmation of the rationale of this study's method.

2.2. Literature Review

The research question is about harnessing emerging data and technologies to fit and optimize the organization's internal control system. The related literature review has two parts: (1) applying emerging analytical tools in the development of internal controls, and

(2) social construction of technology and its four mechanisms.

2.2.1. Internal Control, Continuous Auditing, and Automatic Control

Auditing Standard No. 2201 (PCAOB, 2017, pp.131) defines that IC over financial reporting includes those policies and procedures that "(1) pertain to the maintenance of records that, in reasonable detail, accurately and fairly reflect the transactions and dispositions of the assets of the company, and (2) provide reasonable assurance that permitted transactions are recorded as necessary."

Information technology can add fuels to improve the performance of ICs. Technologies automate processes, measurements, and some controls. IT also changed the organizations' design, including the formation of alliances and related process redesign and information sharing. Many ICs can be automated, including sensing and measuring the world's state, summarizing measurements, identifying exceptions, and the like (Kinney, 2000). The past three decades witnessed significant progress with the implementation of IT. Continuous Auditing (CA) is a significant research branch to apply IT and innovative technologies in the auditing area. Vasarhelyi and Halper (1991) first introduced CA in a project at AT&T Bell laboratory to develop a monitoring tool in an online IT environment. The rationale is to provide more timely assurance by continuously monitoring a company's entire transactional data. CA is designed to measure and monitor large systems, drawing key metrics and analytics into a workstation environment. "This methodology, reflecting the evolution of technology to online, real-time systems, has had slow but progressive adoption both in practice" (Vasarhelyi and Halper, 1991; Vasarhelyi et al., 2012) and also appeared in professional guidance (CICA/AICPA, 1999; ISACA, 2010). This methodology's assumption is based on the concept of "audit by exception," where deviations are flagged as alerts and forwarded to the responsible parties for closer investigation.

Rezaee et al. (2002) pointed out that CA requires auditors to develop client-specific IC templates to (1) evaluate the adequacy and effectiveness of the internal control structure, (2) assess inherent control risks, and (3) provide a detailed set of audit tests. These IC templates can perform electronic testing of sophisticated controls, including firewalls, authentication, passwords, and sensitive information encryption. CA developed a series of novel analytics schema to improve the effectiveness and efficiency of the IC. For example, Kogan et al. (2014) proposed Continuity Equations to set up an adaptive benchmark to detect exceptions in a business purchasing process. A big problem with CA is alarm overloading. Thus Issa et al. (2014) developed "exceptional exceptions" to identify and prioritize the high-risk exceptions when the automated system flags many anomalies.

Many stakeholders exert impacts on internal controls. Few papers have observed how these stakeholders construct and develop ICs with the interaction of emerging data and technologies. The literature needs a new scope to find how the communication of IT, auditors, and management impacts the development of ICs. The SCOT provides a pragmatic scope to observe the communication and interaction between stakeholders and technologies.

2.2.2. Social Construction of Technology and Its Four Mechanisms

The adoption of emerging technologies has been a research topic for more than a half-century in institutional theory. The theoretical background of emerging technologies implementation is resource dependence. Organizations are open systems. They need to actively engage interdependent relationships with their environment to acquire the necessary resources for their survival (Aldrich and Pfeffer, 1976). Facing uncertainties and ambiguity, organizations make technology development efforts to respond to environmental pressure. With more and more available data, organizations can improve their decision-making and risk management by utilizing emerging technologies.

Social Construction of Technology (SCOT) is pioneering in designing and developing emerging technologies (Pinch and Bijker, 1984). After Pinch and Bijker's seminal work, they integrated some subsequent works (Bijker, 1992, 1995a; Pinch, 1996; Pinch and Trocco, 2002; Pinch, 2008) to introduce four mechanisms to understand the social foundations of the application of emerging data and technologies. The four mechanisms can be right fitted into the four stages of technology adoption: initiation and background analysis, introduction, implementation, and institutionalization. The mechanisms include: (1) define existing problems and the relevant social groups, (2) manage interpretative flexibility, (3) closure and stabilization, and (4) social construction. With this problem-solution clue, the following literature review is arranged with these mechanisms.

Mechanism 1: understand the posed problem and the relevant social groups

The resource dependence theory argues that the impetus to search and apply emerging technologies is to extend the limited “resource” to handle the environment (Pfeffer and Salancik, 1978). The starting point is the problem development. Thus, at this stage, the “relevant social groups” (Pinch and Bijker, 1984) play a vital role in developing new technology. The relevant social groups share a particular meaning of the functionality of the new technology. This meaning can then be used to define how artifacts develop along specific paths and may crucially impact the final social-constructed object (Berg, 1998).

When facing uncertainties from the environment, the relevant social groups (the stakeholders of the technology) seek to manage the problem with technologies. This process requires them to define the technologies' functionality for solving the raised issue. Both humans and the technologies remain active according to the logic of their realm: technology performs mechanical functions, and humans interpret and articulate the variants that continually spring up (Berg, 1998).

Internal control-related auditing is in a highly regulated area. It has peculiar social groups for the adoption of emerging technologies. The relevant groups include government agencies (i.e., SEC, PCAOB), professional associations (i.e., AICPA, COSO), auditors, managers at all levels, and staff in almost every business processes. Thus, an IT-supported internal control system needs a "co-construction" (Fujimura, 1996). As Bijker (1995) claims, the SCOT must "figure out a way to take the common evolution of technology and society as our unit of analysis (P. 10)."

Mechanism 2: Manage the interpretative flexibility of the new artifacts

All newly constructed technologies are interpretative flexible; this is an important feature to understand how technology is socially constructed. "One of the key components of interpretative flexibility is that the artifact's actual functionality is not determined by its technological properties, but rather by the meanings attributed to it by relevant social groups" (Leonardi, 2013, pp. 72). Thus, in the process where a designer develops and applies an artifact to solve the problem, the object "works" only if it can be understood to "solve" the problem (MacKenzie, 1996). Thus, relational ontology is the premise to understand how technologies become socially constructed and evolved. The foundation of "things" is first and always a nexus of relations (Slife, 2004). Qualities, properties, and identities do not exist independently or inherently "inside" a substance; instead, these characteristics depend on how, when, and where they are related to each other (Emirbayer, 1997).

The interpretative flexibility emphasizes the interaction between stakeholders and emerging technologies. The technologies are highly consequential through human action, and human action gives technology meaning. Berg (1998) claimed four criteria for the successful application of computer supported cooperative work: (1) the technology has new competence; (2) users can perceive the usefulness of the technology; (3) the application of technology has an equality criterion, which means that a system cannot reduce the workload of some at the expense of others; and (4) the adoption of technologies has cultural and formal languages to communicate with relevant stakeholders.

The interpretative flexibility can cause essential institutional inertia to hinder the application of emerging technologies at this stage. Generally, intricacy and opacity are two interrelated processes that can lead to inaction. Intricacy refers to the number of steps in organizational changes and the complexity of the change (Carroll and Hannan, 2000). The more complicated things are that need to be changed, the more difficult reform will be. Opacity means a "black box" that cannot be explainable. Suppose stakeholders cannot fully understand the functions of the selected technologies and what will change with emerging technology. In that case, they are mostly unable to support the change. Thus, the adoption team must make efforts for a consentaneous understanding of the selected technology. This goal requires that emerging data and technologies need to be accountable and interpretable. Furthermore, the adoption plan requires emerging technologies to work without mainly changing the organization's current operating mechanism.

Mechanism 3: Stabilization and closure

The third mechanism is related to the second: All newly created emerging technologies should reach a stabilization and closure state before becoming social-constructed technologies. As Pinch et al. (1984: pp. 426) demonstrated, “closure in technology involves stabilizing an artifact and the ‘disappearance’ of problems.” The stabilization means that technology is accepted in the organization. After the iterative intergroup negotiation, the artifact stabilizes the meaning from the interpretative flexibility. This consensus signifies that artifacts reach a state of closure. They demonstrated two avenues of stabilization. The first is rhetorical closure, whereby the dominant group defines and emphasizes its finality. The second is closure by redefining the problem, and it is an on-going procedure (Covaleshi et al., 1988).

Barad has an excellent argument about the notion of “stabilization” and “becoming”: “Agencies are not attributes (of either humans or technologies) but on-going reconfigurations of the world” (Barad, 2003, pp.822). For example, the tree rings demonstrate “the sediment materiality of an on-going process of becoming” (Barad, 2007). Another example is the result of a Google search. The information obtained with a Google search done today will shape research practices differently from what the Google search had done last month or will do next week. Furthermore, in certain circumstances, such differences may be quite consequential. Mazmanian et al. (2014) studied how arrangements that produce active forms of agency emerge in on-going work by observing the process of reconfiguration of how the cosmic craft is understood and represented to those charged with its care and maintenance.

Thus, human action shapes technology by defining its function and explaining its output (Leonardi, 2013). The adoption of technology needs to understand how technology is embedded in its social context. Barad (2007) stated that discourse must be related to specific forms and in particular times and places. It is not a separate or static entity, but dynamically produced-in-practice. Therefore, the closure of emerging technology is often periodic and needs to have an on-going refinement.

For example, Orlikowaski and Scott (2012) observed the TripAdvisor, an online hotel evaluation system. They found that shifts in how guest feedback, over time, is accepted in practice, from comment cards to online reviews, are producing different guests, different hoteliers, and various hotels. Moreover, the performance changes, in turn, are serving to reconfigure the hospitality industry. Thus, the stabilization mechanism requires that the emerging data and technologies have a self-adjustment mechanism to adapt themselves to emerging risks and uncertainties.

Mechanism 4: Social-constructed artifacts

The socially constructed technologies include social structure and technological structure. Bijker (1995, pp.123) agrees that technological frames are not characteristics of systems or institutions but exist between actors. The framework elements can impact the interactions of the relevant social groups and lead to the attribution of meanings to the constructed artifacts. Different stakeholders have their specific technical requirements in each procedure of implementing emerging data and technologies. For example, auditors and regulators emphasize the algorithm's interpretative ability, and management may require accuracy. Thus, a technological frame constitutes the shared structure of interpretation of an artifact among members of a relevant social group (Klein and Kleiman, 2002).

Stinchcombe (1965) argued that in attempting to overcome their "liability of newness," organizations built up a stock of resources and processes that locked them into specific structures and rules. The defined routines and technologies are the infrastructure of the interaction of human and nonhuman agencies. Nonhuman entities achieve this closure through their mechanic function (Barad, 2003; Pickering, 1995).

2.3. Assumptions Formulization

This section articulates six necessary strategic actions derived from the literature to highlight how to adopt emerging data and technologies developing internal control intelligence strategically. Internal controls are a process. This "process" lens offers fitted perspectives from the SCOT to observe and study emerging data and technologies in ICs: (1) its on-going character, (2) its interaction, (3) its embedment in social-political contexts, (4) its relation to the capabilities of artifacts, (5) its dependence on shared practical understandings, (6) its refinement capacity to emergent risks, and (7) its enactment of social structures (e.g., their generation, reinforcement, renewal, and transformation) through everyday action (Leonardi et al., 2010). Based on these institutional features, we develop the following six assumptions for adopting the emerging technologies in ICs, and the following participatory case study can provide evidence for these assumptions.

2.3.1. Mechanism 1: Define the existing problem and its related stakeholders

Assumption 1: Accountability and reliability are the logical starting point of emerging data and technologies' adoption. It requires identifiable data flowing within the analytics.

The first strategic action is to arouse the collective consciousness of the existed problem. Technologies' accountability can act as this fundamental function. The ubiquitously available data offers opportunities and challenges for the business to take advantage of the emerging data and technologies in their business world. As discussed in the introduction, there are many stakeholders in the internal control system. Generally, the SCOT considers four stakeholders and their related resources. The first stakeholder is related to economic resources. It requires the balance of cost and benefits of the technologies' application. This balance is called affordability (Leonardi, 2003). It emphasizes that the prospected solution needs to generate an economic gain. Thus, both regulators and management expect that the emerging data and technologies can provide an affordable solution to develop ICs. Economic usefulness is the first consideration. Notably, many critics for the enforcement of SOX 404 emphasize the massive cost for small institutions (Ashbaugh-skaife et al., 2009). The second is the stakeholders from the regulatory resource that considers regulation from the government and industry standards. For example, the Sarbanes-Oxley Act assigns strict requirements over the assessment of ICs. The COSO posted an explicit internal control framework to offer a professional guide on law compliance. The third is the stakeholders from technological resources.

Accountable and effective technologies are the precondition for all stakeholders when they are searching for suitable solutions. The prevalent digital transformation has made data ubiquitously available. Regulators need to enforce laws effectively with reliable analytics output. Management has limited willingness to change the existed decision-making routines and avoid modifying the existed business hierarchy. End-users care more about preventing power abusing from the technologies' opaque. Auditors must examine an electronic process's components about related business transactions, using utilities embedded in the system. The last stakeholder is associated with the cultural evolution of an organization. Relevant stakeholders need to adjust their perceptions about the emerging functionality in internal controls; accountability and reliability can strengthen the perception of its usefulness.

In the intergroup negotiation, the regulators and top management have more influence on introducing emerging technologies in IT-supported ICs. Institutions need to comply with SOX and its requirements over internal control development. Under this regulation, the profession emphasizes the necessity of the COSO's internal control framework and the CobiT objectives. Accountable and reliable results are the fundamental requirement of the regulations. Also, other stakeholders have the same technical demand for their task goals. Thus, data analytics' accountability and reliability can be a logical starting point to harmonize all stakeholders' interest claims. This requirement emphasizes a traceable data flow within the analytics.

2.3.2. Mechanism 2: Manage the interpretative flexibility of the new artifacts

Assumption 2: Co-construction can improve technologies' adaptability.

Assumption 3: Data infrastructure capability is the foundation of the institutionalization of emerging data and technologies.

The second strategic action is to manage the most venerable technology application stage and promote all relevant stakeholders' consensus. In discussing the emerging technologies adoption in ICs, the SCOT emphasizes that we need to focus on what technologies do and what they express, translate, and output (Robichand et al., 2013). This consideration means that the technologies are not only actors but also interpreters.

The emerging technologies do not pre-exist as an entity but are instead created and continually reconfigured through relations. The assumption is that ties are primary and objects secondary (Gergen, 2010). Everything that exists is thus always becoming as a formation of relational effects. This ontology offers a matched angle to observe the relations among the stakeholders, the analytics artifact, and how they intra-act with each other. We should not follow the internal mechanisms separately. It is good to consider the big picture of the relational process of technologies institutionalization: (1) the relations between the institutional pressures and the conventional technologies, and (2) the relationship between the technologies' mechanics and the associated social part. For example, it is necessary to articulate the analytics' specific technical requirement for regulators, management, auditors in the procedures of data selection, algorithm choice, and the interpretation of the results.

The interpretative flexibility management becomes easy if all stakeholders can find an agreeable solution to harmonize different interest claims. Accountability and reliability can help reach the stakeholders' consensus. This goal requires a shareable data platform as the infrastructure. The platform needs traceable data sources and transparent data flow. The expected database needs to meet three requirements. First, it can cover the completeness of the control rules for the entire internal control system. Second, it can be shareable for all analytics tools embedding in each process to enhance its analytical capability. Third, it has a transparent data flow for data preparation.

Under these two assumptions above, the regulators (e.g., SEC, PCAOB) expect to promote emerging technologies to decrease SOX 404 compliance costs. Management can have more confidence to have a useful management tool and simultaneously keep the existing decision-making routines and business hierarchy. Auditors expect to improve their working efficiency but avoid learning complicated IT techniques. End-users can relieve the worries to avoid power abuse via complicated analytical algorithms and achieve a fair institutional environment.

2.3.3. Mechanism 3: Technologies' stabilization and closure management

Assumption 4: The refinement of analytics is an on-going process with relevant stakeholders.

Assumption 5: Combining data analytics can act as a function to develop internal control intelligence.

At this stage, the stakeholders have managed conflicts over the selected emerging data and technologies. The consensus makes the adoption strategy shift into how to refine the mechanical capability of the selected technologies. The criterion of closure and stabilization of an IT-supported ICs is the development and the deployment of control rules. The project can be stabilized after the relevant stakeholders will use the analytical results in their decision-making. The rise of algorithmic practices in auditing and accounting raises critical questions about intervention, surveillance, accountability, and ethics (Scott and Orlikowski, 2012). From the perspective of discursive practices, the following questions are valuable to consider in tuning emerging technologies in ICs: (1) What algorithms are being manifested, and how in particular times and places; (2) What situated outcomes are being produced as a result of their performance; (3) What realities and their functional performance are being enacted in practice over time; and (4) How the different entailments of algorithmic phenomena play out and under what conditions do essential empirical and ethical questions have significant salience for our understanding of management.

The answers to these questions need an on-going refinement. In other words, an IT-supported project can only have periodic closure. It needs to have an on-going process to adapt to a new environment and derive new analytics features. Besides, the business process' digital transformation provides opportunities to run experiments to optimize business processes, so the on-going refinement makes analytics artifacts have technical support (Thomke, 2020).

From a solution perspective, the IT-supported ICs can embed much data-driven analytics in the internal control system to exert all of the control rules. Each analytical schema may have a different algorithm to make it suitable for a specific control activity. The emerging data mining techniques demonstrated many useful skills to combine different analytics to enhance analytical performance (Cao et al., 2012). This combination offers a variety of perspectives to observe potential risks and improve overall IC performance. The AICPA advocates that the data-driven audit analytics procedures for one traditional audit phase can provide evidence for multiple phases. For example, the output of audit analytics in the pre-engagement procedure can support auditing planning, auditing fieldwork, and conclusion with its data-driven evidence (AICPA, 2020). Thus, combining analytics (Cao et al., 2010) can act like a proper data mining technique in the domain of internal control intelligence.

2.3.4. Mechanism 4: Socially constructed IT-supported internal controls

Assumption 6. Ethics issues can cause institutional inertia to adopt emerging technology. It is necessary to manage the moral problems at the beginning of the design process.

The SCOT encompasses multiple potential underpinnings, and the particular foundation is a relational ontology (Latour, 2005; Pickering, 1995). The center of this foundation is that relations are more important than entities (Leonardi, 2013). The notion of SCOT emphasizes technologies in practice. It is integral, inherent, and constitutive, shaping the contours and possibilities of everyday organizing. Thus, the IT application can be more effective if the technology can fit its culture and hierarchy. Even though the technologies have robust mathematic logic, some users may potentially twist and alter what the technology "carries." Data analytics cannot assess moral or ethical concerns (AICPA, 2020). The ethical issues can hinder the adoption of these emerging technologies and needs to be managed well.

2.4.Application of the SCOT to a University's Purchase-to-Payment (P2P) Data Analytics Project

This essay's research question is about how communication among IT, auditors, and management impacts the development of emerging data and analytics projects and how to institutionalize suitable technologies to improve ICs. This question is a strategic matter and is related to a multiple of stakeholders. We initiate a pilot study to develop a Rule-based Continuous Monitoring System for a P2P process at a state university to obtain insights for this question. The one-and-half-year-length participatory case study allowed us to observe the entire procedure from the project initiation to its adoption. The often-used way to identify strategic actions is to discuss the decision situation with decision-makers. However, in this case study, we have a better chance to observe each stakeholder's statement and response to the adoption of the P2P analytics project. We collected evidence from the simultaneous process of taking actions to accomplish specific tasks and doing research. These linked two things can add practical value for problem solving. Keeney and Winterfeldt (2010) successfully used a similar method to extract the terrorists' values by examining their writings and verbal statements. Instead of interviewing decision-makers and stakeholders, we extract the values by capturing the stakeholders' behaviors in the weekly WebEx meeting, the organization's documents, and the project team's emails. Kunz et al. (2016) used the same approach to develop a strategy map of the Balanced Scorecard of a newspaper company. The coding approach keeps consistency with the method - "Integrating the grounded theory method and case study research methodology within Information System research" (Halaweh et al., 2008).

2.4.1. An Introduction to the P2P Project and the Data Coding Protocol

The institution in this pilot study is a state university located in the northeast of the US. It has several campuses in different locations. This university has more than 60,000 students from all 50 states in the US and more than 125 countries. The purchase scale is enormous, and this P2P process has its theoretical value also because of its frequency, amount, and variety. For this study, the author is the principal project designer. The university organized a cross-departmental project team, including the internal audit department, the procurement department, the human resources department, the information system service department, data analysts, and consultants. This team was a "co-construction" team (Fujimura, 1996), as discussed in the previous sections.

The primary timeline for the project was as follows: (1) identify data access (March 2017-November 2017), (2) define metrics (October 2017-December 2017), (3) develop benchmarks (January 2018-March 2018), (4) detect and analyze exceptions (February 2018-April 2018), (5) automate mechanisms (April 2018-June 2018), (6) refine the model to adapt to new environments (on-going).

We collected the data and coded the evidence from the following documents: (1) the university policy files and the relevant documents from the project initiation, (2) the memos of the weekly WebEx meeting that involves the project team, (3) the daily conversations and emails with relevant stakeholders, (4) we also observed the responses from the policy violators in the rules testing, and (5) the transaction data from four primary data sources, including the SciQuest order system, People Human Resource system, Oracle Data Warehouse, and some Excel spreadsheets from associated departments.

We keep using the four SCOT mechanisms as the structure to arrange the evidence about how stakeholders took actions to achieve their goals in the emerging technology's adoption. The evidence coding protocol includes four steps. The first is to identify an initial list of objectives and convert these objectives into a standard form of criteria. By taking this step, we can observe what goals and interest claims are essential for these stakeholders to adopt emerging technologies. It helps the decision-makers clarify the objectives within their power structure and interest claim. The second is to structure the objectives. Objectives structuring can clarify additional insights into the decision-making context. It uncovers the factual information and the causality of the means and objectives. The third is to create options and identify decision opportunities. This step critically evaluates potential alternative means of achieving the fundamental goals beyond the initial means objectives list. The final step is to build the means-ends objectives network (Keeney, 1994). This relation captures the values and interest requirements of a decision-maker concerning a particular decision.

2.4.2. Stakeholders' interest claim and relevant strategic actions in the P2P analytics project

This section demonstrates the SCOT framework's evidence about how the P2P data analytics project evolved based on the perspective of the four mechanisms.

Mechanism 1: Define the existing problem and its relevant social groups

The university's CFO initiated the P2P analytics project under the external and internal contradictions. The institutional pressure pushed him to explore solutions to improve the P2P's control effectiveness and efficiency. The external pressures mainly come from the regulators. Under Sarbanes-Oxley, the CFO of nonprofit organizations is required to sign the form 990 or 990-PF. Non-profit organizations must have effective and efficient ICs to improve these forms' accuracy and completeness.

Nevertheless, the university suffered some problems for the IC's effectiveness. The university's 2016-2017 fiscal year financial report had a qualified opinion from the KPMG. Thus, the university organized several chancellor unit representative meetings to explore the desired solution to enhance the P2P's control performance. Another ambitious goal of the CFO is to develop a data-driven P2P monitoring system that can lead the Big Ten public US state universities to add the university's reputation and leadership in the peer review.

The management cannot have enough experience to respond to risk-related problems by combining three different enterprise Information Systems and running data analytics. In the chancellor unit representative meetings on April 20, 2016, the management figured out " [N]o examples were provided to meet the proposed criteria" (Memo of the meeting, 2016). The university asked technical help from the Continuous Audit and Report Lab (CarLab) at Rutgers Business School. Simultaneously, the management hired a manager, who has data analytics experience in a nonprofit organization, as the project coordinator.

All stakeholders, including the internal audit department, management from all levels, and the human resource department, face the same problem to have an accountable and reliable solution to handle the overloading information. The same goal helps the project to get support from different stakeholders.

Mechanism 2: Interpretative Flexibility Management

This stage is vulnerable to adopting emerging data and technologies because each stakeholder has different considerations and worries that the new technology may hurt their interest. At the beginning of the P2P project, stakeholders only held limited expectations for the project. They took different actions to avoid interest hurts from the potential project. A common phenomenon is that the team participated in the project but with low enthusiasm. The reason for some relevant stakeholders is that they have very little knowledge of the expected data analytics artifact. For example, the team wanted manually extracted training data sets from the SciQuest order system. However, the associated staff extracted unidentifiable data types for three times, and these mistakes hinder the designers' progress. The communications have limited effects, and it looked like we communicate only because the job needs us to do that. The main issue is that it is hard to integrate different data sources as a standardized data platform. The data preparation and test were painful. The university has three segmented information systems: Oracle Data Warehouse, SciQuest Procure Order System, and People Human Resource System. We spent two months exploring how to integrate necessary data from the different systems with the end-users’ help. We successfully prepared the data dictionary for the data platform. The most challenging issue was that the university has no access to extract data from the SciQuest system directly. The internal auditing department had to request periodic transactional data from the software vendor. We can do nothing without qualified data.

Every stakeholder was considering the accountability and reliability of the project. The design team cannot guarantee accountability if the team even cannot solve the data issue, so we spent a half year on solving the data issues. In the stagnant time, both the newly hired project coordinator and the authors of this essay made efforts to demonstrate new evidence for the expected P2P Analytics project. The turning point happened when the analytics detected 961 transactions that violated the business rule that the invoice date must be after the purchase order date. This result got attention from the top management because it accurately captured the problems existing in the operations. After this turning point, the co-construction team was developed. One factor leading to this team is that the analytics project developed a robust domain-based data platform, which can ensure the analytics' accountability and reliability.

Mechanism 3: Stabilization and closure

This stage means that the stakeholders have achieved consensus about the selected technology’s functionality. In the P2P analytics project, stakeholders find that the project can benefit their decision-making and relieve their worries about the analytics’ transparency and accountability. This consensus stabilized the project and aroused the stakeholders’ enthusiasm to refine the functionality of the analytics. At this stage, both the human resource department and the procurement department responded much faster than the initial stage. With support from each stakeholder, we made significant progress with the project. The co-construction team developed 14 control rules for the entire P2P process with the cooperation among the internal audit department, the procure department, the human resource department, and the data analysts. The internal audit department initiated where the institution needs to set up continuous monitoring rules. The authors designed analytics algorithms to test the rules with the training dataset. Then the procurement department confirmed the test results, and some end-users participated in the confirmation. With help from the cross-department, the analytics algorithms achieved high accuracy.

The co-construction team also made several refinements for the Continuous Monitoring System. The refinement for this analytics system is an on-going process. Some interactive communications can be set up to improve its adjustability. The communication with the end-users improved the effectiveness of the parameters setting of some rules. For example, the first communication arrangement lets the end-users confirm and explain the detected rules violations. The end-users responded with the confirmation of the violations or explained why the algorithmic detected mistakenly. Technically, these arrangements act as the function to refine the benchmark and improve its internal control intelligence. Second, we used Tableau to prioritize some risks with the visualization dashboard, and by doing so, to improve the understandability of the communication. This arrangement helps engage all end-users to take advantage of the P2P analytics project to improve their working efficiency. Third, the co-construction team narrows down the analytics to tune the algorithms within some specific areas. For example, it is a time-consuming auditing task to check all Quick Orders. The Quick Order transaction is a procurement order pattern for some urgent operational purchase. Quick Order does not have to acquire the purchase approval from a higher-level manager. The director can decide these product purchases. The Quick Order needs to meet two criteria: (1) the transaction amount is less than $5,000; (2) the products need to be in the category within the university policy. The procurement department requests help to design some algorithms to improve this audit efficiency. The team used combining analytics by embedding text mining and IF-THEN rules analytics. The text mining required preparing a bag of words in the analytics. The co-construction team trained the samples and confirmed the results manually from the rule violators. This arrangement refined the algorithm setting a lot and engaged the end-users' participation and acceptance. An exciting story regarding the co-construction team is that one project member from the internal audit department was promoted as a senior manager at the procurement department because of her prominent performance. This example certifies that co-construction is an excellent way to improve emerging technologies' acceptance.

Mechanism 4: IT-supported Internal Control System Socially Constructed

The socially constructed technologies have two implementation structures, the technical structure and the social structure (Leonardi, 2013). The social part can impact its mechanical part because of the ethical issues of the analytical system. In the P2P analytics project, the design team had a keen consciousness that we need to manage these moral problems at the beginning of the analytical system's design and implementation. The ethical issues can be embedded in the system as parameters or some similar arrangement. We tried to relieve this issue by inserting the project's ethical considerations from the following four parts: data source, training data, algorithms selection, and output interpretation.

2.4.3. The Institutional Value Implications of the P2P Data Analytics

This section extended the evidence coding from the case into a general institutional scope. These derived values can have a general meaning to guide the adoption of emerging data and technologies in other contexts. The values demonstrate what activities can be crucial for the adoption and what approaches can help achieve these goals. Following Sweeny (1994), we derived the value list of P2P audit analytics. Each value listed was converted to an objective. Similar objectives were converged to one objective. After that, the team reviewed the value list several times to get the final results. The results are described by one overall objective and seven specific objectives, listed in Tables 2.1 below.

Understanding the decision-making objectives is an excellent way to find the motivations of these goals. We also can observe what actions the stakeholders took to achieve these goals; by doing so, we can have a coherent observation about the evolution of the emerging technologies. We organized all the derived objectives for four tasks: (1) combine similar objectives to categorize and compose a common general objective, (2) identify approaches to achieve these objectives based on the function of the value implications, (3) explore stakeholders' interest statement and potential institutional inertia for each objective and sub-objectives, and (4) derive value implications with relevant SCOT mechanisms.

As Table 2.1 below shows, the overall objective, in this case, was to improve the effectiveness and efficiency of the P2P monitoring system by using emerging data and technologies. To achieve this goal, we need to harmonize the relationship between relevant stakeholders and emerging data and technologies. The institutionalization process is to find suitable emerging data and technologies and fit them to develop internal control intelligence. Different stakeholders would have different interest claims and

Table 2.1: Institutional Values for the Data-driven Internal Controls

Objectives and sub-objectives	Approaches to achieving the goals	Stakeholders' interest statement and potential institutional inertia	The Relevant SCOT Actions
Overall Objective: Improve the effectiveness and efficiency of ICs	Embedding data-driven analytics in the internal control system by using emerging data and technologies.	*Regulators (SEC, PCAOB)*: enhance law compliance. *Auditors*: increase auditing efficiency. *Management*: increase operational efficiency with affordable costs.	
Objective 1: Find fitted emerging data and technologies to develop internal control intelligence.			
Sub-objective 1.1: Develop traceable algorithms.	(1) Develop "IF-THEN" rule-based algorithms;	*Regulators (SEC, PCAOB) and auditors:* audit evidence needs to be reliable. *Management*: "Black-box" like algorithm can increase the institutional inertia to reject the analytics.	Make algorithms understandable, and by doing so, to increase their acceptance.
Sub-objective 1.2: Increase output interpretability.	(1) Utilize visualization technology. (2) Promote communication among relevant stakeholders.	*Auditor and management*: visualization can increase the communication efficiency for the analytical results. *End-users:* visualization increases user-friendliness.	Make analytical results understandable, and by doing so, to increase the acceptance of data analytics.
Objective 2: Improve data infrastructure capability.			
Sub-objective 2.1: Develop flexible data architecture to	(1) Use meta-data driven integration approach;	*Regulators (SEC, PCAOB) and auditors*: all data sources can be traceable and testable. Doing so ensures the	Let data owners explain data traceability for all stakeholders and enhance

integrate data from different information sources.	(2) Let the data owner help find the transaction identifiers.	accountability of audit evidence.	the consensus of data sources' accountability.
Sub-objective 2.2: Develop a shareable data platform to feed relevant data for different analytics demanding.	(1) Develop analytical techniques that can let analytics artifacts can use analytical results for each other.	*Management*: the data flow within the analytics system can be interpretable; it is unnecessary to change the current decision-making routines. *End-users*: the data attributes are understandable.	Enhance the analytics' capability by taking advantage of the existed available data.
Sub-objective 2.3: Take advantage of ubiquitous data.	(1) Structure Big Data to capture IC behavior. (2) Combining different data sources to extend data availability.		Enhance the analytics' capability by extending data availability.
Objective 3: Develop effective control rules for emerging risks and inherent risks.	(1) Increase the cross-departmental communications; (2) Periodical tuning the existed control rules.	*Regulators (SEC, PCAOB) and auditors*: the system has enough control rules for each node of the entire internal control process. *Management*: the control rules are easy to operate and manage. *End-users*: the control rules are free from power abuse.	Improve all stakeholders' participation in the control rules development, and by doing so, to enhance the effectiveness of these control rules.
Objective 4: Protect the security of assets by improving the digitization of business operations.	(1) Fraud detection by tracing conflict of interest; (2) Assurance segregation of duties by testing the decision-making routines.	*Regulators (SEC, PCAOB) and auditors*: the data-driven solutions can define fraud and segregation of duties as automatic surveillance. *Management*: business policies and rules can be strictly monitored. *End-users*: the monitoring is fair and cannot be abused.	Require recording all necessary decision-making actions and make all of these activities digitized and analyzable.

Objective 5: Engage in total participation.	(1) Arrange a mechanism to listen to the voice of all employees.	*Regulators (SEC, PCAOB) and auditors*: the response from different perspectives can enhance its efficiency and manage biases. *Management*: false analytical results can be found fast. *End-users*: biased analytical results can be filtered by adequate notice.	Increase the collective consciousness that data-driven ICs are tightly related to the collective interests and need everyone's attendance and contribution.
Objective 6: Due diligence.	(1) Gain support from top management; (2) Get response and confirmation from the end-users; (3) Elevate the collective consciousness for operational weaknesses.	*Regulators (SEC, PCAOB) and auditors*: the system needs to be useful and reliable. *Management*: the system needs to be accountable. *End-users*: the system can be helpful and relieve the working load.	The confirmation from the end-users can promote the application of data analytics and accumulate internal control intelligence.
Objective 7: Ethical issues management.	Insert ethical considerations into the design and implementations of the data-driven analytics, e.g., rules parameters setting, training data selection, and data attributes selection.	*Regulators (SEC, PCAOB) and auditors*: The system's design is free of ethical issues embedding. *Management*: data-driven solutions are free of human biases. *End-users*: data analytics is free from power abuse.	Take a futuristic approach to predict and proactively manage the potential ethical issues at the beginning of the design.

worries about the technologies' emerging functionality. The adoption needs a goal-congruence process, like routine weekly meetings, and periodic tuning of the emerging functionality. As we introduced in the previous section, the design team held a WebEx-based weekly meeting for 16 months. We developed and confirmed 14 query-based control rules.

The SCOT emphasizes developing the algorithms' traceability and interpretability with a standardized data platform. The algorithm traceability and interpretability directly impact stakeholders' perception of the technologies' acceptance. No one is willing to put risky decision-making into an opaque machine. Boolean function-based "IF-THEN" control rules and visualization can be two technical features that improve the system's accountability and reliability. A standardized data platform can act as two essential functions to improve its institutionalization procedure. On the one hand, it can extend data availability and take advantage of ubiquitous data. On the other hand, it can improve the perception of the technologies' accountability and reliability with a transparent data flow within the analytics. The data platform also influences the algorithms' traceability and fitted architecture. Both the conceptual model and the meta-data driven approach can impact the construction of the data platform.

The evidence above from the case study is about preparing the technical part of the emerging data and technologies. It demonstrates that both the standardized data platform and the traceability of algorithms act as essential functions in the IT-supported ICs. We present a detailed discussion of these two value propositions in the following section.

After preparing the technical part, the institutionalization needs to fit the emerging functionality into the institutional context by harmonizing the relations with relevant stakeholders. The design and implementation of data-driven solutions need voices from all relevant social groups. Top management's support will give a clear signal to arouse the collective institutional consciousness to search for new solutions for the existing problems. The confirmation from the end-users is a direct and effective way to improve the technologies' adaptability. The data preparation needs experts' professional judgment for its relevancy and usefulness. The communications from all relevant stakeholders can manage the negative impacts of the "social" part and significantly improve the "technical" part's positive impacts. Specifically, effective control rules can bridge the "technical" part and the "social" part. It needs cross-departmental communication and periodical tuning. Ethical issues can be significant institutional inertia to adopt emerging technologies. It can be embedded into the system from the very beginning of the design. We need to consider the potential moral problems thoroughly in the design and implementation process.

In summary, the institutional values generation shows three essential institutional inertias. 1. Organizations do not want to change their existed business decision-making routines on a large scale. 2. The existed enterprise Information Systems wants to keep their independence. 3. All stakeholders resist that the transition from the traditional ICs into IT-supported ICs is in "Black box." In the following section, we extend the case to develop a theoretical framework with six principles to manage these institutional inertias to promote the application of emerging data and technologies to develop internal control intelligence.

2.5. Construction of a Framework for the Institutionalization of a Data-driven Audit Analytics Project

The case study provides a thorough insight into the initiation and development of emerging data and technologies. In this section, we extend the value propositions of the P2P project as a theoretical framework. As we discussed in the previous section, the impetus to adopt emerging technologies is operational contradictions from inside and outside. The pressures push top managers to put a coercive mechanism into the management system to propel the initiation of a better solution. This institutional evolution routine keeps consistent with the current literature, and also extends the SCOT scope.

Specifically, the forces come from resource dependence, reflected as four contradictions: efficiency, non-adaptability, misaligned interest contradictions, and inter-institutional incompatibility (Seo et al., 2003). Simultaneously, some successful cases from peer organizations accelerate the pressure for management (Wagner, 2010). At this stage, mathematical logic and calculative ability make algorithms achieve more effective decision-making than humans. Algorithmic objectivity can signify the organizations' contradictions and elevate the consciousness of contradictions (Mittelstadt et al., 2016) and accelerate the process of developing IT-supported internal controls. A suitable strategy to harness emerging technology's potentials is to take the "co-construction" procedures. Following the assumptions, we develop six propositions to guide the development of the IT-supported internal control system in a sequence of development stages with the SCOT framework.

2.5.1. Arouse collective consciousness in the initiate stage by defining the problem and engaging the problem-related stakeholders (the SCOT mechanism 1).

Proposition 1: Accountability and reliability are the logical starting point of emerging data and technologies' adoption, and it requires identifiable data flowing within the analytics.

In this initial stage, the adoption action's underlying goal is to arouse the collective consciousness to explore suitable solutions for the existing problems. The motivation for the implementation of emerging data and technologies in ICs comes from various stakeholders. The first task is to find a bridge to link the stakeholders' interest in the technologies' emerging functionality. The technology's accountability and reliability can act as the bridge because it can help all relevant social groups meet their essential requirement for the emerging functionality. If the technologies are reliable and explainable, the application can manage the technologies' intricacy and opacity.

The emerging technologies' algorithmic objectivity contributes to arouse actors' collective consciousness of the existing contradictions and signifies the institutional pressures. Seo et al. (2003) indicated that collective consciousness acts as an active catalyst to pursue a new solution when the institution suffered emerging problems. With the algorithm's accountability, the output of audit analytics can add more persuasiveness. Thus, audit analytics can accelerate developing an IT-supported ICs if, and only if, the project can ensure the accountability and reliability of the technology. This requirement emphasizes that we need to avoid using algorithms with complicated logic. As we illustrated in the previous section, Boolean function based IF-THEN rules can be a good option.

Technology may reshape organizations and institutions to conform to its logic better because technologies' mathematical logic can avoid human judgmental biases. Institutions have routes or logics embedded in operating routines, bureaucratic politics, norms, cultural beliefs, and social networks. The application of emerging technologies can clarify these logics and embed these rules into analytics. As a result, accountable and reliable technologies avoid human biases and execute these rules much better than humans.

2.5.2. Accelerate the process to achieve stakeholders' consensus by managing the new artifacts' interpretative flexibility (the SCOT mechanism 2).

Proposition 2: Data infrastructure capability is the foundation of the institutionalization of emerging data and technologies.

The primary goal is to persuade all associated stakeholders to achieve consensus about the selected emerging data and technologies' functionality in the introduction stage. Accountability and reliability is the logical starting point of the proposed theoretic framework, and data is the technological capability's foundation. All well designed IT technologies and algorithms would lose power without qualified data. Internal controls can be anywhere in the organization's operations. Data related to ICs can come from different Information Systems, or even external Big Data, and the data can be in a different format. A standardized and shareable data platform makes the potential emerging technologies more achievable than without the standardization. It can help restructure the entire IC process but does not require changing the current decision-making routine and the arrangement of the current information systems. This data platform helps relax the institutional inertia from the current political hierarchy. Also, more available data, such as legacy data, can be utilized in this data infrastructure to develop internal control intelligence.

Furthermore, because IC is a complex process, we often need to observe the effectiveness and efficiency of ICs based on an entire process, which often crosses several departments or institutions. The standardized data platform can integrate data from different sources and offers opportunities to run necessary experiments and explore how to improve the efficiency and effectiveness of ICs. These explorations usually have challenges when extracting data from different data sources, usually distributed in different information systems (e.g., SAP, Oracle, spreadsheets).

In practice, the "standardization" stays at the architecture level, offering flexibility for a potential opportunity for different operational systems. This solution also provides potentiality to structure Big Data and other un-structural data. Besides, combining analytics, which uses the result from other analytics, can lead to various benefits to developing internal control intelligence. However, different data formats may hinder the application of combining analytics. A shareable database can directly help to solve this issue. With a shareable database, the data-driven solutions can fuel external auditors to improve audit efficiency and effectiveness in internal control assessment.

Proposition 3: Co-construction can improve the adaptability of emerging technologies.

The nature of the internal control system is a cooperative system. The internal control intelligence needs all stakeholders to offer their support. Any slack behavior may paralyze the function of the controls. Thus, the design and development of IT-supported ICs need a co-construction team.

The following example from the P2P project can provide readers a vivid understanding of the co-construction team's effectiveness. In the training process, after the project team developed the analytical rules for the P2P process, we ran the designed algorithm to detect the transactions of 2017. Then the related management verified the analytical output. After that, we prepared a report and posted it to the related department for confirmation. In the report, we developed three columns in a spreadsheet for the rule violators to (1) confirm the violation or verify the result is wrong; (2) explain why this happened, and (3) provide comments about how to avoid similar abuses in the future. This simple arrangement collects precious qualitative evaluation data for the future. It improves the relevant staffs' enthusiasm to participate and make significant progress for the P2P analytics project.

2.5.3. Keep refining the technologies' mechanical functionality in the technologies' stabilization and closure (the SCOT mechanism 3).

Proposition 4: The refinement of analytics is an on-going process with relevant stakeholders.

An analytical function may have different performances in a different context. We need to adjust the technologies' mechanic functionality to fit them in a specific operational context and ultimately to maximize its usefulness. As Orlikowski et al. (2008, pp. 462) argue, "Relations and boundaries between humans and technologies are not pre-given or fixed, but enacted in practice." In the design and implementation process of the IT-supported ICs, we need the intra-action between "exceptions" and "benchmarks" to refine the automatic control system and improve its adaptability. The performance of models is a particularly relevant technology entailing "algorithmic configurations" (Callon and Muniesa, 2005), forms of work, and orders produced by the agencies of humans and technologies.

It is valuable to periodically reassess the following elements of the data-driven solutions in the internal control system: the manifested algorithms, the situated outcomes, the context of the analytics, and the algorithm's combination. These adjustments and new processes are necessary "as the functionalities of technologies are negotiated, shaped, bypassed, or undermined in the situation" (Kallinikos, 2011, pp.12). This reassessment process means to optimize the decision-making routine in the defined ICs. Feldman and

Pentland (2003, pp.97) define an organizational routine as “a repetitive, recognizable pattern of interdependent actions, involving multiple actors.” Organizational routines are concrete, empirical phenomena that can be viewed as social and technical assemblages (Pentland et al., 2011).

These “socio-digitized” structures “exhibit dynamics of their own that derive from technological capacities that enable specific patterns of interaction” (Latham and Sassen, 2005, pp.4). Technological information is an embedded element of institutional life. Thus it is crucially involved in the reconstitution of organizational reality in various novel forms.

Proposition 5: Combining data analytics can act as a function to develop internal control intelligence.

Combining analytics combines some analytics functions to achieve an enhanced analytics performance, like combining data source, domain expertise, and results from different analytics. This emerging data mining technique can act as the primary analytics function to develop internal control intelligence. First, it can extract useful knowledge by taking advantage of ubiquitous data. Many electronic activities can leave an action trace, e.g., event log data and timestamps. These automatically recorded data cannot be manually doctored. The built-in unchangeable feature can provide significant value to capture managerial behaviors. Combining analytics can make full use of the ubiquitous data and make them analyzable and useful. Second, this technique can let each analytics artifact help each other. For example, social network analysis can help find some group behaviors. This result can provide a big help for other analytics, like fraud detection analysis.

2.5.4. Manage ethical issues and improve the technological functionality after the technologies are socially constructed.

Proposition 6: Management of ethical issues needs to be embedded in the design and implementation of emerging technologies.

Information technologies can extend humankind's physical existence and the power of perception. This extension serves to forward an expansion of human senses and actions (Weiner, 1954). Data-driven internal control solutions have moral problems. Furthermore, the analytics result may be the basis of decision-making. Thus, this ethical issue can scale up its impact on the institution's political hierarchy. The moral considerations can cause large institutional inertias to hinder the adoption of the CADA project. Thus, it is essential to take proactive actions to manage these moral problems.

2.6. Discussion and Summary

With the demonstration from the value coding, we develop a theoretical framework, integrated with the six propositions, to guide the development of IT-supported ICs. We document the system's implementation process to observe how data-driven solutions can change institutional inertia and use emerging data and analytics to develop an effective and efficient system. The combined evidence illustrates how technology signifies the external and internal institutional pressures and how these pressures motivate the internal audit department to apply emerging data and analytics to relieve the loads. The design and development of the data-driven P2P solution face challenges from technical and social forces. This essay demonstrates six propositions to guide how to harness data-driven technology in the internal controls auditing area. The following chapter applies the six propositions to develop a prototype to assess ICs and develop internal control intelligence. To exemplify the theoretical framework to adopt and maximize IT-supported internal controls, we develop the selected university's P2P process prototype.

Generally, the application of emerging data and technologies in ICs is a co-construction process. It requires co-construction with its “relevant social groups.” It is an on-going procedure to institutionalize data-driven systems and optimize organizations’ decision-making routines. The fundamental condition is to find suitable algorithms and qualified data. With qualified data, the analytics can output an accountable and reliable result to signify the institutional pressures, which acts as the starting point. The case study also finds evidence that integrating “relevant social groups” and technology helps circulate emerging technology. The most challenging task of emerging technology adoption is to deal with a variety of data issues. The case study demonstrates techniques and mechanisms to prepare the integrated data platform and the shareable database.

This essay develops a theoretical adoption strategy to promote emerging data and technologies in ICs. It contributes to the literature by preventing the emerging technology usage trap and improves its analytics performance. This work answers the call for a study from the PCAOB to investigate emerging technologies application in auditing. The four social construct procedures provide practical value to institutionalize emerging technologies. The six propositions also provide technical guidance on developing a more flexible data-driven analytical schema.

Chapter 3. Embedding of Audit Data Analytics in Continuous Monitoring Systems to Develop Internal Control Intelligence:

--A Prototype for a University's Procure-to-Payment Process

3.1. Introduction

This chapter develops a prototype to validate the theory that we proposed in Chapter 2. The proposed approach in the development of Continuous Audit Data Analytics (CADA) is to break down the complexity of internal controls into many simple sub-processes until they can be monitored by algorithmic analytics. Simultaneously, the separated analytics is expected to aid each other and be integrated to gain enhanced control performance. The beauty of this approach is that it allows each analytics artifact to operate in its most suitable area to improve performance. From the perspective of design, CADA is far beyond a simple data analytics artifact. It includes a set of analytical objects that have interactions with the audited institution and the related stakeholders. We can describe the entire systematic procedure as the data value chain (GSMA, 2018). It acts as a conceptual framework to specify where data is identified, acquired, processed, stored, analyzed, and finally utilized by decision-makers to add value. This chapter demonstrates how CADA needs to manage the entire chain from data generation through data-driven audit evidence, and the specific techniques as well.

In "*The Next Frontier in Data Analytics*," Tschakert et al. (2016) described four types of data analytics in auditing: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. The proposed analytical paradigm is to embed the four types of analytics in the Continuous Audit and Monitoring System with a sharable database as a bridge. The database can be shareable by all internal control (IC) stakeholders to restructure the entire process of ICs by its data elements. This analytical schema is exemplified by using a Procure-to-Payment (P2P) project conducted jointly with a public university's internal audit and procurement departments. We illustrate the thorough process of the proposed schema, including data preparation, substantial test, exploratory analytics, and confirmatory analytics. Each analytics is managed as a module, and we embed the four analytics to act as an integrated CADA project.

This essay seeks to contribute to the Continuous Audit (CA) literature in three new ways. First, CA emphasizes filtering transactions to find "exceptions." This study provides solutions to discover new potential risk areas to identify inherent risks, extending the analytical scope. Second, we develop an approach to digitize the analytics of those internal controls by formalizing control rules within each transaction. A sharable data platform provides opportunities to integrate each module of the CADA project. This arrangement can let each analytics work in its most suitable area. It also integrates and improves the overall performance of the CADA artifact. Last but not least, the combining analytics schema extends CA's tolerance to take advantage of ubiquitous intelligence, which is mined from the area around its domain.

Technically, we borrow the risk discovery logics from MineSweeper, a famous computer game from Windows. Players need to explore and signify which squares have bombs using the knowledge about the squares around the targeted square. If management has enough knowledge in a specific area and develops clear control rules, we apply an algorithm to test the rule compliance and assign a violation score for each transaction in the specified area. The violation score can act as a benchmark to explore potential operational risks in adjacent areas, similar to the MineSweeper game, where the number is shown how many bombs are adjacent to it.

Specifically, following the six application propositions of social-constructed CADA in the previous chapter, we describe three techniques around internal control intelligence. The first is the digital transformation techniques to construct a shareable data platform. The second is how the combining CADA schema can ensure analytics traceability and interpretability. We also demonstrate how diagnostic analytics and prescriptive analytics can identify potential risk areas requiring additional controls. The CADA project provides an enhanced analytical performance to assess internal controls and develop internal control intelligence as well.

The nature of internal control is a (complex) process (COSO, 2013). To achieve an enterprise's mission, the board of directors needs to define strategies and then derive concrete operations from the plan. Finally, management has to report on the activities and check whether their actions comply with the laws and regulations. Internal control assessment needs to integrate all related operations and activities to observe its effectiveness and efficiency. Thus, technically, to achieve actionable knowledge discovery, a data-driven solution needs to manage all corresponding domain constraints (Cao et al., 2010). Under this constrained condition, the combination of domain knowledge and data analytics techniques can provide an enhanced solution that allows each analytics to be used in its suitable area. This essay proposes a modular-based schema to break down the complexity of internal controls to handle specific IC's domain constraints and improve its efficacy. Thus, from the solution perspective, CADA can be defined as an actionable knowledge discovery (Cao et al., 2008) with three features: (1) The goal is to find audit evidence or actionable business policy to improve business performance, (2) It requires active data mining mindsets to explore ubiquitous data, which is mined from the areas other than accounting, and (3) It has “social” and “technical” sides and needs auditors’ involvement and confirmation.

Data infrastructure is necessary for CADA, and a shareable data source can offer traceable and reliable data flow for the combining analytical schema. This chapter exemplifies developing an agile domain-based database as the shareable data platform to access data from all relevant sources related to the internal control system. A practical challenge for this digitization is that the scope of the database should be big enough to cover all of the activities in reality. CobiT 5 acts as a theoretical basis on this topic. The ITGI (IT Governance Institute) and the ISACA (Information Systems Audit and Control Association) developed CobiT (Control Objectives for Information and Related Technologies) framework to facilitate the implementation of SOX with IT technology. From the perspective of technology, CobiT is a more concrete form of the COSO framework, and it offers detailed control objectives for the internal control system and governance in the IT area. The CobiT framework groups control objectives into eight categories, which consist of 13 processes. It offers a supportive list of specific control objectives for the rules digitization to ensure its completeness.

Theoretically, a fundamental requirement of the internal control system design is to set up necessary control nodes to manage risk. In each business process, management needs to know which actions or policies can mitigate risk and improve business performance. CADA is required to explore and decide the actions (data attributes) that can act as conditions in the rule expression, and certify that the actions have impacts on business performance.

Many of the requirements derived from the SOX can be expressed through the rules of these processes. Control rules provide a common language to improve communication between the managers and auditors. A digital transformation of ICs by formalizing business rules provides significant benefits for all relevant stakeholders (Ross, 2003; Taylor, 2012). The regulations benefit from the exhaustive investigation of law compliance by testing all of the rules' compliance. The external auditors can accumulate fact-based evidence for their judgment from the investigation of the enforcement of the rules. Thus, the management and internal audit obtain a platform to optimize their decision-making. Furthermore, this rule repository also improves audit evidence's adaptability, traceability, auditability, reusability, and manageability (Boyer and Mili, 2011).

This essay is arranged as follows: the following section is the literature review, and we illustrate the technical mechanism of combining CADA schema after the review. The fourth section is a prototype that exemplifies how to apply the schema in a P2P process. We conclude the paper in the fifth section. In Appendix 1, we describe a conceptual mechanism to apply the schema to implement SOX 404.

3.2. Related Literature Review

3.2.1 Internal Control Assessment and Continuous Audit and Monitoring System

Internal control is a hot topic in the audit and management area. A considerable amount of research has already been done from auditor behavior and audit opinion (Ashton et al., 1980; Bloomfield, 1997; Krishnan, 2005). One outcome of internal control assessment was a specific research branch to focus on regulatory and corporate governance. Moreover, most studies are based on the risk of misleading financial reports and concentrate on errors or manipulation of data. Another branch of research is a transaction-based audit to test rule compliance over each transaction (Vasarhelyi and Halper,1991; Kim and Vasarhelyi, 2012; Kim and Kogan, 2014). Papers in this branch documented the evaluation of ICs from a company's operational features, such as internal fraudulent activities. The transaction-based approach provides more circumstantial evidence and helps collect more detailed audit evidence than traditional audits. However, this approach practically needs detailed transactional data, which is limited for academics. We follow the transaction-based research branch in this study. The strength of this research grows from the P2P data that the researchers accumulated from the case subject. The following literature review focuses more on the perspective of the transaction-based approach than other aspects.

Three decades ago, Grimlund (1982) proposed an audit function on assessing the risks of financial reports. Srivastava (1985, 1986) then developed the application of Grimlund's function by considering the evidence interaction and independence of internal controls. Following this line, Mock et al. (2009) developed an evidential reasoning approach to Sarbanes-Oxley-mandated internal control risk assessment. They proposed an integrated, five-level taxonomy to assess the quality of the internal control system. Each statement, which claims the related IC quality within the taxonomy, may be threatened by one or more risks. For each management assertion, there are several potential risks. For each risk, there may exist more than one internal control to mitigate the risk. Based on this model, every risk can be mitigated by one or more controls. The existence and effectiveness of each control are expressed as the fifth-level sub-assertion. However, they demonstrated only the taxonomy of the financial report's risky accounts, but they did not link the taxonomy to the related specific business processes.

Vasarhelyi and Halper (1991) initiated Continuous Auditing, and this methodology has evolved from academic theory to a process applied in practice (Vasarhelyi et al., 2012). Some significant novel schemas have been studied for the development of ICs. Kim and Vasarhelyi (2012) proposed a fraud detection model based on potential anomaly indicators in the wire transfer payment process of a major insurance company in the US. They developed a severity index to detect exceptions for the following investigations. Kim and Kogan (2014) proposed a multi-step fraud detection model to screen transactions at various levels to decrease the amount of "exceptions." This schema improves the analytics' practicality: audits are possible to apply a broad screening criterion at the highest level while utilizing more detailed rules at the lowest level.

The "process" nature of IC carries a significant dynamic feature of the research about the evaluation of the ICs. This notion inspired several studies to observe the ongoing dynamic behavior involving IT. Jans et al. (2014) demonstrated the use of event logs to provide audit evidence about the effectiveness of ICs. They extracted meta-data from the event logs of the ERP system from a leading global bank and analyzed the procurement data with process mining of event logs. The study identified numerous transactions as being audit-relevant, including payments made without approval, violations of segregation of duty, and company-specific procedures. This "process mining" recognized the anomalies that the internal auditors cannot use conventional auditing procedures on the same data.

3.2.2. Business Rule Formalization Review

A business rule is a statement to be used to guide managerial behavior or lead the information flow in an organization (Steinke and Nikolette, 2003). From an IT point of view, “a business rule is an atomic piece of reusable business logic, and business rules are the encoded knowledge of business practices” (Ross, 2003, 165). The rules are used to guide and constrain the behavior of management. Reaction rules are the standard IF-condition-THEN-consequence type, and they are often used to act as a trigger that activates the evaluation of rule compliance. When a condition is evaluated and meets a specified criterion, a subsequent activity can be carried out (Wagner, 2005).

The languages of business rule modeling are typically based on formal logic, like Boolean function. They must have stable and precise expressive power (McBrien and Seltveit, 1995). In his influential book, *Principles of the Business Rule Approach*, Ross claimed that business rules could create a seamless, never-ending, self-training environment by a real-time delivery of business logic to knowledge. Moreover, this documentation can prevent the loss of knowledge when people leave the company (Ross, 2003). The integration of rule and process modeling has gotten attention in recent years. Krogstie et al. (1991) first introduced the rule-modeling concept. They suggest that the combination of the business process and rule modeling approaches can improve capturing temporal information. Their proposal demonstrated a top-down approach for the specification of process logic at the lowest level of decomposition. McBrien and Seltveit (1995) further enhanced this integration by defining the structure of rules within the process model. Then reaction business rules can be used to set up the coordination in workflow systems (Kappel et al., 1998). In the rule formalization process, meta-data acts as an essential function. It can combine and integrate important business components (e.g., objective, process, events) and technical components (e.g., data objects, workflow, data attributes) (Kovacic, 2004).

The business rule can determine what happens in a particular situation (Morgan, 2002). Along with this pipeline, Grigoria et al. (2004) proposed business process intelligence. They combined the process analysis and process behavior tables to provide a data set that includes process instance attributes and labeling information. The table then contains all of the information required by the taxonomy tool to generate classification rules. By this arrangement, the designed table makes technical progress, and the components of the table can map the behavior analysis problem into a classification problem. In another prominent work in this stream, Lee (2008) introduced a rule-based and case-based reasoning approach for a bank's internal audit. First, Lee uses rule-based reasoning to determine if a new problem should be further inspected in the screening stage. Then, in the auditing stage, Lee uses case-based reasoning, which performs similarity-based matching, to find the most similar case and use case-based reasoning to solve new problems. Lee's data-driven method helps detect risks in ICs from the pattern of the extracted data, which can help to discover the potential risks hidden in the daily operations.

The literal specification of rules is just data. In general, we are familiar with how to manage data, so it is easy to utilize data elements to express business rules. The domain-oriented database can digitize rules and document the accumulated expert knowledge in the internal control area. It also offers an experimental platform to test-and-learn potential important business rules (Taylor, 2012). The ultimate goal of the business rules approach is to guide and improve decision-making performance. This domain-based database can also help to understand the components of decisions and how the decisions worked.

3.2.3. D^3M Methodology, Risk Discovery, and Feature Extraction

Domain-driven data mining (D^3M) is an emerging research line that explores actionable business insights by considering the function of domain knowledge in problem solving (Cao et al., 2009; 2010). The traditional knowledge discovery data mining is an algorithm-focused method that uses refined data to find any hidden data pattern in an automatic trial-and-error process. The evaluation focuses mainly on whether the technical metrics meet the statistical criteria. Unlike the traditional data mining approach, the ultimate goal of D^3M is to discover actionable knowledge to satisfy real user needs. D^3M emphasizes to use real-life data and domain-oriented factors to tell the story. It encourages humans to be involved in the analytics on discovering actionable insights. D^3M encourages taking advantage of ubiquitous data (Cao et al., 2010) surrounding a domain problem to relieve the domain constraints.

Cao et al. developed a series of personalized data mining approaches to extract a variety of intelligence to achieve specific goals. For example, when the task needs one analytics to use other analytics' result, combining data mining can integrate multiple relevant mining techniques to enhance analytics performance (Cao et al., 2009). Suppose the goal is to observe the detailed behavior of a small group. In that case, the in-depth data intelligence is to have depth exploration to narrow down the commonly mentioned data patterns in transactional or demographic data (Cao et al., 2008). Besides, organizational and social intelligence is a mining technique that is used to discover hidden knowledge by involving social factors in the business processes, business rules, and things like social networks (Cao et al., 2011). Moreover, we can use human intelligence to accurately analyze human-related factors, such as social roles, demographic factors, expert knowledge, and user preferences. Human intelligence can add specific knowledge about decision-makers' behavior in data mining to improve analytics performance in other related intelligence (Cao et al., 2013). Network and web intelligence is to merge patterns from multiple data sources to cater to factors related to networks for data allocation (Cao et al., 2014).

D^3M pays significant attention to domain knowledge and qualitative intelligence from domain experts to emphasize the usefulness of analytical results. This requirement can be achieved "through selecting and adding business features, involving domain knowledge in modeling pattern significance and impact, supporting interaction with users, tuning the parameters and data set by domain experts, optimizing models and parameters, adding organizational factors into technical interestingness measures or building domain-specific business measures, improving the results evaluation mechanism through embedding domain knowledge and expert guidance (Cao et al., 2010, P.109)."

After the internal controls digital transformation, huge data can be accessed and used for the development of internal control intelligence. We need to reduce the data dimensionality to gain attention on the riskiest (important) control activities. Statistics literature has many feature selection approaches. Principal Component Analysis (PCA), an eigenvector-based multivariate analysis, demonstrates its usefulness for continuous-scale data (Jolliffe, 1986). Both Logistic PCA (Collins, 2001) and Robust PCA (Candès, 2006) can be used for categorical data. Greenacre (2006) proposed the Multiple Correspondence Analysis (MCA) to extract useful categorical data features by maximizing the variance across cases.

Internal control assessment is similar to conducting a survey, and the management and internal auditors answer the questionnaire and provide data to express what they did to serve internal controls. Then we need to select useful features and reduce data dimensionality to afford data analytics. The critical information of internal controls assessment is about Who, and When does management take What actions to assure the quality of accounting information. Categorical attributes are the most frequent data to reflect control activities. Thus, MCA is a more suitable approach in internal control evaluation than PCA and the like. Technically, MCA combines the PCA (latent grouping factors for variables with the horizontal axis) and cluster analysis (grouping observations with the vertical axis). MCA offers a more general perspective to observe the data pattern than just using PCA. Researchers often emphasize the first two dimensions because they usually can provide information about which data attributes are the most important in the targeted dataset. The first dimension of MCA maximizes the variance of the studied dataset. It is also coined as the first principal inertia. The second dimension maximizes the variance subject to the scores being uncorrelated with those on the first dimension. The detailed statistical property and mathematical basics of MCA are illustrated in Appendix 5.

3.3. Embedding of Continuous Audit Data Analytics

As introduced above, the proposed approach is to perform the proposition in Chapter 2 using analytics practice. One primary goal is to break down the complexity of the internal control system and transform all control rules into data attributes, which are used to develop an agile database. At the same time, the design ensures to keep the data flow to be traceable and interpretable. The design follows a loop control mechanism to apply its ongoing refinement. The proposed CADA schema offers opportunities to combine different types of analytics to enhance auditing performance. The philosophical notion is to break down the evaluation tasks into pieces and subsequently apply the most appropriate analytics in each assessment. The proposed modular arrangement provides flexible analytics to manage domain constraints and extract relevant intelligence from other sources (Cao et al., 2008).

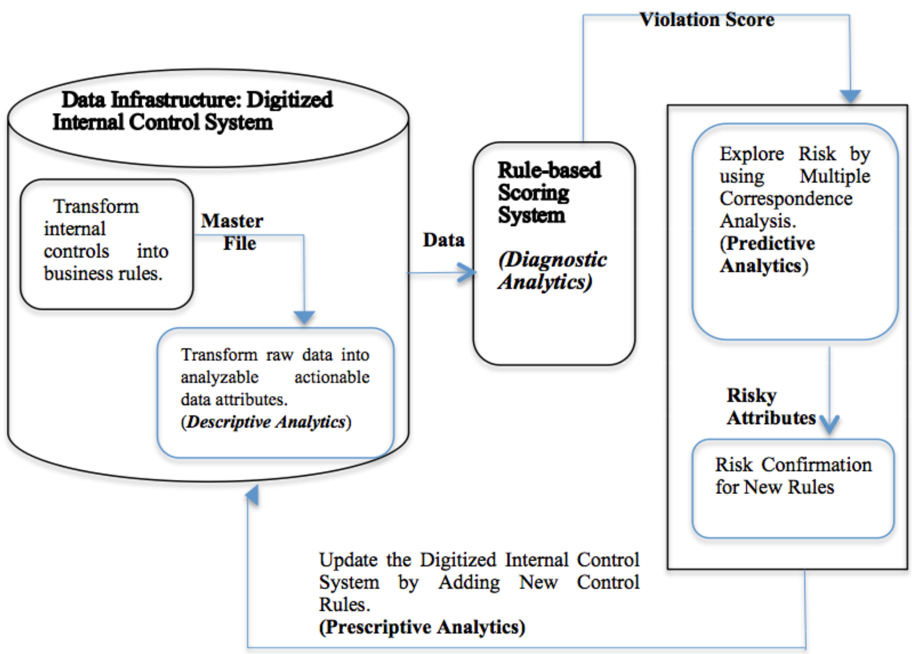

Figure 3.1: The architecture of the modular-based analytics schema

As shown in Figure 3.1 above, the core component is an agile database, which digitally transforms the internal controls by utilizing three types of elements (Objects, Users, and Control actions). In this transformation, the business rule is a suitable linkage because it has a straightforward expression about what must be done, or what cannot be done in cases of uncertainty. Rule-based enterprise knowledge can be codified as a series of data attributes (Ross, 2012). With this data platform, auditors can exert analytics to collect evidence for internal control evaluation. The detailed logic for the data preparation and the four types of analytics is demonstrated as follows.

3.3.1 **Module 1:** Develop an Agile Data Platform to Bridge Audit Data Analytics and Continuous Monitoring System

Data is the basis of analytics, and any powerful emerging technologies will be paralyzed without reliable data inputs. The module's objective to have a data platform that can digitize the control rules with a suitable data structure for the specific IC goal. It also can offer data infrastructure to combine analytics and enhance overall performance. Each Information System may have its specific function and technical arrangement to achieve its targeted goal. However, we often need to observe the performance of ICs by crossing several departments to collect the necessary data. This investigation often needs to impact the current decision-making routine and the arrangement of some Information Systems. A shareable data platform can solve this issue. Different stakeholders can observe the entire internal control process with the shareable data platform. Because the database can restructure the entire IC process by its data elements, it also offers opportunities to run analytical experiments to observe potential solutions for the current ICs. The data platform's importance deserves some detailed discussion about the data integration because it acts as an indispensable function in CADA.

As discussed in the previous section, CobiT 5 addresses a potential issue and ensures that the scope of the domain-based database can be big enough to cover all of the activities in the internal control system. With the guidance of CobiT 5, we break down the internal control system into processes and sub-processes until the processes can be represented as a series of business rules. Using three types of data attributes can digitize these rules: decision-makers, decision-making time and control actions, and decision results. The decision-makers are used to answer the “who” question. A timestamp for each control action is essential for the efficiency evaluation, and the time can answer the “when” question. The “what” question depicts the specific actions for a company to exert ICs. Control rules can capture all business actions with answers to the “who,” the “when,” and the “what” questions. This formalization can deliver business logic to a database, which can make data analytics achievable (Ross, 2003). With this transformation, the knowledge is no longer tacit. Even if the company loses the person, it cannot lose the knowledge. This transformation also makes rules automation achievable.

As mentioned above, an implicit issue is that this transformation lacks a theoretical basis to handle the generality of most situations. CobiT 5 has 164 illustrative control objectives, and it acts as a guide to digitize the internal control system. This objective list has been extensively used in the auditing professions and industries for IT-supported internal controls. In Appendix 2, the study exemplifies the transformation of the 164 control objectives into necessary specific data attributes.

The ultimate goal of ICs is to implement the Generally Accepted Accounting Principle logic for the calculation, evaluation, and explanation of certain financial statement items within external reporting. Thus the transformation of the ICs needs to satisfy the following six criteria: completeness, existence, accuracy, valuation, ownership, and presentation (Chuprunov, 2013). Figure 3.2 below shows the module that transforms and then documents all of the control actions as data elements. It is critical to ensure that the conversion does not miss important information. Any control activity information missing could impact the subsequent analytics and even the whole project’s power and reliability. The completeness criterion in this transformation can ensure that all of the data attributes can express risk-related users and actions in a specific process. The goal of the mechanism of Figure 3.2 is to break down the complexity of internal controls into many simple activities until we can use data attributes to express the specific control actions. With this digital transformation, the internal controls can be monitored by algorithmic analytics.

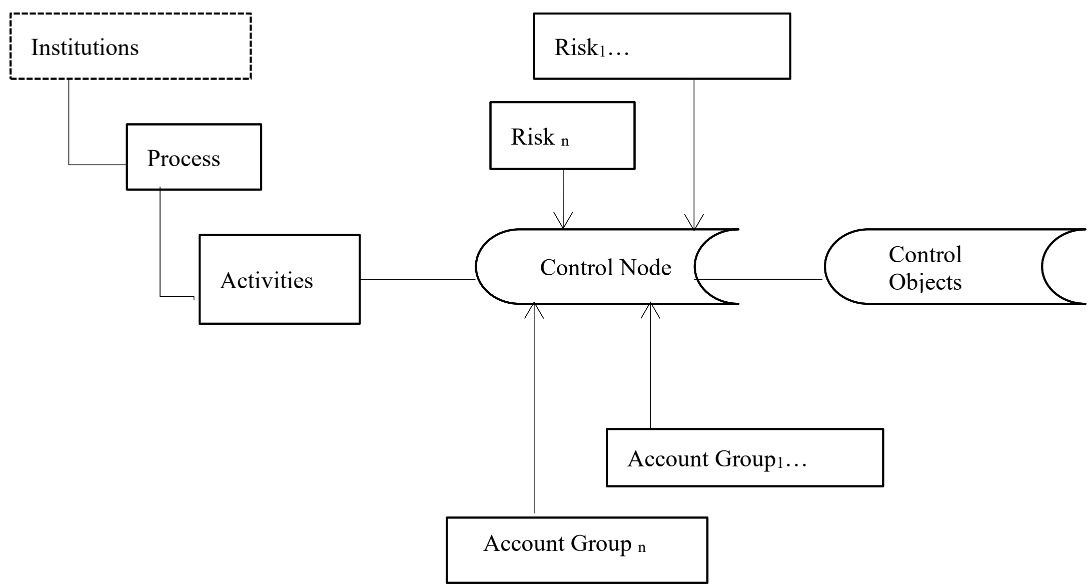

Figure 3.2: The Schema of the ICS Breakdown

The OUA is an object-oriented semantic data model, which holds its advantages over understandable logic traceability and execution transparency. As Ross (2003, 217) claims, "[A]ll decision services should, therefore, have the ability to log how decisions were made—which rules [were triggered], what data was used, what results were generated by analytic or optimization models. These logs can be written to some long-term storage by the decision service directly, or the organization can have a standard design pattern for decision services that ensures this information is." This model logically ensures the completeness of the business rule formalization. As Ross (2003, 256) explains, "[E]ach object and each attribute for each objective have a technical definition that allows it to be accessed. The mapping gives each object and each attribute a description useful to someone familiar with the business but not the underlying technology. This mapped vocabulary will replace technical IT terminology with something that matches the way the business team thinks about the things being [controlled] with business rules."

The "process" and "questionnaire" notion makes it clear that the OUA (Object, User, Action) data model can capture all relevant control behaviors and digitizes the internal control system. This digital transformation meets the requirement of D^3M to develop data infrastructure capability (Cao et al., 2010). The general description of the model includes: (1) the *Object* depicts the control agents of each process, (2) the *Users* are the stakeholders who have related responsibility in these processes, and (3) the *Action* is actionable decision-making attributes that can capture the management's control activities. By using this transformation, we can obtain all the necessary information within the financial reporting network.

One output can be derived from this module is the master file of the digitized internal control system. This master data demonstrate explicitly the management's decisional actions to safeguard the resources. Moreover, it offers a description of the data attributes for the stakeholders, objectives, and related control activities. As Figure 3.3 shows, the master file includes metadata, timestamp data, data about control actions, and data about resources change. With the objects' arrangement, sub-objects, control functions, and attributes, both auditors and the management know what data attributes the evaluation needs, and how the control function works with the extracted data.

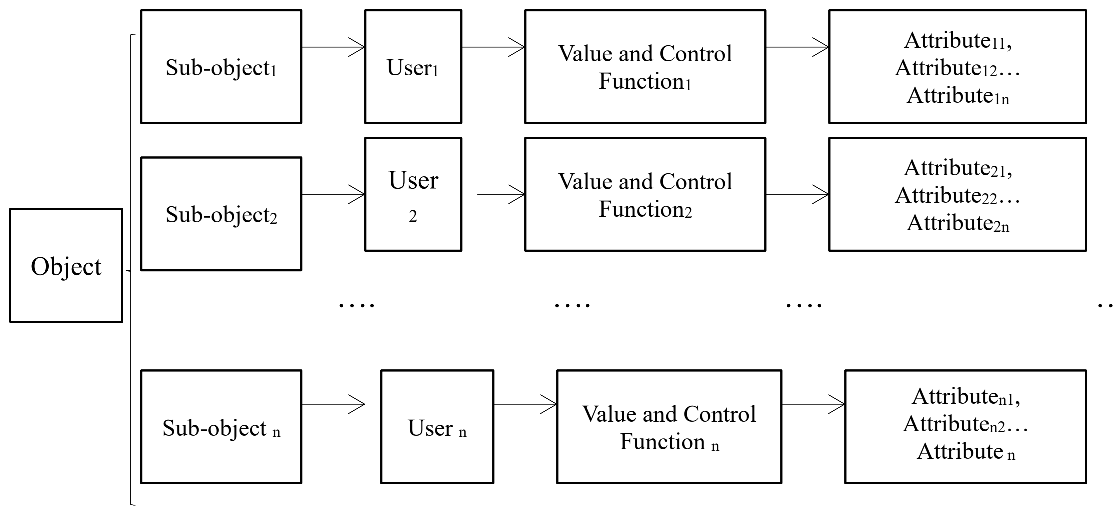

Figure 3.3: The Description of The Master Data of the OUA Model

3.3.2. *The Sub-module to Transform the Raw Data into Analyzable Actionable Data*

The raw data, sometimes, is helpless to develop control rules. For example, the approver is a categorical data with a name list. This list cannot completely express the control actions and cannot be used to develop understandable control rules. We need to apply feature engineering to derive a composite attribute or index for these control rules. After this transformation, these attributes can be used as a direct ingredient of control rules for real-time monitoring. For example, we coded a new attribute to express whether the approver's name is the same as the requestor. Then this attribute can be directly utilized as the condition part of an "IF-THEN" control rule. Thus, this transformation aims to prepare the condition part of an "IF-THEN" rule. This transformation signifies all actions and risk-related policies in a row of the domain-oriented database. This data structure provides robust data infrastructure to experiment with any available analytical models in the following test and exploration.

Two often-used attributes coding methods in CADA include social network analysis and discretization. The first is to apply a social network analysis. For example, in fraudulent analytics, the biggest challenge is the collusion of the stakeholders in the internal control system. Thus, it is highly encouraged to deduce some measurements to describe the network of relevant stakeholders. This analytics is called social-network intelligence in D^3M. The raw data related to actions can be developed into composite attributes to capture the decisional alternatives. The second is to utilize discretization and similar tactics, which stratify a continuous dataset into several specific subgroups. For example, continuous data is hard to set up straightforward control rules, but it is analyzable after the amount is discretized as "1" or "0" based on the criterion of $10,000. We can explore whether the transactions with an amount of over $10,000 hold more risks than less. The final digitized internal control system is composed of these derived data attributes. Much off-the-shelf software, such as ACL and IDEA, can offer fast solutions for this engineering feature.

3.3.3. **Module 2**: Application of Diagnostic Analytics with a Rule-based Scoring System

This module uses diagnostic analytics, establishing the application benchmark by applying substantial testing. Application benchmarking was addressed by the PCAOB in its November 2004 guidance, stating that benchmarking is an acceptable practice as long as certain conditions are met. This module is diagnostic analytics: it involves documenting and testing the relevant controls that support the accounting information to confirm their configuration and operating effectiveness. According to SOX 404, the company management states what and how they applied internal controls to assure the quality of the financial statement. Each transaction's compliance score is qualified as a benchmark to show the risk level for each transaction.

This module runs data analytics to accumulate evidence about the situation of the company's rules' compliance. We apply the rule-based-scoring mechanism (A Macro can automate the test mechanism in some off the shelf products like IDEA and ACL) to practice an exhaustive test of the business rule compliance for each transaction. The rule-based-scoring-system is an investigation on the transaction population about their rules compliance. The logic basis is the Boolean function based on the "IF-THEN" query. The mechanism gives each transaction an effectiveness score; the score is decided by the violated business rule and its related risky weight. Risky weight needs management or auditor judgment. Auditing Standard No. 2201 offers materiality-based scoping and a risk-based scoping to decide risk weight. We utilize materiality-based scoping in the following pilot study and prototype development. The system tests the rules' compliance and operates the scoring mechanism for each control activity in the internal control system. The output from this module is a violation score for each transaction. The score accumulates the number and the severity of rule violations based on the current internal control rules. This violation score acts as a benchmark for the following exploratory analytics.

3.3.4. **Module 3:** Application of Prescriptive Analytics and Dimensionality Reduction to Explore Potential Risky Control Areas

3.3.4.1. Phase 1: Exploratory analysis with MCA

According to SOX 404, the company's IC statement is the starting point of evidence to support the IC's evaluation. Management may misreport it, despite the existence of IC limitation. However, the analytics might still detect these internal control weaknesses even if management, later on, decides not to report them. Furthermore, the emerging marketing environment may have new potential risks, which can produce hazards for companies' development, but managers cannot find these risks in a timely manner. Thus, these potentials may produce severe control risks for the institution's development. For example, 30 years ago, the emerging digital technology changed the customers' preference for photography in a short time. Kodak is the then biggest film company, but then management could not have clear risk consciousness about this change and lost the precious opportunity to win the strategic opportunity. We combine exploratory analytics and a risk comparison design to discover these potential risks to solve this issue. Exploratory analytics matches to the "predictive analytics," which is one of the four types of data analytics in Tschakert and others (2016). One primary goal of this module is to develop necessary control rules to set up monitors for the important potential risks.

According to the Continuity Equation (Kogan et al., 2014), the transaction flows would be in a steady-state without control rules violation. However, if the transaction flow has one internal control defect, some other control nodes would have more opportunity to violate rules. The reason is that one violation needs other violations to seek a dynamic balance. For example, a manager wants to trade with his family business. He would try to avoid other managers' interference and violate the rule of segregation of duties. He may also seek to complete the transaction on the weekend to prevent more people from noticing these transactions. Thus, these attributes have some relationship, and the attributes that are not specifically included in the current rules are related to the attributes that are already in the current internal control rules. Based on this understanding, we develop the following assumption: the fraudulent behaviors are bound to leave marks in the transactional information trace.

A practical challenge is that we need to define the “result” part for the “IF-THEN” rules because the definition of “result” may be different in different organizations. We need to apply CADA to understand the “result” part and investigate the relationship between the “condition” and “result.” Technically, the violation score that comes from the rule-based-scoring-system is like the number of MineSweeper game, which indicates how many bombs are adjacent to it. It is advantageous to utilize the knowledge that we already knew from the rule-based-scoring-system to explore potential operational risks. We assumed the violation score from the rule-based-scoring-system as the virtual “result” part.

There are a huge number of data attributes that can be developed as control rules. The exploratory analysis aims to reduce the attributes dimensionality and select the attributes that have potential risks. Theoretically, the analytical model’s choice in the exploratory analysis is dependent on its domain context (Cao et al., 2010). As we discussed in the previous section, MCA can deal with a categorical data type, and the investigation of ICs includes many category variables. Thus, MCA is a suitable exploratory approach to reduce the data dimensionality by discovering complex relationships behind the data pattern. MCA depicts the relationships among observations (rows) and variables (columns) by joining a cluster analysis for the rows and a PCA analysis for the columns (the statistics basics are illustrated in Appendix 5). Geometrically, the cloud of the data points has a barycenter that acts as the center to observe the relationship between the rows and columns. We explore the knowledge of the quality of ICs by studying the similarity among the data points. Observations are compared on the basis of the presence-absence ("1" or "0") of the features that they presented in the operations.

Geometrically, the first two dimensions from the MCA analysis usually hold the essential knowledge for the data set. These two dimensions construct a coordinate system, and each quadrant of the coordinate system includes a group of variables. Thus we use these two dimensions to discover potential risky control nodes. As illustrated in Appendix 5, MCA has two kinds of data attributes. Active attributes are used to construct the reduced dimensions and coordinators. The supplement attributes were not used to construct the reduced dimensions. The supplemental attributes will be added into the analytics when the analyzers want to understand the relationship of the active attribute data pattern and the supplement variables. We took the investigation result (violation score) from the rule-based-scoring system as a supplementary variable that was not used to construct the two dimensions. After the development of the coordinate, the violation score can be transited into the coordinate to see the data pattern between the violation score and other exploratory variables.

The initial variables, which are input into the analysis, are control-action-related attributes to come out of the feature engineering analytics module. We follow Greenacre (2017) to define the following three criteria to select potential risky attributes: (1) contribution of the variable is more significant than 1/k, and k means the number of variables that are joined in the analytics, (2) the square correlation between a variable and the dimensions is bigger than 0.5, and (3) the absolute value of the coordinate value of the variables is more significant than one.

3.3.4.2. Phase 2: Confirmatory analysis

This study embeds a risk comparison in this module to confirm the potential essential controls. We follow Ronald Fisher, the early pioneer of the experimental design (Fisher, 1971), to define the training dataset and the testing dataset. The training dataset is that the transactions violated at least one control rule that lists in the IC statement. The testing dataset is that the transactions include risky features that come from the MCA analysis (The testing dataset may also have transactions that violated other rules that already existed in the current IC statement because of the interaction among these control actions). This classification criterion is whether the testing dataset's risk level is higher than the average risk of the training dataset. The new potential control rules need to be included in the internal control system if the tested sub-group is riskier than the control group. Thus, the objective function in this analytics is to search for relevant potential control rules based on the quantifiable risk level.

According to AS 2201, the risk assessment needs to consider two aspects: the transaction amounts and the risk's statistical opportunity. We define two criteria for accepting the potential control rules: (1) the mean violation score of the testing dataset is significantly different from the training dataset, and (2) the risk index is higher than 1. We choose two proxies, violation probability and transaction amount, to develop a combined risk index. The index acts as the threshold of whether to include new potential control rules. We define the risk index as follows.

Definition 1: Risk Index (RI_i)

Risk Index quantifies the risk based on the transaction amount *(the left half of Equation 1)* and the violation probability *(the right half of Equation 1)*. The violation probability is computed as the Z-score by which the mean violation score of the testing dataset is above the mean score of the training dataset.

$$RI_i = \frac{Mean(TransactionAmount_i)}{Risk\ Threshold} + \frac{Mean(NRVS_i) - Mean\ (VS)}{std\ (VS)} \qquad (Equation\ 1),$$

where (1) Mean (TransactionAmount$_i$) is the average amount of all the transactions that violate the new rule; (2) Risk Threshold is a parameter decided by the auditors' judgment; (3) Mean (NRVS$_i$) is the mean violated score of the testing dataset (the transactions that violate the new rule); (4) Mean(VS) is the mean violated score of the training dataset; and (5) std(VS) is the standard deviation of the violated score of the training dataset.

Let's take a simple example to clarify the Risk Index. In the rule compliance test from the third module, we obtained a distribution of violation scores for all of the transactions that violated rules; the mean is 9, and the standard deviation is 0.78. We found that there were 20,000 transactions that have a feature that is related to the new rule, and the average transaction amount was $15,000. We define these 20,000 transactions as the testing dataset. The mean violation score of the testing dataset is 9.6. If we set up the risk threshold as $30,000, and also, the mean violation score of the testing dataset (9.6) is *significantly* higher than the mean of the training dataset (9), then RI$_i$, in this case, is *RI = 15,000/30,000 + (9.6 - 9)/0.78 = 1.13*. This new rule should be seriously considered in the internal control system when we set the Risk Index threshold as "1".

The confirmatory analytics goal is to assess the potential risk by using data analytics, which is derived from the risk comparison. Because the sharable agile data platform can provide a robust data-driven mechanism, external auditors also can use this proposed combined analytics schema to implement SOX 404. The details can be seen in Appendix 1.

3.4. A Pilot Study of the Prototype of the Combined Data Analytics Schema

A pilot study was done to validate the proposed data analytics schema by using data from the institution described in Chapter 2. The data in this pilot study include the transactions from October 2016 through April 2018 for a public university in the United States. Moreover, the project team is a cross-departmental group, including the internal audit department, purchasing department, human resources department, information system service department, data analysts, and consultants.

3.4.1. Breakdown of the Internal Control System and Identify Control Rules

Following the demonstration of the schema from Figure 3.4, the project team separated the P2P process into sub-processes and identified related risks for the transformation of the rules. The main risk in the process is from unauthorized purchase orders and the associated outflow of funds, including purchase order creation and approval, ordering, vendor selection, invoice verification, payment, and bank-related accounting. To counteract these risks, managers restrictively handle data transfer and goods receipts by strictly using a purchase-order-number. The highly risky three accounting accounts include accounts payable (based on the invoice amount), products received, and payment.

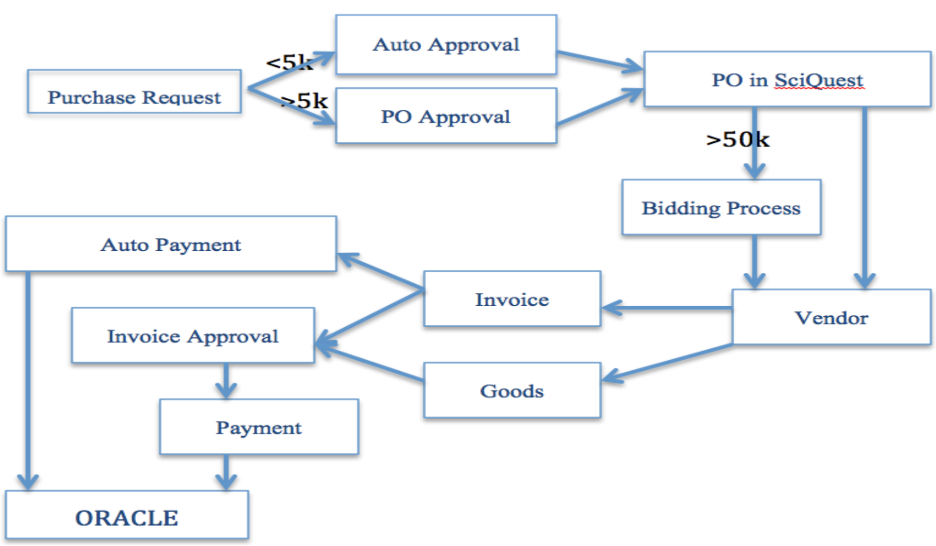

Figure 3.4. The data flow map of the P2P process

Table 3.1 illustrates the formalization of the rules by applying the OUA data model. We used the 21 control objectives of the purchasing cycle of CobiT 5 as a benchmark. Unfortunately, the university did not document the products received, and the purchase order numbering system has technical issues. Thus, the current internal control system misses the following five rules, which need to be included in the system to ensure effective controls. The five missed control rules include: (1) All purchase-order issued are input and processed (No. 74 in Appendix 1). (2) Amounts posted to accounts payable represents goods or service received (No. 75 in Appendix 1). (3) Amounts for goods or services received are recorded in the appropriate period (No. 77 in Appendix 1). (4) Disbursements are made only for goods and services received (No. 83 in Appendix 1). We have 11 final rules to prepare the master file for the specialized database. The missed documentation may be a severe control weakness, which is important evidence of the internal control quality because the system failed to record all data about products received to verify the completeness of the transaction. The procurement department had a separate investigation of this case.

With the guidance of the OUA data model, the master file contains three kinds of necessary attributes, which include the identity of users, actions, action timestamp, and action related measurement, to prepare the domain-based database. These elements theoretically capture all of the control activities that management has done. We illustrate the description of these data attributes in Appendix 3.

Table 3.1: Control rules and the mechanism to calculate violation score

Illustrative Control Objective	Financial Assertions	Risks	Users (stakeholders)	Raw Data Attributes	Control Attributes from Feature-engineering (Code in IDEA)	Segregation of Duties	Violation Score	True Positive Rate
1. The invoice date has to be later than the purchase order approval (creation) date.	Existence	Purchasing activities happen before approval.	Order Requestor; Order Approver	Invoice date; Approval date; purchase department; Vendor number; Product category.	@age(InvoiceDate, PO_CreationDate) < 0	The approver cannot be the same person as the requestor.	V-score = 5	100%
2. The payment date has to be later than the invoice approval date.	Existence	The payment is not approved.	Payment Approver; Cashier	Payment date; Invoice approval date; Payment type; Check number.	@age(PaidDate, Approval_workflow_date) < 0	The approver cannot be the same person as cashier.	V-score = 5	100%
3. If not under a prepaid contract, the payment date has to be later than the invoice date.	Existence	Payment is before the invoice date (excluding prepaid purchases).	Invoice Owner; Cashier; Payment Approver.	Payment date; Invoice date; Payment type.	@age(PaidDate, InvoiceDate) <0	Approver cannot be the same person as cashier.	V-score = 2.5	98.1%
4. It is less than five days from invoice system creation date to payment date under the purchase payment terms.	Valuation	Invoice Payment staff holds invoice too long and delays the payment.	Invoice Payment staff; Cashier; Payment Approver.	Payment Terms; Payment Due Date; Payment Approval Date.	DelayDays= (@age(PaidDate, System_creation_date) - (5 + VendorTermDays)	Not Available.	If Z-score_DelayDays < -0.675, then V_score = 7.5; Else if -0.675 =< Z-score_DelayDays < 0, then V_score = 5; Else if 0< Z-score_DelayDays < 0.675, then V_score = 2.5.	96.3%
5. Payment is less than five days from the vendor invoice	Valuation	The staff holds the invoice too	Invoice Handling Staff; Cashier;	Payment Date; Invoice management staff; Cashier;	DelayDays = (@age(PaidDate, VendorInvoiceDate) - (5 +	Not Available	If Z-score_DelayDays < -0.675, then	95.5%

date to payment date under the purchase payment terms.		long and delays the payment based on the vendor invoice date.	Payment Approver.	Payment Approver	VendorTermDays)		V_score = 5; Else, if -0.065 < =Z-score_DelayDays < -0.19, then V_score = 2.5.	
6. The sum of the same buyer's purchase from the same suppliers on the same date is bigger than $5000 (the university agrees to utilize Quick Order if the total purchase amount is lower than $5000).	Completeness	Requestors split one transaction into several purchases to avoid higher-level managers' approval.	Purchase Requestor; Purchase Approver;	Supplier Number; Invoice Amount; Invoice Date; Invoice number; Product description; purchase order creation date.	Step 1: Extract transactions: $Vendor_i$ = $Vendor_j$ .AND. $Buyer_i$ = $Buyer_j$ .AND. $CreationDate_i$ = $CreationDate_j$ .AND. $PO\_line\_NO_i$ = $PO\_line\_NO_j$ Step 2: Invoice_total < $5000 .AND. Sum (by buyers) > $5000 Note: i,j mean any two different transactions	Approver cannot be the same person as requestor.	V-score = 10	100%
7. Detect and stop duplicate payments.	Completeness	Duplicate payment for the same invoice.	Payment Requestor; Cashier; Payment Approver	Vendor Number; Invoice Total; Invoice Date; Invoice_line_extended price; Invoice Number; Purchase Order Number; Invoice Status.	Same-Same-Same-Different code: Condition1: The transactions have the same Vendor, Invoice Total, Invoice Date and Invoice_line_extended price, but the transactions have a different invoice number and purchase order number. Condition 2: The invoice status is "paid," and the invoice total is bigger than 0.	The approver cannot be the same person as the requestor, and approver cannot be the same person as a cashier.	V-score = invoice total * 0.001	78%
8. Investigate the transactions that two of the requests, approval, and	Entity Level Control	The condition that both payment	Payment Approver; Payment Requestor.	Order request date; Approval date; Accounting date; Payment date.	@weekday (Accounting_date) = 0; @weekday (Approval_date) = 0;	Approver cannot be the same person as requestor, and	If simultaneously satisfy two of the three conditions, then V-score =	Not Available

system input, happened on the weekend.		and related requests and approval happen on the weekend has more chances for fraud.			@weekday (Paid_date) = 0	approvers cannot be the same person as the cashiers.	2.5	
9. It has a conflict of interest that the active vendors have the same address as the employees.	Entity Level Control	Employees buy products from their own companies.	Employees; Active Vendors	The detail address for employees and active vendors	Step 1: Extract vendors that have the same address as employees. Step 2: Detect the transactions from the violated vendors.	Conflict of interest exists in the transactions that happen with the employees' own company.	V-score = 10	98.1%
10. The purchasing staff has to speed up the processing time to earn the available discount.	Completeness	The available discount is wasted because the invoice was held too long.	Invoice management staff; cashier; payment approver.	Payment Terms; Payment Due Date; Payment Approval Date. Paid Date.	If Discount term > 0 .AND. process_length = (@ age(paid_date, system_creation_date)) – vendor terms day > 0	Not Available	V_score = 10	79.2%
11. Only valid changes are made to the supplier master file.	Completeness Existence	Invalidated vendors were accepted as active vendors	Vendor master file change requestor; Vendor master file manager.	Vendor Name; Vendor Number; Vendor Address; Vendor record date;	$VendorName_i$ = $VendorName_j$; $VendorNumber_i$ = $VendorNumber_i$ $VendorAddress_i$ = $VendorAddress_j$; Note: i means the mater file; j means the transaction database	The vendor master file manager cannot be the same person as the requestor	V_score = 5	100%

3.4.2. Data

After the project team prepared the master file that states all necessary data attributes needed in the analytics, we developed a domain-based database by extracting data from various data sources. The data architecture used in the scoring system can be seen in Figure 3.5. The prototype includes data from four sources: Oracle, SciQuest, People Software, and Excel spreadsheets from related departments. The university has three information systems as the data sources for internal audit departments to access necessary data. The Oracle Data Warehouse is the source of financial data, the SciQuest system is used for purchase orders, and People Software is the human resources system. The team collected the whole transactional data population from April 2017 through June 2018 by utilizing the meta-data-driven approach (details can be seen in Appendix 1).

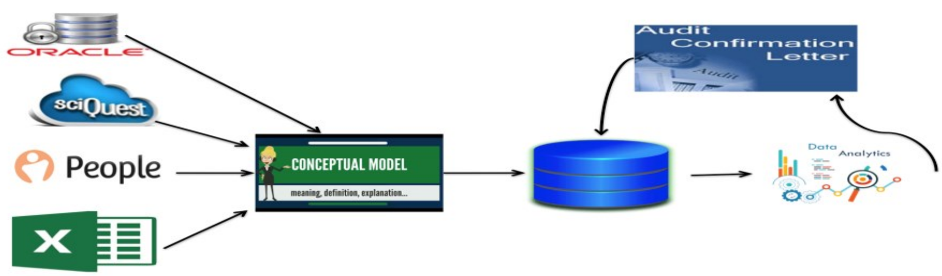

Figure 3.5 The Data Architecture of the University Internal Audit System

Some data were lost in the cleaning process because of the wrong input and other reasons (we undertook other investigations for the missing data). The final dataset has 193,103 transactions, and each transaction had 198 attributes. The data description is given in Appendix 3. The significant timestamp and date sequence of the P2P process were as follows: (1) Invoice_system_created_date, (2) Invoice_Record_Date, (3) Accounting_Date (4) Workflow_step_Date, (5) Invoice_Approval_Date, (6) D_Oracle_Input_Date, (7) D_Oracle_Creation_Date, (8) D_Payment_Date.

3.4.3. A Substantial Test with Combining Rule-based-Scoring-System and Visualization

As discussed in section 3.3, the analytics schema separates the internal control system into two areas to define and assess risks. If management has a clear understanding of risk and has developed control rules, we can exert a substantial test to investigate the compliance of these rules. With a thorough understanding of each control action's potential risks in business transactions, the project team developed a Rule-Based-Scoring-System with Boolean Function and the "IF-THEN" rule presentation. It is technically achievable to exert a substantial test for the compliance of the rules in each transaction. The system can provide a rule violation score for each transaction to demonstrate internal controls' compliance in this transaction. The violation score has its flexibility to be attributed to different stakeholders, like approvals, purchase requestors, vendors, and so on. Visualizing the violation score with its multi-dimensions can improve various stakeholders' understanding of diagnostic analytics' results.

Information overload and interpretative persuasiveness can challenge diagnostic analytics. "Information overload is a state in which a decision-maker faces a set of information comprising the accumulation of individual informational cues of differing size and complexity that inhibit the decision maker's ability to optimally determine the best possible decision" (Roetzel, P. G., 2019). The data-driven Rule-Based-Scoring-System can cause severe information overload because it needs to define the condition part based on delicate considerations of daily business operations. The literature demonstrated much anecdotal evidence for this challenge in a real-time monitoring system (Alles et al., 2006; Issa et al., 2014).

Soucek and Moser (2010) proposed two approaches to manage information overload.

The first is to filter the amount of generating information. The other is to enhance the ability to process information. Guided by these two approaches, we advised filtering incoming information by tuning the condition parts' threshold weight. An adaptable threshold criterion can manage the amount of "exceptions and alerts." Domain knowledge and expertise are necessary to improve the design of the parameters and weights of the Rule-Based-Scoring-System. An interactive education process is an effective way to enhance information process ability. This education needs all stakeholders' participation and knowledge input. Visualization can act as the education function in communication with the end-users of diagnostic analytics. The dashboard of the Rule-Based-Scoring-System not only demonstrates the big picture of the compliance of control rules; it also visualizes the detailed violation actions. Each stakeholder can learn how the control rules work with flexible visualization.

In the case study, the project team completed the rule development in a clear area. It confirmed the applicability of these rules by holding a series of weekly WebEx-based meetings. As Table 3.1 illustrates, we verified the true positive ratio by manually confirming the output exceptions. The true positive ratio is the rate of the violated transactions, confirmed by related managers and rule violators, to cases from the analytical output (True positive ratio equals the number of rule violations confirmed by management divides the number of rule violations from data analytics).

Then the team used the Caseware software IDEA to construct the prototype with the Internal Audit Department's convenience because the department was using the same software. And also, a real-time visualizing dashboard was developed with Tableau for the prototype. Each rule can be an independent module embedded in the information system as a real-time control. After the system identified the rules compliance violation for each transaction, it could calculate the violation scores. The average of the effectiveness score demonstrates the internal control compliance for these groups. It is easy to find specific violations in each control rule. There is a technical issue to assign a weight of the violation score because the weight of importance needs management and experts' judgment. We follow Issa et al. (2014) to use the co-construction project team as the expert panel to prioritize "exceptional exceptions" to decide each part of the control rules' weight. The project team then developed several real-time dashboards for the rule-based-scoring-system from the perspective of departments, vendors, and approvers.

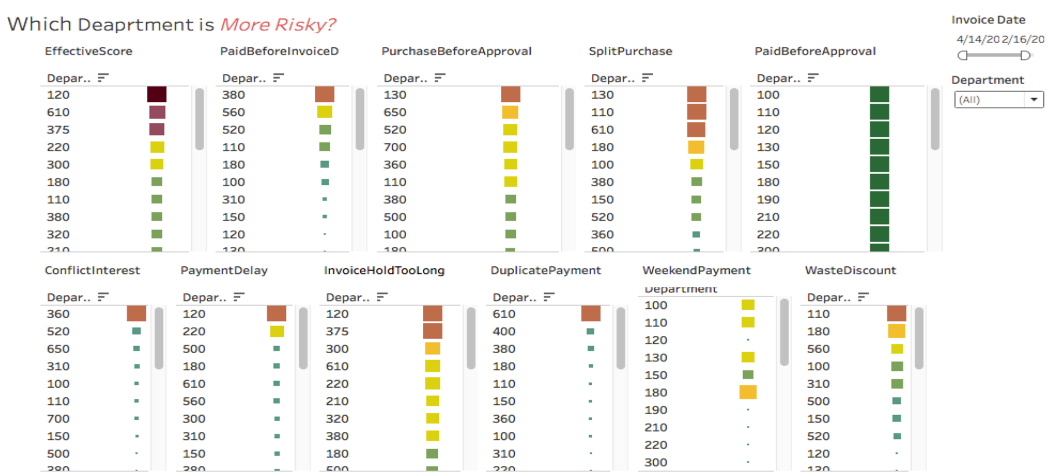

Figure 3.6: The Dashboard of Each Department's Compliance of Business Rules

As Figure 3.6 shows, the dashboard offers a direct understanding of the internal control system. We can clearly see each department's performance for each business rule at any time or within any specific period (by filtering the Invoice Date and Department). The combining analytics not only provides a holistic picture but also gives a detailed explanation of rule compliance.

Figure 3.7 below also demonstrated the effectiveness score for related vendors, buyers, and approvers. It shows a similar real-time dashboard from the perspective of vendors, and a detailed report for the 35 riskiest vendors is included in Appendix 4. Auditors can sort the risk level for each vendor based on the performance of the rules' compliance. For the 35 riskiest vendors, it was not surprising to find that 63% of these risky vendors were employees, which was detected by testing whether a vendor had the same address as that of an employee. It is understandable that conflicts of interest can violate the IC rules. Another significant finding is that all of the payments to these risky vendors were delayed. It is necessary to closely investigate transactions from the vendors that have a conflict of interest.

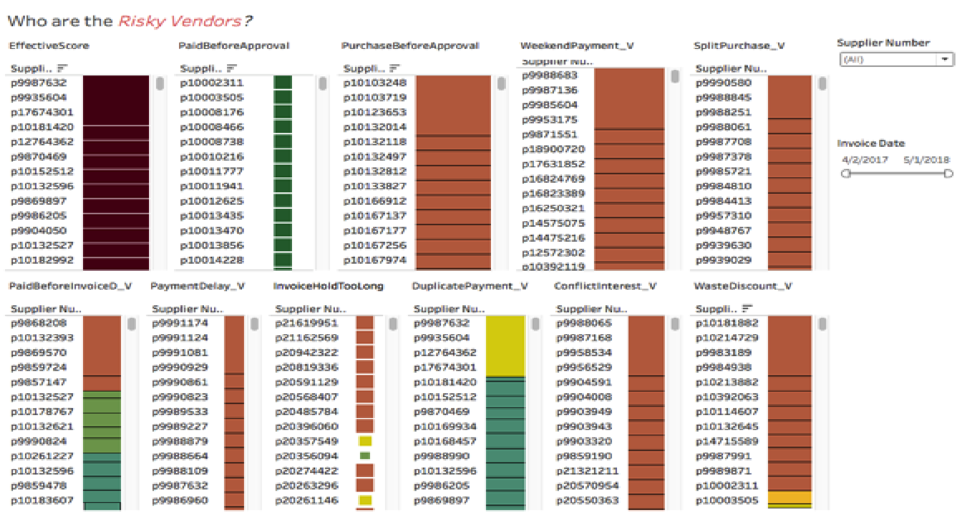

Figure 3.7: The Dashboard to Show Each Vendor's Compliance of Business Rule

3.4.4. Application of Exploratory Analytics and Dimensionality Reduction to Discover Potential Risky Areas

As discussed above, it is necessary to explore potential risks if management has uncertainties about which business rules need to be set up as controls. The objective is to select risky features (variables) that capture risk-related decisional actions by reducing attribute dimensionality. The following analytics uses these attributes to develop new rules to enhance the control performance. Audits need to reduce the data dimensionality and select attributes to exert exploratory analytics. The precondition of the selection is that the attributes can act as the condition part in the "IF-conditions-THEN-results" rule expression. These variables often can express decision-related alternatives or control actions. Table 3.2 shows the analysis result and the definition of each variable.

As suggested in section 3.3, we conducted an organizational network analysis to observe the relationship among vendors, purchase requestors, and approvers. The result produces two new variables: Requestor GEL and Approver GEL (GEL is short for the gestalt element link. The online Merriam-Webster dictionary defines gestalt as a structure, configuration, or pattern of physical, biological, or psychological phenomena so integrated as to constitute a functional unit with properties not derivable by summation of its part). These two variables help to know the number of transactions that each purchase requestor (approver) had with each vendor expressed as a ratio. Any high GEL ratio means that more transactions were done with a particular vendor, which may indicate a special relationship.

Both Effectiveness Score and Violated-Transaction are the attributes that are calculated from the test result from the rule-based scoring system. These two variables can act as supplemental variables of MCA to observe the relationship between the violated scores and potential risky attributes. As we introduced in the MCA methodology part, the supplemental variables are not used to calculate the distance and develop the two dimensions' coordinate system. After active variables developed the coordinate system, we can transit the supplemental variables to see the relationships between the supplemental variables and the data pattern derived from active variables.

Table 3.2: Definitions of Exploratory Analytics

Item	Definition
InvoiceNumber	The number of the invoice, which is the identification of the transaction.
SupplierNumber	The identification of the vendor.
EffectiveScore	The effectiveness score of each transaction, which is the analytics result from the rule-based scoring system. "3" means the effectiveness score is below 80; "2" means the score is between 80 and 90; "1" means the effective score is between 90 and 100; "0" means the transaction has no rule violations within the current internal controls.
AmountRange	If the transaction amount is larger than $5,000, the variable is "1"; otherwise, it is "0".
UnitPrice	If the unit price is larger than $100, the variable is "1"; otherwise, it is "0".
ProductType	If the product is a catalog product, the variable is "1"; otherwise, it is "0".
ConflictOfIntrest	If the vendors who have the same address as the active employees' home address, the variable is "1"; otherwise, it is "0".
WeekendTransaction	If the approval of purchase request or payment request happened on the weekend, the variable is "1"; otherwise, it is "0".
POCreationToInvoice	If the date of the invoice is earlier than the creation date of the purchase order, the variable is "1"; otherwise, it is "0".
InvoiceToPaid	If the payment date of the purchase order is earlier than the date of the invoice, the variable is "1"; otherwise, it is "0".
SystemCreationToPaid	If the payment date of the purchase order is earlier than the date that the invoice was recorded into the information system, the variable is "1"; otherwise, it is "0".
InvoiceToSysCreation	If the invoice date is earlier than the payment date of the purchase order, the variable is "1"; otherwise, it is "0".
SegregationOfDuties	If the approver of the purchase order is the same person as the requestor of the purchase order, the variable is "1"; otherwise, it is "0".
Discount	If the transaction has discount terms, the variable is "1"; otherwise, it is "0".
InvoiceSource	If the invoicing source is manual, the variable is "1"; otherwise, it is "0".
ViolatedTransaction	If the transactions violated at least one of the rules that are illustrated in Table 1.1, otherwise, it is "0".
RequestorGEL	If the vendor in the transaction meets the condition that the requestor's 70% of the purchase orders are from this same vendor, the variable is "1"; otherwise, it is "0".
ApproverGEL	If the vendor in the transaction meets the condition that approver's 70% of the purchase orders are from this same vendor, the variable is "1"; otherwise, it is "0".

The result is illustrated in Table 3.3 and Figure 3.8 (Both the Burt plot and Burt table use the name of mathematician Burt (1953), who developed MCA. The Burt plot is a graphic plot based on Burt matrix, which is the symmetric matrix of all two-way cross-tabulations between the categorical variables and has an analogy to the covariance matrix of continuous variables). We follow Greenacre's (2017) criteria to determine the selected attributes for the following analysis. The squared correlation between the variable and the constructed dimension is more significant than 0.5, and the contribution to the size is more significant than the average of variables that were included in the coordinate development. We included eight variables, so the contribution criterion, in this case, is 0.125. Thus, six attributes meet the two thresholds.

Table 3.3: Result of Exploratory Analysis

Method: Burt/adjusted inertias	Number of observations: 193103		
	Total inertia = 0.99		
	Number of axes = 2		
Dimensions	Principal Inertia	Percent	Cumulative Percent
Dimension1	0.079	80.05	80.05
Dimension2	0.00017	0.17	80.22
Total	0.099	100	

Categories		Dimension1 Coordinate	Dimension1 Square Correlation	Dimension1 Contribution	Dimension2 Coordinate	Dimension2 Square Correlation	Dimension2 Contribution
AmountRange	0	0.14	0.74	0.037	0	0	0
	1	-1.238	0.74	0.328	0	0	0
InvoiceToPaid	0	0	0.124	0	0	0.498	0
	1	-0.82	0.124	0	1.643	0.498	0.502
SegregationOfDuties	0	0.125	0.75	0.03	0	0	0
	1	-1.325	0.75	0.32	-0.002	0	0
Discount	0	-0.001	0.658	0	0	0.366	0
	1	0.113	0.658	0	-0.084	0.366	0.086
InvoiceSource	0	-1.379	0.975	0.157	0.009	0	0.003
	1	0.056	0.975	0.006	0	0	0
ConflictofIntrest	0	-0.0001	0.01	0	-0.001	0.604	0.001

	1	0.062	0.01	0	0.487	0.604	0.491
	0	0.028	1.031	0.002	0	0	0
ApproverGEL	1	-2.02	1.031	0.119	-0.013	0	0.002

As illustrated in Figure 3.8, AmountRange, SegregationOfDuties, InvoiceSource, and ApprovalGEL are along the dimension 1 axis. ConflictOfInterest and InvoiceToPaid are along Dimension 2 axis. ApproverGEL is in the farthest along the first dimension, which means that the feature is risky if more than 70% of transactions of the purchase order approver are with the same vendor. SegregationOfDuties (requestors are the same person as an approver), AmountRange (bigger than $10,000), and InvoiceSource (manual invoices) have similar impacts on the first dimension. The InvoiceToPaid and ConflictOfInterest have a big contribution to the construction of Dimension 2.

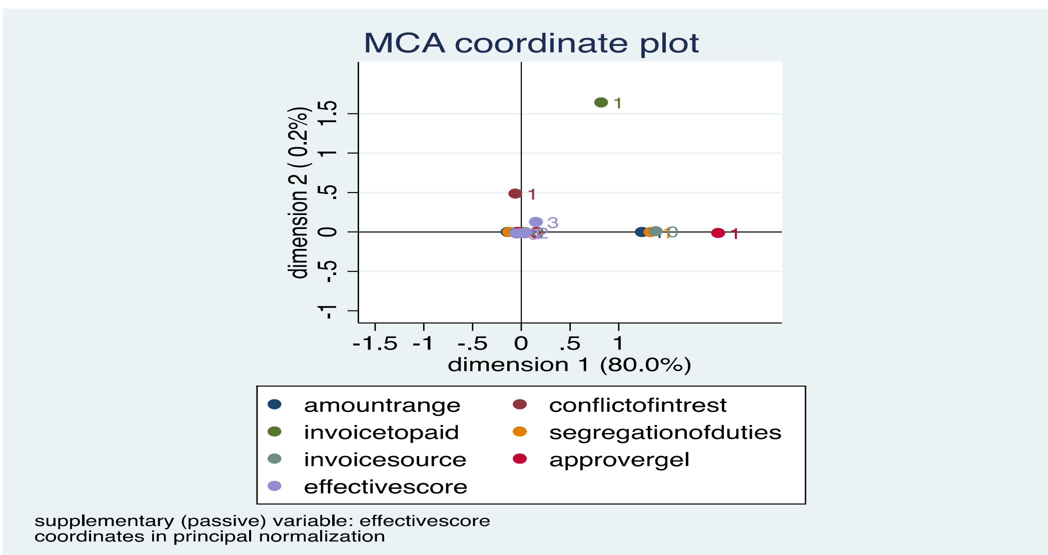

Figure 3.8: Burt plot of the two dimensions of the MCA

3.4.5. New Control Rules Development by Using the Experimental Comparison

The objective of this confirmatory analysis is to compare whether a sub-group of transactions, which have a risk-related feature, is riskier than the benchmark group, which is defined as the group that violated at least one rule within the current internal control system. As illustrated in section 3.3, the criterion is a composite index developed by two proxies: violation probability and transaction amount. Technically, we completed the dimensionality reduction and selected the risky attributes to develop new control rules.

3.4.5.1 Potential control objectives selection and encoding

The new rules are developed from exploratory analytics based on the first two dimensions of the virtual coordination. We developed the following five potential rules and rule-related groups: (1) the group had the same requestors as approvers (Rule1); (2) students, casual employees, or part-time employees own companies that are active vendors (Rule2); (3) the vendors in the transaction meet the condition that approver's 70% of the purchase orders are from these same vendors, and the invoicing source is manual; (4) the vendors in the transaction meet the condition that approver's 70% of the purchase orders are from these same vendors, and the approvers are the same person as requestors; (5) the approvers are the same person as requestors, and the invoicing source is manual.

3.4.5.2 Result of confirmatory analytics

We calculate the Risk Index (RI) according to Equation 1 in section 3.3 by setting up the minimum risk amount. We chose the minimum risk amount to be $15,000, according to a discussion from a WebEx meeting with the project team. Two conditions need to be met to accept a new rule: (1) the mean violation score (calculated by the currently existing rules) of the new rule-related group needs to be significantly higher than the mean violation score of the benchmark group; (2) the risk index is higher than "1."

Table 3.4 shows the analytics results. The confirmatory analytics reject Rule5 and

Rule7 because Rule5 cannot meet the first condition, and the risk index score of Rule7 is smaller than "1." Thus, these two rules are excluded. All other rules meet the two conditions, which means that the average violated score is significantly different from the benchmark, and the risk index score is higher than "1." These included rules were perceived as risky because these activities are often related to some rules violation behavior and need to have close monitoring in the future.

Table 3.4: Results of Welch *t*-test between Control Groups and Tested Groups

Potential rules	Violation Score		Transaction Amount Score	Risk Index Score
	Z-score	*t*-statistics		
$Rule_1$: Requestor cannot be the same person as the approver.	-0.5155	-47.24^{***}	1.9724	1.4569 (accepted)
$Rule_2$: The companies owned by active students (casual employees, or part-time employees) have a conflict of interest as an active vendor.	6.48	18.39^{***}	0 (the mean, $1679, is less than the threshold)	2.24 (accepted)
$Rule_3$: If the vendor the invoice is the manual type, approvers have to be separated from requestors.	0	-0.97 (The difference is not significant.	2.53	Rejected (The rule-related group cannot meet the first condition.)
$Rule_4$: If the approver has 70% or above transaction with the same vendor, the invoice cannot be manual type.	-0.329	8.25^{***}	2.47	2.041 (accepted)
$Rule_5$: If an approver has 70% or above transaction with the same vendor, the requestor must be different from the approver.	-1.03	9.25^{***}	1.15	0.02 (Rejected)

3.4.6. Report on the Application of the Designed Prototype

In conclusion, audit evidence can be accumulated by the implementation of the CADA schema. In this case, the internal controls in the P2P process have at least four severe control issues. First, the responsible department did not document the products received, directly leading to the paralysis of five necessary control rules. Auditors should do specific investigations and solve this problem. Second, the P2P process should implement segregation of duties, meaning that the approvers must be different people from the requestors. The evidence shows these transactions are riskier than the control benchmark group. Third, the management should prohibit transactions with conflict of interest. Last but not least, the manual type invoice shows a significant risk if the approver has a high threshold ratio (70% in this case) or above transactions with the same vendor. In this situation, the invoice type would have fraudulent potentials.

3.5. Discussion and Conclusion

3.5.1. A summary of the Combining Analytics Schema

The CADA schema, which is a combining modular-based analytics artifact, emphasizes the evaluation of ICs by exploring potential risks within the internal control system. The goal is to provide a data-driven solution to enhance internal controls and related accounting information quality. The beauty of the CADA schema is that it breaks down a complicated solution into independent modules, including the agile data platform module, the diagnostic analytics module (Rule-based-scoring-system), and the prescriptive analytics module. Each module can accumulate identifiable evidence and support each other to exert the whole investigation as well. Each module has its own domain constraint and needs different D^3M approaches to handle specific challenges (Cao, et al., 2007, 2010). The data module can feed robust data input for the other two analytics modules. It can ensure that the data flow used in the analytics is traceable and transparent. The rule-based-scoring-system provides accurate evidence for the investigation of the rule violation based on the “IF-THEN” rules testing. The prescriptive analytics module can combine the other two modules' outputs with exploring potential emerging risks by using traceable statistics deduction.

We achieved two goals in the first module. The first is that we can prepare the master file to request necessary data from the operating information system. The validation is to assure the completeness of the master file. The 164 entity-level control objectives from the CobiT 5 provide a strong theoretical basis for the master file. This digital transformation demonstrates the traceability and interpretability within the analytics. This module exemplifies the first principle of the adoption of emerging technologies from Chapter 2. It emphasizes the accountability and reliability are the logical starting point, and it requires traceable data flow within the analytics. In addition, in the data platform module, we answer some initial questions about how to derive analyzable attributes for the CADA. An obstructed scenario often results when audits cannot apply data analytics even though a huge amount of raw data exists. In this design, we exemplified how to utilize feature engineering to transform raw data into straightforward decision-making actions, which can act as the condition part of the "IF-THEN" rules. After this transformation, these attributes can be used as a direct ingredient of control rules for real-time monitoring. The analytical logic in the module of the rule-based-scoring-system is straightforward. The validation is directly from the management and the violators of these control rules. The analysis shows a high true positive ratio in the case study. The combination of the rule-based-scoring-system and the visualization strongly improve the analytics' persuasiveness.

The prescriptive analytics module achieves the goal of identifying the best option to choose the desired outcome through the related risk comparison. It acts as the primary analytics part of the design. The combined MCA and risk comparison show its ability to discover potential risky control areas by using dimensionality reduction. Unlike supervised learning algorithms, this analytics depends on the assumption that the result of the test of the current ICs acts as a “benchmark” of the quality of an internal control system. But the “benchmark” is dynamic and may change after pruning the current internal control system. The result can only show significance per confidence level to arouse the management's attention for the data-driven findings. The management will act as the authority to verify the data-driven conclusions. This combination demonstrates its practical value in developing internal control intelligence.

3.5.2. Limitations and Possible Future Research

The design of the CADA schema provides a practical tool to develop internal control intelligence. It also can be used to implement SOX 404 and Auditing Standard No. 2201 by exhaustively testing the transaction population. The schema offers a road map with evidentiary deduction, which links the operational features to a violation score by the bridge of the control rules. By using the business rule approach, the schema digitizes the entire internal control system. The CobiT 5 acts as a theoretical basis to determine the necessary data attributes for this sharable database.

A challenge for the development of internal control intelligence is that the management cannot determine all of the necessary controls and related control rules because of emerging risks and potential inherent risks. This paper offers MCA exploratory analytics as a solution to observe potential risky business actions. This solution assumes that the risky decisional alternatives are bound to leave a mark in the transactional information trace, and we still can figure out the potential risks using only parts of the information. With data pattern recognition, the MCA can filter out some possible risky decisional attributes to develop new control rules. The solution effectively explores inherent risks and quantifiable audit evidence for the quality of internal controls.

In this pilot study, we selected the P2P process to experiment with the design of the combining analytics schema and rule-based-scoring-system. The P2P process has limitations because it is related only to the organization's cost side, and external changes have no real-time influences within the selected process. However, the CADA prototype still illustrates its strength in developing internal control intelligence. In future research, we may implement the prototype in more complex procedures to observe the impact of interactions among different types of processes. As far as we know, this solution is the first exploration to solve data issues in the auditing area. Besides the practical value for the implementation of SOX 404, this study exemplifies its constructive function in the risk assessment and how to explore potentially risky internal control areas.

Chapter 4. The Ethical Implementation of Algorithm-based Decision-making: *An Analytic Roadmap to Examine Accountability and Responsibility of Designing and Implementing Audit Data Analytics*

4.1. Introduction

This essay aims to mitigate the potential ethical issues around the design and implementation of audit data analytics. We pay attention to identifying potential root causes that may occur throughout the entire design and implementation process, from data collection to analytics output. Munoko, Brown-Liburd, and Vasarhelyi (2020) proposed the ethical implications of using artificial intelligence (AI) in auditing. They present an extensive conceptual analysis of potential ethical issues that the audit profession should consider as the audits adopt AI. This essay follows Munoko et al. (2020) and uses a futuristic (i.e., forward-looking) approach (Brey, 2012). However, the difference is that we take a practice-relevant perspective to identify and manage ethical issues from both the design and application process. We specifically focus on algorithmic Audit Data Analytics (ADA), an augmented AI in the framework of Munoko et al. (2020). They discussed the ethical implications of AI in auditing based on the technology level, artifact level, and application level. Moreover, they also find that ethical issues may overlap across the three levels. This essay aims to complement Munoko et al. (2020) by studying the dimensions from the four components of audit data analytics: data source, training data, algorithm development, and output interpretation. Thus, direct insight into managing ethical issues of audit data analytics in the analytics schema design and application is gained.

The focus on ethics in this essay is on the ethics particular to the audit analytics system, which includes activities in the development and application process. In data science, the data value chain provides a conceptual framework to describe where data is identified, acquired, processed, stored, analyzed, and finally utilized by enterprises to add value (GSMA, 2018). Ethical issues can be produced in each procedure of the data value chain. The ethical subject includes algorithms, agents, data input, and analytics output. The source of ADA's ethical issue comes to both data and algorithm, which can initially inject biases and risks to hinder the ADA's effectiveness. From a designer's perspective, two issues are the root causes of these risks. The first is that there are understanding variances about data inputs. Emerging data and technologies feed ADA a variety of non-financial data and unstructured Big Data. This expanded data input requires strict guidelines about defining data and managing data traceability (according to ISO 9000, data traceability is the capability to verify the history, location, or application of the used data by means of documented identification). These potential risks can violate the ethical principles of beneficence and non-maleficence. The data issues can directly paralyze the ADA system and impact the auditors' professional competency and justice. The second is about the complexity of the algorithms. The opaque of algorithms offers some stakeholders an opportunity to take advantage of other users of the ADA project. We follow the definition of *algorithm* from Hill, who coined the term as a mathematical construct with "a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given provisions" (Hill, 2015, P.11). Algorithms can extend humans' physical existence, and this scale-up can amplify and accelerate justice and auditors' autonomy in auditing.

Data and algorithms can impact the moral problems in the design process of the ADA project. Furthermore, the root cause can be embedded in the system at the very beginning of the design. Thus, it is necessary to take a proactive guide to mitigate the ethics issues at the beginning of the design with a futuristic approach (Brey, 2012). We forwardly predict what can happen and mitigate the negative impact in advance. These issues are technically common and legal, but potentially damage the effectiveness of decisions based on algorithmic functions. It is difficult to manage these issues practically because the problems may be unintentionally ignored but have destructive consequences for ADA. Thus, the developers need to shoulder the responsibility to mitigate these potential risks by identifying their root causes from the peculiarities of the analytical mechanism of ADA.

The issues about data and algorithms also come from the interaction between the emerging technologies and audit context. This interaction can intensify the uncertainty of the ADA's consequences. It can be handled by taking proactive action if the potential result could be foreseen (Gillespie, 2016). The ADA causes some ethical issues in related decision-making also because it can scale up potential human bias. Algorithms' biases (Luca et al., 2016) can impact audit evidence's relevance and reliability. Biased analytics can increase risks that negatively impact decision-making routines and business policies, scale-up beneficence and non-maleficence problems. Besides, there are negative social perceptions about data analytics, including potential privacy invasion and possible discrimination. These risks can elevate institutional inertia to accept ADA as a routine decision-making tool even though audit analytics, technically, has a reputation for an excellent performance. Managing these potential issues can proactively help optimize the ADA's strength.

Therefore, we need a particular methodology to handle related ethical issues in the development of ADA. We intend to increase awareness of the problem that may arise when decisions are made using algorithmic functions in general and specifically in an auditing context. Auditing data analytics uses emerging data and technologies to perform a series of auditing procedures to collect audit evidence (AICPA, 2017). This consideration is consistent with calls for research about how emerging technology impacts accountants' ethical decision-making (PCAOB, 2019). For example, there is a need for research to gain an understanding of ethical issues related to the amount and types of data recorded, how the information is shared, and ethical concerns about the decision rules based on an algorithm-based software application, including data privacy, and abuses of investigative power.

One precondition of the management of ethical issues is to identify the root cause of the potential risks. It is necessary to "give comprehensive and thorough evaluations of computer systems and their impacts" (ACM, 2018, P. 8). The literature has provided viewpoints for this examination. We combine Floridi's (2008) information ethics framework and Martin's (2018) four algorithm ethics dimensions to theorize eleven ethical principles for audit analytics. Both of these two theories provide practical dimensions to trace potential biases that can cause issues for the accountability and responsibility of the ADA. As ACM states, "a system for which future risks cannot be reliably predicted requires frequent reassessment of risk as the system evolves in use or not be deployed. Any issues that might result in major risk must be reported to appropriate parties" (ACM, 2018, P.8). This essay constructs a four-phase ethics assessment roadmap to guide how to evaluate and deal with these potential moral problems based on Wright's (2011) assessment map. The four stages are precisely matched to the peculiarities of the ADA. Thus, this assessment map helps to recognize and take special care of the analytics systems that become integrated into the infrastructure of the audited institution. The four stages include the series of assessments about the data source, training data selection, the algorithm's traceability, and the interpretation of the output. The roadmap opens a detailed discussion about proactively identifying and managing the potential ethical issues for audit evidence.

We use a case study to demonstrate how these ethical issues were considered and addressed in the development of an ADA project. The case was to design a Continuous Monitoring and Control System for a state university's procure-to-payment process, to exemplify the implementation of the four-phase management roadmap. The case study provides specific evidence for the validation of the analytical roadmap of the ADA. We exemplify the approach by using the data from the design process, including weekly WebEx meetings, some specific tests from the aspects of the eleven mentioned ethical dimensions. We code the evidence from the case study by taking the Value-Focused Thinking methodology (Sweeny, 1998). The case study exemplifies how these ethical issues were considered and managed throughout the process to develop the Continuous Monitoring and Control System.

This essay is arranged as follows. The second section introduces the futuristic ethics analysis framework. Following this approach, we conduct a literature review about how the technical components of ADA can potentially lead to moral problems in accounting and auditing context. In the fourth section, we use the combined ethical impact assessment checklist from Floridi (2008) and Martin (2018) to identify the eleven ethics dimensions from ADA's development and implementation. Then, we construct an assessment framework to explore how to avoid these moral problems in ADA's design and implementation. The following section is a participatory case study to exemplify our methodology.

4.2. Futuristic Ethics Framework

We follow Munoko et al. (2020) to adopt a futuristic ethics framework (Brey, 2012) to analyze and forecast the potential factors that can lead to ADA's ethical issues. The futuristic approach initiate dialogue on the potential factors derived from the data value chain of ADA. Specifically, we follow the Anticipatory Technology Ethics (ATE) Framework (Brey, 2012) to explore the root cause of ethical issues from the technology level, artifact level, and application level. As Table 4.1 shows, at the technology level, we focus on data and algorithms to explore how these two components of ADA raise potential moral problems. The analysis at the artifact level pays attention to the risks from the interaction of data and algorithms. The application-level analysis takes care of the ethical issues aroused by stakeholders in the implementation process. These three levels of ethics analysis provide an opportunity to use an ethical checklist to enable the contemplation of ethical questions at each level of ADA.

Table 4.1: Ethical Framework for Audit Data Analytics

(Adopted from the combination of Table 2 and Figure 2 in Munoko et al., 2020)

Steps followed in this paper	ATE Framework	Description of Analysis Performed at Each Level		
		Capabilities	Constraints	Social and Ethical Impact
Step 1: Technology analysis	Data: the potential impact of data inputs for the ADA ethics concerns.	Describe how data inputs impact the ADA's analytical effects.	(1) Data Traceability; (2) The effectiveness of training data; (3) Data attributes have ethical implications.	(1) Improve auditor's professional competence. (2) Improve auditing due care. (3) Improve the justice of the attributes selection.

	Algorithm: the potential impact of the algorithm for the ADA ethics concerns.	Describe how algorithm impacts the analytical effects of ADA.	How do we balance the effectiveness and the transparency of used algorithms?	(1) Improve auditor's professional competence. (2) Improve the audit evidence's accountability. (3) Manage the ADA's non-maleficence.
Step 2: Artifact analysis	Artifact: The potential impact of the interaction of data and algorithms for the ADA ethics concerns, including parameter and control rules settings.	Describe how the interaction of data inputs and algorithms impacts the analytical effects of ADA.	How do we improve the justice of the control rule and the parameter setting?	(1) Improve professional Skepticism. (2) Improve the justice of the ADA project.
Step 3: Application analysis	The effectiveness and friendliness of the ADA.	Describe how the ADA system improves the auditing's effectiveness and efficiency.	How do we improve the easiness and friendliness of the design?	(1) Improve the ADA system's accountability. (2) Improve users' autonomy. (3) Improve professional skepticism.
Step 4: Assessment and management	Ethical issues mitigation	How to explore and utilize effective design skills to avoid causing moral problems.	(1) Making the ADA's working mechanism transparent. (2) Using the technological standardization extend the ADA's fairness design.	(1) Manage institutional inertia to adopt an ADA project. (2) Improve technology's persuasiveness.

In general, the futuristic analysis method adopts a four-procedure approach to discuss the potential root causes of the ADA and how to avoid the moral problems in its design and implementation.

Procedure 1: We surveyed related literature to explore how data and algorithms lead to ADA moral problems. The discussion is on the technology level, and the scope is from the nature of the technological components. We pay attention to the relationship between the inherent peculiarity of data and algorithm and the potential ethical issues.

Procedure 2: We used an ethics checklist and related exploratory questions to discuss ethical implementation for each part of the ADA system. The scope is from how to use ADA in auditing context. The benefit of following a set of predefined ethical questions is that we can have practical guidance about how to deal with potential issues in a given user context. We combined Floridi's framework of information ethics (2008) and Martin's algorithm accountability model (2018) to develop the technology ethics checklist (Figure 4.1 in section 4.4 provides a summary of the ethical questions).

Procedure 3: Develop an assessment framework to explore how to avoid moral problems in the design and implementation of the ADA.

Procedure 4: Use the developed framework to inject the ethical solutions to design and implement a P2P audit data analytics project in a state university.

4.3. The Nature of the Ethical Issues Arising from the Design and Implementation of Audit Data Analytics—a General Literature Review

4.3.1. Audit Analytics Has Ethical Issues

Beauchamp (1982) derived four properties associated with ethical judgments. First, the decision-making alternatives need ethical judgments to assess these potential choices. Second, the judgments are prescriptive rather than descriptive, meaning they are used to suggest what actions should be taken. Third, judgments are universally valuable to any person in similar situations. Fourth, they pertain to the welfare of others, rather than simply to the effects on oneself (Beauchamp, 1982). These four properties provide useful guidance to gain an understanding of the ethical issues in algorithmic audit analytics. These issues are related to moral responsibility and accountability, and they usually lead to a series of consequential ethics (Wright, 2011), like beneficence, discrimination, and justice.

The nature of auditing is tightly related to accountability. According to the Merriam-Webster Dictionary, the definition of *Accountability* is “the quality or state of being accountable.” Primarily, it means an “obligation or willingness to accept responsibility or to account for one’s actions.” Accountability in the business world means an explanation or justification to stakeholders for one’s judgments, intentions, and actions. Responsibility is defined as a bundle of obligations associated with a role. Accountability is also firmly associated with a recognized responsibility (The Arthur W. Page Center, 2020, pp.2). In the auditing context, accountability is about answering to clients, colleagues, and other relevant professionals about the evidence of the auditing judgment, and explaining how to develop the related evidence. It is also associated with responsiveness to all stakeholders' views, which includes a willingness to explain, defend, and justify actions.

In the Audit Data Analytics context, emerging technologies may transform and exchange information in a profound way. This context may add impressive new dimensions to old problems and lead to new ethical issues. This obscurity to trace the algorithm's mechanic mechanism makes it difficult to trace the lines of responsibility and accountability. There are a variety of technical topics that address this issue. For example, the facial recognition system's mechanism will depend on both the training data selection and the algorithms’ mechanical capability. Thus, the system cannot work well for other racial people if the training data is dominated by one kind of racial people (Dwoskin, 2015). This chapter focuses only on some ethical issues that are technically common and legal but potentially damage the effectiveness of algorithmic audit data analytics.

4.3.2. Data-driven Information Technology and Moral Philosophy

The relationship between information technology and moral philosophy dates back to Norbert Wiener (1954). Information technology can extend humankind's physical existence and the power of perception. This extension serves to forward an expansion of human senses and capabilities of action (Weiner, 1954). Thus, Weiner foresaw that machines would be integrated into the social fabric. They would create, store, and receive messages from the external world. They would also make decisions and take actions. Additionally, they would be merged with human bodies to form beings with vast new powers. This process exchanges information and generates moral issues (Weiner, 1964). Further, Moor (1996) argues that the "logical malleability" of computers (adjustable algorithms in this case) leads to so-called policy vacuums that require careful ethical analysis to fill. Thus, it is necessary to work through the moral issues in these new areas (Moor 1996).

Moor explained that ethical issues related to information technology arise because we need to use the "digitized" sense to understand the activities in old forms (Moor 1996). This "informational enrichment" can change the meanings of old terms, creating "conceptual muddles" that have to be clarified before new policies can be formulated (Weckert et al., 2008, P. 135). From the perspective of value creation, the ADA is to apply emerging data and technologies to transform data into useful knowledge to generate audit evidence (AICPA, 2017). Data is the raw materials, and emerging technology is the tool. In data science, the data value chain provides a conceptual framework to describe where data sources are identified, acquired, processed, stored, analyzed, and finally utilized by enterprises to add value (GSMA, 2018).

Ethical issues can be produced in each procedure of the data value chain. First, in the data acquisition procedure, ethical issues can appear if data providers and data analysts have an inconsistent understanding of what the data stand for. This inconsistency can create bias and conflict of interest (Weckert et al., 2008). It provides auditors' potentials to override professional or business judgment. Second, in the data storage procedure, data traceability can raise ethical issues by hindering auditors from tracing the data source of audit evidence. Traceability is the capability to verify the history, location, or application of a data item using documented identification. Thus, data traceability can impact the ADA's professional competency (Moor, 2008). Third, in the analysis procedure, training data can produce potential moral problems. A remarkable peculiarity of ADA is that the analytical results only are responsible for the input of training data. ADA needs to manage ethical issues by guaranteeing that the training data's context has no material difference from the working data. The moral issue from training data can reduce professional audit competence and due care (Floridi, 2008). Last but not least, in the analysis procedure, attributes selection also can produce a justice ethical issue of the ADA. The ADA design needs to avoid gender, race, and even Zip Code because these attributes are highly correlated to human social and economic groups (Wright, 2011).

The data issues can cause a beneficence issue if the ADA cannot take positive steps to prevent harm with its blurry data. We need to manage these issues at the beginning of the design process. The solution to address the ethical implications is to provide traceability and accountability of the evidence from the "digitized" process. Moor (2008, P. 77) demonstrates this problem in "[M]oor's Law: As technological revolutions increase their social impact, ethical problems increase."

4.3.3. Algorithms can Scale Up Human's Biases and Arise Ethical Issues

As introduced in the previous section, *an algorithm* is a mathematical construct with "a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given provisions" (Hill, 2015, P.11). Algorithms, like other computer systems, are designed to represent the interests of their users. Thus, algorithms are typical surrogate agents. Thus, the result of the algorithmic solution mainly depended on both data inputs and the mechanical capability of the machine. There are two ways that the surrogate system can transmit ethical problems. The first is incompetency, which means that the systems may be useless for some stakeholders, and they are only useful for some specific stakeholders. Most research focuses on the mechanical parts, which meet the system's accuracy and predictive ability, not on how the system can help achieve a practical task. The other is misbehavior, which means that surrogate agents can harm when used to intentionally violate one of the role constraints, e.g., the system ignores the violation of segregation of duties. One user can access two roles, and this behavior may lead to fraudulent activities. This misbehavior is the designer's responsibility, but the system acts as the surrogate agent (Johnson et al. 2008).

In audit data analytics, algorithms can scale up human ethical biases from those who have a dominant voice in rules development. This dominance is especially true in the context of internal controls if the control rules that were developed as algorithmic control to exert real-time monitoring. The rules can act as guidance to lead managerial behaviors. The moral problems can be scaled up if the algorithmic rules had been injected ethical biases. Thus, in ADA, the emerging technology can potentially be used as one interest group to take advantage of some other groups.

On the one hand, emerging data and technologies can improve information processing efficiency. On the other hand, algorithmic decision-making can easily lead to bias and increase risks. Gillespie (2013, P. 189) cautions that business leaders need to pay “close attention to where and in what ways the introduction of algorithms into human knowledge practices may have political ramifications.” Four decades ago, Winner (1980) claimed that technical things are political. First, the users can selectively choose some features in the application. Second, some stakeholders can claim their interests by setting parameters in the algorithm. Also, Keeney (1984) states that ethical implications could be inherently part of the methodology being utilized in the specific evaluation and choice from a series of alternatives.

It is unnecessary to worry about the computer system getting distracted, stealing, being lazy, or going on strike because of its strict mathematical construct. Computer systems exactly do what they are programmed to do. The sound mathematical logic and calculative ability make algorithms achieve more accurate predictions that can support more effective decision-making. As a surrogate agent, the algorithms must be designed to avoid engaging in actions that violate the users' intentions. As such, the algorithms should be programmed to serve the interests of some groups other than their users. They may even be designed in ways that conflict with or undermine the interests of their users. In other words, misbehavior can occur when analytical algorithms are developed in ways that achieve the benefits of the third party other than the users.

Luca (2016) stated that algorithms, unlike humans, are only machines and have no ethical constraints, and only pursue a specified objective single-mindedly. In other words, computers can inject bias in decision-making without suitable ethical restrictions. Thus, first, to address the ethical issues, we must have the reasonable control of responsibility, liability, and blame in situations in which multiple and diverse agents are at work. There will typically be at least three agencies at work—end-users, systems designers, and analytical algorithm programmers in cases of audit data analytics. Second, we must fully understand the computer logic and the traceability of the audit evidence.

4.3.4. The Ethics of Audit Data Analytics Is Consequential Ethics

The algorithms in audit analytics "mediate social processes, business transactions, governmental decisions, and how we perceive, understand, and interact with the environment (Gillespie, 2016, pp.5)." Hence, algorithms can potentially act as the function to claim the interest for one or more stakeholders (Gillespie 2016). The gap can have severe consequences if there is a difference between the design of algorithms and the users' values. In audit analytics, the use of "algorithms" also includes implementing the mathematical construct into technology and applying the technology configured for the specific task. Given this clarification, the configuration of an algorithm to a specific dataset can include a further tweaking of the algorithm's operation related to the specific tasks and queries. Fountain (2001) showed that information technology has been impacting institutional change and the organization's hierarchy. The "political" feature of IT can negatively influence the decision maker's behavior, which has ethical implications. Orlikowski (2008) suggests that the IT application in the decision-making process has both social and technical sides.

Based on the discussion above, we consider ethical issues related to the two sides of the ADA system: the mathematical construct, the generally reliable part, and the subjective human capacities for action and comprehension. As such, we explore suitable solutions to constrain these potential problems. Ethics, in this case, is "consequential ethics" (Wright, 2011) from the designer's role. Two essential principles are used to define consequentialist ethics. First, it requires that the ethical actions depend only on the relevant consequences to society, not on any prior concept of moral rights. Second, the ethical choice is the one that can maximize the value of the consequences to society, where the consequences to be evaluated are specified independently of any moral duties (Keeney, 1984). Specifically, the system should be designed using the principles of consequentialist ethics. In other words, we need to foresee and take proactive actions for some potential issues that could lead to negative results. The consequential ethics emphasizes the necessity to take proactive actions to mitigate the potential negative consequences that may occur when designing algorithms for audit analytics techniques.

4.4. Dimensions of Ethical Issues in Audit Analytics

4.4.1. A framework for the Ethics Checklist of Audit Data Analytics

It is common to use an ethics checklist and exploratory questions in artificial intelligence ethics research (Munoko et al., 2020). Floridi (2008) proposed a framework of information ethics. He separated the ethics framework into three parts: information-as-a-resource ethics, information-as-a-product ethics, and information-as-a-target ethics. Our study extends Floridi's (2008) category to four pieces, including data source, training data, algorithm, and outcome. This framework can include all of the activities in the data value chain of the ADA. We develop an ADA ethical framework with the ADA's four parts. The strength of this framework is that we can trace the source of the potential ethical issues within the four components of the ADA in the development and implementation process. The framework that we use also keeps consistent with an algorithm accountability model by Martin (2018). As Figure 4.1 shows, it has four parts: training data, source data, algorithm, and output. Wright (2011) proposed an ethical impact assessment of information technology with the perspective of values stated in the EU Reform Treaty. The benefits include human dignity, freedom, democracy, human rights protection, pluralism, non-discrimination, tolerance, justice, solidarity, and gender equality. These values are also stated in the Charter of Fundamental Rights of the European Union. Furthermore, this essay integrated technology, artifact, and application levels by observing the root cause of ethical issues.

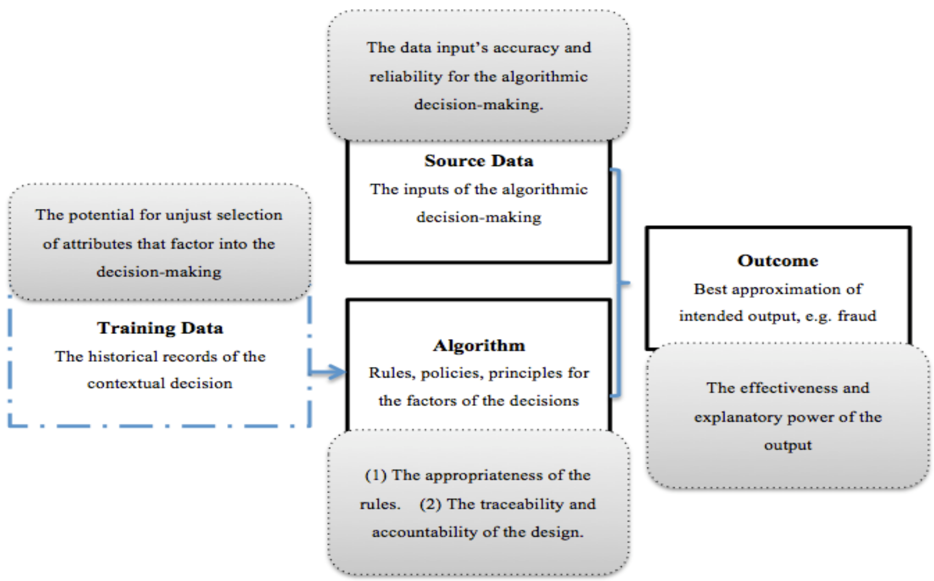

Figure 4.1: Framework of the Ethics of Audit Analytics (*this figure is modified from the Figure 3 in (Martin, 2018), and the shaded areas indicate potential ethical issues)*

In the next section, we follow Martin (2018) to develop eleven ethics dimensions of audit data analytics from the perspectives of training data, data input, analytical algorithm, and analytical output. Also, we follow Wright (2011) to use exploratory questions to address potential ethical concerns in each dimension. As Wright (2011, pp.204) claims, "to some extent, the issues and questions set out here should be regarded as indicative rather than comprehensive." In each dimension, we borrow a modified Value-Focused Thinking (VFT) approach (Keeney, 1996) to derive the ethical implications. The approach is called a modified Value-Focused Thinking approach because it follows the logic principle of this approach, but we did not develop the fundamental objective metrics, the means-ends objective network. The indicative questions are considered as a guide for the fundamental objectives and their attributes. Specifically, we categorize the process of conducting the ethical objective (value) and its attributes analysis by the following five steps. The first step is to identify the ethical values when concerning training data, data input, analytical algorithm, and analytical output. The second step is to specify ethical attributes and potential measures to indicate the degree to which they are achieved. In the third step, we assess the likely ethical implications of each alternative in terms of the ethical attribute. Then, the fourth step is to develop a series of questions to indicate potential ethical issues. The last step is to integrate the indicative questions into the ethical value structure.

4.4.2. Data Source and Its Ethical Implementation

Gitelman (2013) states that "raw data" is an oxymoron. Indeed, all data provide oligopolistic views of the world. Different scopes or different measurement instruments can lead to different data, which cannot be an all-seeing, infallible God's eye view (Amin and Thrift, 2002; Haraway, 1991). As such, it is necessary to understand data from the way they are generated and from where data come before performing data analytics. Data are not merely natural and fact-based elements abstracted from the world in neutral and objective ways and can be accepted at face value. In the measuring process, the collection instrument and related approach can shape their constitution (Ribes and Jackson, 2013).

It is an illusion that data analytics itself can automatically discover insights without testing scientifically for validity and integrity. In other words, algorithms need managers too, and only managers can understand the nature of the targeted data in the specific domain (Luca et al., 2016). Furthermore, data cannot directly speak for themselves free of human bias or framing. We always hold some conceptual framework, which may be intuitive or has biased understanding, to the task of data analytics and related interpretation. Understanding the quality of data is always relevant; data have to be examined through a particular lens that influences how they are transformed and interpreted. Moreover, we are required to use domain knowledge to understand the real economic meaning of data. Specifically, in audit analytics, data preparation is a fundamental requirement and often poses a variety of challenges, including reliability assessment, access detection, and data standardization. Practically, this issue is heavily related to the existing data legacy and current information system in the ADA project.

From the discussion above, we can derive two ethical dimensions related to the data source: the first is "data quality," and the other is the "transparency" of the data source.

Ethics dimension 1: the quality and completeness of data input

One source with highly risky bias, technically, is missing values. Emerging data and technologies act as the fundamental function in the ADA system to conduct audit procedures and collect evidence (AICPA, 2017). But some statistical programs that can automatically deal with missing data cannot be used in the ADA system. The reason is that some commonly used statistical techniques may cause significant ethical problems in auditing. For example, it is common to use an average number to substitute missing data or moving the records with missing data for statistical purposes. However, the fact that data is missing may be caused by fraudulent behavior. Audit data analytics cannot directly ignore the issues of missing data, and audits need a specific investigation of the missing data. We may miss the opportunities to detect fraudulent behavior if we cannot proactively understand the issues of missing data in ADA. We can imagine a scenario: Z is a senior manager who has access to create an active customer master file. Z successfully put his own company as an active vendor. Then Z filed several purchase orders to his private company. Two months later, he forged an item of the received document and paid a sum to his own company even though no products have been received, and simultaneously Z deleted his own company from the list of active vendors. In the auditing system, these transactional records have missing values. Z impacted the audit department's integrity, which removed the transactions because they have missing values.

Thus, an investigation is necessary for incomplete data input. The analysis of the data pattern of the missing data is a useful approach to detect potential fraud and inefficiencies of internal control. The following four questions are the related exploratory questions for data quality: (1) What assurances exist that the data collected are accurate and complete? (2) Have the data been obtained from others than the person to whom they pertain? (3) Can we find a particular pattern for the records that have missing values? (4) If the data collected are not accurate or lost, what consequences might ensue?

Ethics dimension 2: data traceability and its transparent data source

According to ISO 9000, data traceability is the capability to trace something. It means the ability to verify the history, location, or application of a data item utilizing documented identification. Data traceability acts as an essential function in the context of

ADA. A transparent data source means that auditors must understand the data source that is used in audit analytics. Transparency is required to understand where the data inputs come from, who can update the data sources, and how the data's values have been defined. It is a precondition for auditors to have an understanding of the meaning and the source of the data used in audit analytics. For example, Z is an internal auditor in a university. He proposed using Big Data analytics to compare the price from the online search with the purchase order's price. His proposal got substantial resistance from managers because the data source has noise, and the information entropy is high. The use of the data with high entropy impacts the independence of the management system. In other words, the noisy audit opinions from blurry data analytics can constrain managers when they make decisions.

The following three exploratory questions are for the "transparency" of data sources: (1) What information elements are necessary to apply audit data analytics? Which data source can provide the necessary information elements? (2) Is the targeted data source the original, or was it transformed from other sources? How can one ensure its reliability if it went through a transformation? (3) Can anyone separate the "clean" data from the "noisy" data? Can one ascribe the responsibility to related stakeholders about the unqualified data?

4.4.3. Training Data and Its Ethical Implementation

In algorithmic decision-making, training data can lead to ethical issues from three perspectives: social sorting, attribute selection, and historical biases. As Martin (2018, pp.13) states, "training data is often gathered from people who inspect thousands of examples and tag each instance according to its category. The algorithm learns how to classify based on the definitions and criteria humans used to produce the training data, thus potentially introducing human bias into the classifier."

Ethics dimension 3: social sorting and discrimination

According to Wright (2011), profiling technologies lead to ethical or legal issues, including privacy, equality, security, and liability. Profiling technologies are, by their very nature, discriminatory tools. They can produce non-parallel types of social sorting and segmentation, which could have adverse effects. For example, a facial recognition program could recognize white faces but be less effective in detecting faces of non-whites. The scientist "eventually traced the error back to the source: In his original data set of about 5000 images, whites predominated" (Dwoskin, 2015, pp.128). The algorithm designers did not write the algorithm to focus on white individuals; however, the training data used to train the algorithm contained predominately white faces.

Thus, in the procedure for the selection of the training sample, it is necessary to avoid discrimination and social sorting. The following three exploratory questions are for the dimension of discrimination and social sorting: (1) Does the project or service facilitate social sorting? (2) Could the project be perceived as prejudice against some groups? If this probability exists, what measures could be used to help avoid discrimination? (3) Will some groups have to pay more than other groups?

Ethics dimension 4: attribute selection biases

Human decision-makers usually have a mental shortcut (heuristic), like anchoring or availability. Familiar assumptions can trigger data analysts to choose familiar attributes as factors in the data model unintentionally. This tendency can lead to biases. For example, a manager is preparing to select attributes to assess the risk of potential loans.

The accumulated experience makes the manager believe that the postal code can be a useful factor because no loan violations happened in the community that he/she has worked. This attribute can lead to severe discrimination; the postal code could be highly correlated to race and family income.

In the exploration of attribute selection, we need to ask the following questions: (1) Are the selected attributes identifiers of particular groups, and do those attributes have the potential to cause discrimination? (2) Are the characteristics highly correlated to particular groups? (3) How can one assure the relevancy of the factors to the expected decisions?

Ethics dimension 5: historical biases

Data analytics often uses historical data as a training sample. The algorithm will learn historical biases in the training data in the training stage, so past discrimination will be coded into the algorithm. It is necessary to periodically compare the new context of algorithmic decision-making, and refine the data model with matched training data. For example, in a financial decision, most data models avoid using the data from 2008 to 2009 because the period of the financial crisis has severe historical biases.

In the exploration of the training sample, we need to ask the following questions: (1) Did the context of the decision-making substantially change from the training data period? (2) Does the training sample have prominent historical biases?

4.4.4. Algorithm and Its Ethical Implementation

According to Luca et al. (2016), machine learning can lead to biases, which are insidious and more challenging to identify because the preference is yet another level removed from the outcome. In the audit area, the initial motivation to use emerging data and technologies is to improve the effectiveness and efficiency of auditing tasks. In other words, the algorithms help improve accuracy and achieve real-time monitoring. However, algorithms are only tools; the devices' effectiveness is highly related to their application contexts. Thus, algorithmic decision-making also is related to four ethical dimensions, including beneficence, sustainability, justice, and transparency. We demonstrate the four aspects with related exploratory questions.

Ethics dimension 6: the beneficence and sustainability of algorithms

Beauchamp and Childress (1994) claimed that the principle of beneficence is that machines must be useful, and it needs to be perceived useful as well. They claimed two principles of beneficence. The first is positive beneficence, which requires benefits. The second is utility, which requires the trade-off between benefits and drawbacks. This dimension shows that audit analytics should benefit the organization. The algorithms used in the ADA project need to be traceable. The algorithm's traceability takes a positive step to prevent harm. Otherwise, the outputs of audit analytics could cause institutional "political muddle."

The traceability of the algorithms helps to achieve the goal of independence. The traceable mechanism of audit analytics is a self-evident explanation when managers have disagreements with internal auditors. However, the phenomena can be in chaos if the internal auditor cannot give a clear interpretation of the mechanism of the algorithm used. One can imagine the chaotic situation if the auditors cannot explain the analytical results that provide a lousy performance measurement to a manager who feels that the result is unfair.

Thus, it is necessary to be mindful of the following exploratory questions about the beneficence of ADA: (1) Will ADA benefit individuals? If so, how does this mechanism work? (2) Who can benefit from the IT-supported analytics, and how? (3) What are the most prominent consequences, including the benefits and drawbacks of not proceeding with the analytics development? (4) Does the expected analytics serve broad community values or only a specific group? What are these, and how are they served? (5) Does the expected audit analytics have a negative impact on the institution's political hierarchy?

Ethics dimension 7: justice of the algorithm

This dimension aims to achieve fairness in the conventional design. Let's imagine the following scenario: Z, a senior manager, holds responsibility in the purchasing department. Z also can impact the design of the monitoring system. Z morally hates staff that holds the invoice too long and delays the payments. Z adds an unfairly higher punishment weight for this specific behavior. Thus, it is necessary to ask the following exploratory question that is related to the dimension of justice. Does the expected audit analytics confer benefits on some groups but not on others? If so, how is it justified in doing so?

Ethics dimension 8: transparency of the algorithm

This dimension aims to avoid the "Black-Box" in audit analytics. The opacity of algorithms leads to more severe ethical issues than traceability. The issue of transparency has gotten big attention in other areas, like risk management in finance. As Gillespie claims, transparency can provide a platform to allow individuals to "game" the system. People with priorities may make themselves algorithmically recognizable and orient their data to be viewed favorably by the algorithm (Gillespie, 2016). Thus, the potential goal could help some groups have more advantages than others, thereby creating a new disparity and "political muddles" to reconcile (Bambauer, 2017). Furthermore, algorithmic opacity is also framed as a form of proprietary protection or corporate secrecy (Pasquale, 2015). This consideration is more critical in the audit area, where intentional obscurity is designed to avoid scrutiny (Burrell, 2016; Pasquale, 2015). For example, an internal audit department developed a practical scheme to detect fraud by the employees who violate the rule to do business with their own companies. It is better to keep it secret because the potentially fraudulent employees could design a new mechanism to outmaneuver the mechanism.

Thus, it is good to think through the following exploratory questions about the transparency of the algorithm: (1) Has an analysis been made of who are the relevant stakeholders? (2) Are studies about the pros and cons of the project or technology available to the public? (3) Do one or two of the stakeholders of the project dominate the design? (4) Can the mechanism of the audit analytics be traceable for its reliability and relevance?

4.4.5. Analytical Output and Its Ethical Implementation

As discussed above, algorithms and related audit analytics unavoidably have two features: the sense-making of the algorithm and the understanding of data inputs. These two features can lead to some unintentional biases, which is hard to detect from the analytical process. The explanation of the results is an inseparable part of data analytics; the outputs of algorithms require reasonable decipher (i.e., what one should do based on what the algorithm indicates) (Luca et al., 2016). For example, correlation detection is a widely used algorithm in analytics, but without domain knowledge, the result of the highly significant correlation is nothing that can help decision-making. Different metrics "make visible aspects of individuals and groups that are not otherwise perceptible" (Lupton, 2014, 859).

In this area, the ethical issue is about the willingness to explore better job performance if the available technology can help achieve a higher goal. In other words, the designer and end-users have ethical discretions. They can submit more understandable visualized analytical results, or output the simple statistical analysis results and require them to learn to understand the consequences. Thus, the ethical issue is also tightly related to "ethical laziness."

Two considerations are related to the interpretation of the outputs of data analytics for the final automatic control. The first is about how to post the result to appropriate management. For some control activities, visualization can help understand the rule violation (e.g., the expense changes over a long period). Some others need the absolute numbers for the indicators (e.g., it takes a long time to hold the invoice payment, which wastes the available discount). The second is what kinds of results of data analytics can be the indicators of electronic monitoring. The choice is extensive within Big Data analytics; more and more behavioral data have been added to the analytics. This action is an ongoing exploratory process and needs relevant managers' involvement. A periodic revision mechanism is necessary to refine the selection of the control indicators and control rules.

Ethics dimension 9: accessibility (user-friendliness)

This dimension adds to the interpretative ability of algorithm-based decision-making. Auditing requires the interpretation of the evidence's relevancy and reliability. As discussed above, verification requires that analytical evidence has relevance and credibility. Audit Analytics is very different from predictive analytics in marketing areas, which pays more attention to predictive ability. The auditing needs more reliable and persuasive evidence for the end-users. Visualization can help auditors to communicate with management. Fekete et al. (2008, pp.23) showed that "[R]esearchers and users of Information Visualization are convinced that it has the cognitive benefit and perceptual support. These values can easily be communicated to others in a face-to-face setting, such that this value is experienced in practice."

The following exploratory questions relate to the dimensions of accessibility: (1) Does the new technology or application require a certain level of knowledge of IT that some related users may not have? (2) Do the end-users of the analytics understand how the result comes out clearly? (3) Do all of the stakeholders know the responsibility related to the analytical output?

Ethics dimension 10: the availability of assessment feedback

This dimension seeks to avoid a decrease in the negative impact of the false positive of the analytical output. Management needs a mechanism that alerts them of the wrong results from the machines. This feedback can help the ongoing operation of the IT-supported system and improve its effectiveness and efficiency. The feedback design provides end-users with a channel to engage them and solicit their voice to potentially keep them from "exiting" the current audit system. And further, this feedback also generates "loyalty" consistent with Hirschman's classic economic analysis (Hirschman 1970). Audit Analytics is an ongoing operation, so the engagement with the end-users can fine-tune the current mechanism and accumulate intelligence.

The following exploratory questions are related to the dimension of transparency. (1)

Who is responsible for the feedback from the end-user of the audit analytics? (2) Does the system have a mechanism to find the exact cause of the wrong result? (3) When the operation of audit analytics fails at its assigned task, who takes the blame? The programmers? The end-users?

Ethics dimension 11: data privacy protection

Data privacy is a critical issue in data analytics. Over eighty countries, including nearly every country in Europe and many in Latin America, have adopted a comprehensive data protection law (Greenleaf, 2014). In the context of audit analytics, data privacy is an ethical issue, mostly related to how to use and protect the analytical results. Auditors have access to a relevant data source in the analytical process to extract information for a specific audit goal, so no privacy violation exists. However, auditors should be very careful to ensure that the result is posted only to the relevant managers and end-users. The content cannot have identifiers, like SSN, name, address, and the like. We can imagine a scenario where Z is the director of the purchasing department in a company. Z received an internal auditor's email this afternoon, and Z found four transactions that violated the purchasing rule last week. But he was also surprised to see that the report has all of the identifier information, including SSN and date birth, of the four employees who violated the business rule. He asked the internal audit department, and the auditor explained that the disclosure was unintentional, and they would be careful next time. This accident indicates a bigger ethical problem for auditing than a mere mistake; the problem is systematic.

Thus, it is necessary to ask the following exploratory questions related to the dimension of data privacy: (1) What is the required content in the report of analytical results for the employees responsible for the violations? (2) How can one identify the employees responsible for each kind of a breach? (3) Who will trigger the action to post the analytical report? (4) Does the system have a mechanism to filter sensitive information from the analytical report?

4.5. The Construction of the Ethics Assessment Roadmap

Machines have ethical issues. Algorithmic ADA is inevitably value-laden because the data input and operational parameters are specified by designers and configured by users. The designers of the system have the discretion to filter information to privilege some values and interests over others (Brey and Soraker, 2009; Friedman and Nissenbaum, 1996; Johnson, 2006; Kraemer et al., 2011; Nakamura, 2013). The ethical issues can arise in some phases of the design and implementation of algorithmic decision-making project. Thus, machine ethics research should tightly relate the machine mechanic function with the application to specific domains. This combination provides an opportunity to observe how to apply ethical principles to particular real-life cases (Anderson and Anderson, 2007). In the following section, we discuss each moral dimension in each ADA procedure to develop the ethics assessment roadmap.

The proposed assessment map is structured based on how the analytics is organized and how algorithms operate. This ethical roadmap is a four-phase solution, which includes data quality, attributes selection, algorithm logic, and interpretation consideration. We want to map the data value chain to understand how ethical issues impact the evidence through the data flows and to find solutions to constrain the system design's potential problems. This framework verifies whether the targeted system has ethical issues and deals with the problems when ethical issues exist. The evaluation of the algorithms' performance, like accuracy, false negative, and false positives, is beyond this paper's scope.

Phase 1: Data traceability and data quality assessment

This stage assesses the data input of the analytics. We theorize two issues that need quality assurance. The first is to investigate the roots of the missing data. The second is to investigate the data traceability and reliability and separate questionable data from the clean data.

As discussed in the previous sections, audits must investigate the cause of missing data. Some fraudulent behaviors may hide behind the missing data, so the assessment has to ascribe the responsibility to related stakeholders. If no detailed investigation would be exerted for the missing data, the ethical issues exist because the system ignores the potential biases and attributes the responsibility to algorithms and machines. Moreover, ignoring missing data constitutes a loophole in the design where the fraudulent actor(s) can take advantage of this bias and have severe consequences. Audits also need to assess the accuracy, consistency, and completeness of the raw data. One of the solutions is to isolate flawed data by locating critical areas, identifying the impact domain associated with each instance of poor data quality. To improve the data quality in the following operation, we need to attribute the responsibility to related stakeholders.

Data quality is a multi-dimensional concept (Pipino et al., 2002). Audits need to assess data quality from both subjective and objective perspectives. Objective data quality assessment is to evaluate: (1) if the data recording provisions include all relevant information to identify data flow within the system; (2) if the regulatory requirements have a functional traceability system. Subjective data quality evaluation reflects stakeholders' needs and experiences: the collectors, database managers, and consumers of data products. The accurate assessment is related to the measurements based on the data set in question.

Practically, we need to develop both subjective and objective metrics to improve organizational data quality from the three steps: (1) Performing subjective and objective data quality assessments; (2) Comparing the results of the evaluations, identifying discrepancies, and investigating the causes of the differences; (3) Determining and taking necessary actions for improvement.

Phase 2: Attribute selection and discrimination assessment

This assessment is both a technical and an ethical issue about what kind of information elements should be chosen in a logical schema. Moral action preference is ultimately dependent on the selected features that actions involve, such as harm, benefit, respect for autonomy, and the like (Anderson and Anderson, 2014). For example, in a fraud detection schema, it is not ethical to choose gender and age as elements in the logical schema.

Discrimination means that some attributes can intentionally or unintentionally cause discriminatory decisions. Discrimination can be either direct or indirect. Direct discrimination consists of rules that explicitly mention minority or disadvantaged groups based on sensitive discriminatory attributes related to group membership. Indirect discrimination includes provisions that, while not expressly specifying discriminatory attributes, intentionally or unintentionally could generate biased decisions (Hajian et al., 2013).

Hajian et al. (2013) demonstrate a three-stage discrimination prevention method.

Stage 1: Transform or remove the discriminatory data attributes from the original data in the preprocessing stage. Stage 2: Change the algorithm in the processing stage. Stage 3: Modify the resulting data analytical model in the post-processing stage.

Phase 3: The assessment of algorithms' Accountability and traceability.

The *AccountAbility* coined Accountability as: "to account for something is to explain or justify the acts, omissions, risks, and dependencies for which one is responsible to people with a legitimate interest" (AccountAbility, 1999, pp.8). It belongs to that area of causal responsibility having to do with blame and punishment.

The goal of this assessment is to assure algorithmic accountability, which allows the results to be traced to the raw data sources and offers evidential links across the evidence. It is accurate to assess whether the algorithm's logic is traceable and transparent. In other words, the audit automatic control does not allow a "Black-Box" phenomenon, and all of the analytic philosophy can be interpretable. Audit Standards No. 15 mandates that audit evidence has to be relevant and reliable (PCAOB, 2009). Thus, this assessment is to assure that analytic outputs are free of ethics problems. On this issue, the literature discussed some methodologies, e.g., Allen et al. (2000, 2009) introduce Moral Turing Testing (MTT). Nevertheless, so far, MTT is still in the explorative stage. In this roadmap, we utilize subjective judgment to assess the logic with two procedures. First, the design can explain the reasoning in plain English. Second, the users can understand and agree that the output indicators can show the exceptions of the targeted operations.

Phase 4: The assessment of the interpretation of the analytical output

The fourth phase is to assess the ethical issues in information communication. The outputs of audit analytics, mostly, are simple indicators, like index or score. The bridge between the signs and the violations of the business rules should be understandable and acceptable. In this phase, we have two considerations. The first is the subjective judgment from the users about the transparency of the explanation. The second is about the ethical dimension of user-friendliness. In other words, the goal is to improve the effectiveness of the output interpretation. Some empirical results illustrate that some effective communication methods, like visualization, can add an understanding of the effect more directly and readily (Dilla et al., 2010).

4.6. The Participatory Case Study to Insert Ethics into the Design of a Procure-to-Payment Audit Analytics Project

4.6.1. Introduction to the Environment of the Project

We test the ethical assessment map in the same participatory case study as in the previous chapter by observing how to implant the ethical dimensions in a specific data-driven solution. The institution in this pilot study is a state university located in the northeast of the USA. The goal of the project is to develop a rule-based Continuous Monitoring and Control System for the assurance of the effectiveness and efficiency of the P2P process. The university constructed a cross-departmental project team, including the internal audit department, the purchasing department, the human resource department, the information system service department, data analysts, and consultants.

4.6.2. A Specific Assessment about the Ethical Issues about Data

The researchers deployed two quantifiable mechanisms to assure data quality. The first mechanism compares the aggregate amount of the transactional data and the balance of the account from the financial statement. The second mechanism is to trace the data flow in the analytical system with the identifier of the transactions. We identified and matched each transaction with the Purchase Order, Invoice Number, and Item Received Number. In these two tests, we found 12,234 transactions having no record for the item received. Technically, this missing information indicates a severe data quality issue. The post-investigation showed that the item users did not input the necessary data and documents into the system, and no supervisors monitored the documentation. This missing is a severe data issue in both the auditing and ethical viewpoints, so the project team reported this issue and removed the analytical audit protocol in this control activity. Simultaneously, we require enforcing the necessary information documentation for products received to activate the future monitoring function. Also, we exerted a qualitative survey to assess the safety and quality of the data repository. Specifically, for each information system used as data storage, we figured out the job position in charge of the data input and ensuring data security.

4.6.3. Evidence from the Design Process

We collected the data from the following documents for the ethical assessment: (1) the university policy files and the relevant documents from the project initiation, (2) the memos of the weekly WebEx meeting that involved the whole project team, including one audit manager, one audit consultant, two purchase managers, and two data analysts, and (3) the daily conversations and emails with relevant stakeholders. The evidence of the ethics assessment in the case study is demonstrated in Table 4.2.

Table 4.2: Evidence from Excerpts of WebEx Meetings and Communications in the Participatory Study

Ethics Dimension	Evidence	Status
Quality and completeness of data input	To assess data traceability and the source of each attribute, the design team spent twenty weekly meetings to ensure understanding each data attribute. The design team also communicated with the staff, which is directly responsible for the data quality, to ensure that we have a consistent understanding of the data.	Design Team
	The team suffered missing data about payment_date, and the audit manager responded, "Stop doing any analytics if we cannot assure the input data is qualified."	Audit Manager
Transparency of data source	The data analyst merged the data from Oracle Datawarehouse and SciQuest Purchase System and missed 5,346 transactions. The audit manager objected to train the remaining cleansed data for the monitoring model until we knew the reason that the data were missing: they were reversed transactions.	Audit Manager
	To monitor seeking qualified vendors, one data analyst proposed to compare the real unit prices with the prices from the analysis of the website search results. But the purchase manager rejected the proposal because the data from the big data search were not qualified data.	Purchase Manager
Social sorting and discrimination	The team discussed each attribute for the control model to make sure no social sorting and discrimination would occur.	Design Team
Attribute selection biases	The data analyst tested all attributes and assured that the selected characteristics were not highly correlated to gender, department code, vendor number, or similar.	Data Analyst
Historical biases	The team excluded the transaction data from December 2015 through March 2017 even though the team had the data in this period, and the team treasured more data to improve the performance of the model. The reason is that the university was applying a new ERP system at that time, and the data have quality issues at the initial stage.	Design Team
Beneficence and sustainability of algorithm	The purchase manager stated, "The audit analytics helps me improve efficiency to audit Quick Order (the transactions can be exempted from the high-level managers’ approval if the purchase amount is lower than $5000, and it is in the Quick	Purchase Manager

	Order list). Now we spend only 30% time to do the same work before the audit analytics application."	
Justice of Algorithm	The design team provided a solution for this issue by reviewing all of the feedback from the responses of the business rule violators. Purchase managers claimed, "Most Quick Order violators admit the violation and say sorry about the unintentional mistake."	Purchase Manager
Transparency of Algorithm	The audit manager clarified, "To avoid some anti-anti-fraudulent action, we need to keep the secrecy of some fraudulent analytical algorithm in a limited group."	Audit Manager
	The design team undertook a half-year exploration and tested about the choice of the algorithm. The final version of the algorithms of these queries can be expressed as the "if…then" business rule. The chosen algorithms can assure traceability and accountability; each end-user can trace the analytical result into a specific violation.	Design Team
Accessibility (user-friendliness)	To help the end-users have a good understanding of the analytical result, the audit department transferred and visualized the outputs using Tableau. The visualization can help produce direct knowledge about the analytical result.	Audit Department
Availability of Assessment Feedback	The audit manager explained, "We send emails to the related business rule violators and require them to confirm the violation. We require them to give feedback or explanation if the result has false positive. Then the feedback was input into the database to tune the analytical model periodically."	Audit Manager
Data Privacy Protection	The design team took two measures to assure data privacy in the audit analytics: (1) hide all of the sensitive data, like SSN, Date Birth, and the like; (2) limit the analytic result only to relevant end-users and business violators.	Design Team

4.7. Summary

This essay intends to increase awareness of ethical issues in audit analytics, which is based on algorithmic function. The study ascribes these problems to the four components of audit analytics, including data source, training data, algorithm mechanism, and output interpretation. Then we derive eleven ethical dimensions from the four aspects of audit analytics. Examining ethical considerations of the algorithmic audit analytics reveals essential changes to management's understanding of how the ADA impacts audits. Also, we provide a theory-informed systematic four-stage process assessment based on the eleven ethical dimensions. This evaluation roadmap provides a reliable means for enhancing self-governance and the quality of managerial ethics, so governance is not arbitrary.

Chapter 5. Conclusion

Continuous Audit Data Analytics is far from being a simple technological tool. The successful institutionalization of a CADA project requires integrating the technology and the institutional political hierarchy and culture. Furthermore, the internal control system is a complex system that covers almost every corner of an organization and requires all management levels. The adoption of CADA project in ICs is very challenging because of the complicated domain constraints. This dissertation argues for the effectiveness of CADA, demonstrates how its institutionalization can effectively develop internal control intelligence, and demonstrates related analytics techniques as well.

Primarily, the study contributes to the literature by theorizing the CADA institutionalization to develop internal control intelligence. The ultimate goal of CADA is to find audit evidence or actionable business policy to set up actionable controls in the internal control system. It affords actionable knowledge discovery by using emerging technology. Moreover, audit evidence strictly requires accountability and reliability. This constraint brings pressures and inertia for organizations to institutionalize emerging technologies even though these technologies showed persuasive ability in other areas. This study developed a theoretical framework to harness the power of IT-supported internal controls. The adoption of emerging technology requires a co-construction process, and six propositions have been theorized to manage the negative impacts of CADA in internal controls.

Second, the study suggests practical analytical techniques to promote the application of a CADA project. Data issue has fundamental constraints upon CADA, and any robust analytics algorithm will be paralyzed without qualified data inputs. The dissertation proposed an agile shareable data platform to digitize the internal control system. The beauty of this data platform is that it deploys a progressive mindset for the development of the domain-based database. (1) The database requires actively searching for what kind of attributes are needed for the potential analytics solution. This feature enhances the extensibility of the data infrastructure, which has the flexibility to search ubiquitous data around the domain. The CobiT 5 acts as a technical frame to ensure that the domain-based scope can completely digitize the whole internal control system. (2) The agile database can be shareable for all analytics artifacts embedded in the CADA system, as well as all stakeholders of the CADA. The dissertation demonstrated a series of feature engineering techniques to transform raw data into analyzable data, directly acting as the controls in the internal control system. (3) The agile database virtually restructures the entire internal control process, which often crosses several departments. It provides institutional value to enable an in-depth investigation of potential risks. The agile database provides the potential for real-time monitoring and also can be used for experimentation to explore potential risky controls. We also exemplified the potential risk exploration by combining an MCA approach and an experimental comparison. This mechanism is matched to the "ongoing process" metaphysics of SCOT and provides intelligence for the CADA.

The complexity of the internal control system decides that the CADA requires integrating the "social" and the "technical" side to maximize its performance. Thus analytical traceability and interpretability are significantly emphasized. The modular-based analytics schema achieves this integration and enhances its flexibility for user-participation. The schema has no complicated algorithm or data model, but it can afford an obvious evidentiary deduction that links the operational features to the effectiveness score by the bridge of the control rule. The mechanism is straightforward, and we can trace any problems to violators.

The study also exemplifies that the integration of visualization and data analytics can improve the acceptance of the CADA. Unlike data analytics in other areas, CADA needs to ease the deployment constraints to communicate the results with all related stakeholders. Auditors need to enhance the persuasiveness of the analytics result. Visualization improves communications among different stakeholders by combining the "social" and "technical" sides. With the robust domain-oriented database, it is easy to integrate visualization tools with the analytics schema. The institutional value of this integration can help handle inertia and improve acceptance of the CADA project.

Finally, the algorithmic solution has potential ethical issues. This moral problem is a significant hindrance for stakeholders wanting to embrace the dominant technologies. Users may worry about the ethics uncertainty from the emerging technologies because the machine has no ethical constraints, and the power can be abused. We discussed and developed basic approaches to deal with these issues. This work theorized eleven potential ethical dimensions for audit analytics and then constructed a four-phase ethics assessment roadmap to guide how to evaluate and deal with these moral problems.

This dissertation initiates the research pipeline to obtain insights into the use of emerging data and technologies in the internal control assessment. This pipeline can be extended to the future in two directions. One is about how to adopt CADA artifacts to a variety of institutional ecosystems, and the other is to develop a more effective analytics algorithm to balance the domain constraints and computational strength. It is important that both the development of data infrastructure capability and ethics issues are ongoing dynamics and need new exploration and development.

BIBLIOGRAPHY:

Abbott, A. (1992). From causes to events: Notes on Narrative Positivism. *Sociology Methods Research,* 20(4): 428-455.

AccountAbility, (1999). What is accountability, available at: https://www.accountability.org/insights#insights (accessed on Sep. 2018).

ACM. (2017). Statement on algorithmic transparency and accountability, Washington, DC: ACM US Public Policy Council. ACM. (2018).

ACM Code of Ethics and Professional Conduct. Available at: https://www.acm.org/binaries/content/assets/about/acm-code-of-ethics-booklet.pdf(Retrieved on April 15 2020).

AICPA, (2017). Guide to Audit Data Analytics. *American Institute of CPAs*.

AICPA, (2020). The Data-Driven Audit: How Automation and AI are changing the Audit and the Role of the Auditor.

Aldrich, H.E. & Pfeffer, J. (1976). Environments of Organizations. *Annual Review of Sociology,* Vol. 2, pp.79-105.

Allen, C., Varner, G., & Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence*, 12(3), 251–261.

Alles, M. A., Brennan, G., Kogan, A., & Vasarhelyi, M.A. (2006). Continuous Monitoring of Business Process Controls: A Pilot Implementation of a Continuous Auditing System at Siemens. *International Journal of Accounting Information Systems*, 7 (2), 137-161.

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). A four-process architecture for assessment delivery, with connections to assessment design. *Journal of Technology, Learning, and Assessment*, Vol. 2 (3), 1-5.

Amin, A. & Thrift, N. (2002). Cities: Re-imagining the Urban. *Polity Press,* Cambridge.

Anderson, M., & Anderson, S.L. (2014). Toward ethical, intelligent autonomous healthcare agents: A case-supported principle-based behavior paradigm. Available at: http://doc.gold.ac.uk/aisb50/AISB50-S17/AISB50-S17-Anderson-Paper.pdf (accessed on 24 August 2015).

Ashbaugh, H., Collins, D., Kinney, W., & Lafond, R. (2009). The effect of SOX internal control deficiencies on firm risk and cost of equity. *Journal of Accounting Research*, 47, 1-43.

Ashton, R.H. & Brow, P.R. (1980). Descriptive modeling of auditors' internal control judgments: replication and extension. *Journal of Accounting Research,* Vol.18 (1), pp.

269-277.

AuditNet, (2012). Big Data and its Usage in Auditing, Available at: https://www.auditnet.org/recordings (Accessed on September 2018).

Bambauer, D. E. (2017). Uncrunched: Algorithms, Decision making, and Privacy. Second Annual Digital Information Policy Scholars Conference, George Mason University Antonin Scalia Law School, Arlington, VA. (Apr. 28, 2017).

Barad, K. (2003). Posthumanist performativity: Toward an understanding of how matter comes to matter. *Journal of Women in Culture and Society,* 28(3), 801-831

Barad, K. (2007). Meeting the Universe Halfway: quantum Physics and the Entanglement of Matter and Meaning. Durham, NC: Duke University Press.

Batini, C., Cappiello, C., Francalanci, C., & Maurinoacm, A. (2009). Methodologies for Data Quality Assessment and Improvement. *Computing Surveys*, Vol. 41, No. 3, Article 16.

Berg, M. (1998). The Politics of Technology: On Bringing Social Theory into Technological Design. *Science, Technology, & Human Values*, Vol. 23 (4), 456-490.

Beauchamp, T. L. (1982). Philosophical Ethics, McGraw-Hill, New York.

Beauchamp & Childress (1994). Principles of Biomedical Ethics, Fourth Edition, Oxford.

Bijker, W. (1995). Sociohistorical technology studies, in: Jasanoff et al. (Editors), Handbook of Science and Technology Studies (Sage Publications, London) pp. 229-256.

Bijker, W. & Law, J. (1992). Shaping Technology/Building Society: Studies in Socio-technical Change (MIT Press, Cambridge, MA).

Bijker, W., Hughes, T. & T. Pinch (1987). The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology (MIT Press, Cambridge, MA).

Bloomfield, R.J. (1997). Strategic dependence and the assessment of fraud risk: a laboratory study. *Accounting Review*, Vol. 72 (4), 517-538

Boyd, B. (1990). Corporate Linkages and Organizational Environment: A Test of the Resource Dependence Model. *Strategic Management Journal,* Vol.11 (6), 419-430.

Boyer, J., & Mili, H. (2011). Agile Business Rule Development: Process, Architecture, and JRules Examples. Heidelberg: Springer, 2011.

Brey, P., & Soraker, J.H. (2009). Philosophy of Computing and Information Technology, Elsevier.

Brey, P. A. (2012). Anticipating ethical issues in emerging IT. *Ethics and Information*

Technology, 14(4), 305–317.

Bumgarner, N., & Vasarhelyi, M.A. (2015). Continuous Audit—a new view, Audit Analytics and Continuous Audit: Looking Toward the Future, Copyright © 2015 by American Institute of Certified Public Accountants, Inc. New York, NY 10036-8775, ISBN: 978-1-94354-608-4

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society,* Vol. 3(1).

Callon, M. & Muniesa, F. (2005). Peripheral Vision: Economic Markets as Calculative Collective Devices. *Organization Studies*, Vol. 26(8), 1229–1250

Candès, E.J., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory,* 52(2), 489–509.

Cao, L. (2007). Agent & Data Mining Interaction. *Tutorial for 2007 IEEE/WIC/ACM Joint Conferences on Web Intelligence and Intelligent Agent Technology*, 2007.

Cao, L. (2009). Data Mining and Multi agent Integration, *Springer*.

Cao, L. (2007). Domain-Driven Actionable Knowledge Discovery. *IEEE Intelligent Systems*, Vol. 22(4), 78-89.

Cao, L.; Zhang, C.; Yang, Q.; Bell, D.; Vlachos, M.; Taneri, B.; Keogh, E.; Yu, P.; Zhong, N.; et al. (2007). Domain-Driven, Actionable Knowledge Discovery. *IEEE Intelligent Systems,* 22 (4), 78–89.

Cao, L. (2009). Domain-Driven Data Mining: Empowering Actionable Knowledge Delivery. *PAKDD 2009 Tutorial*.

Cao, L. and Dai, R. (2003). Agent-oriented meta synthetic engineering for decision making. *International Journal of Information Technology and Decision Making,* Vol. 2(2), 197-215.

Cao, L. & Dai, R. (2008). Open Complex Intelligent Systems. *Posts & Telecom Press*.

Cao, L., Luo, C. & Zhang, C. (2007). Developing actionable trading strategies for trading agents. *IAT 2007*, 72-75.

Cao, L., Yu, P. S., Zhan, C. & Zhao, Y. (2010). Domain Driven Data Mining. *Springer*.

Carroll, G. R. & Hannan, M. (2000). Why Corporate Demography Matters: Policy Implications of Organizational Diversity. *California Management Review,* Vol. 42(3), 148-163.

Cecez-Kecmanovic, D., Galliers, R. D., Henfridsson, O., Newell, S., & Vidgen, R. (2014).

The Socio-materiality of Information System: Current Status, Future Directions. *MIS Quarterly* Vol. 38 No. 3, pp. 895-925, (September), *Special Issue: Socio-materiality of IS & Organization*.

Cheng, Z. & Chen, D. (2002). Integrated strategy of pattern classification and its application. *Journal of Zhejiang University (Engineering Science),* 36(6), 601–606.

Chuprunou, M. (2013). Auditing and GRC Automation in SAP, *Springer-verlag*, Berlin, Heidelberg.

CICA/AICPA. (1999). Continuous auditing, Research report, Toronto, Canada. The Canadian Institute of Chartered Accountants.

CobiT: Control Objectives for Information and Related Technologies. www.isaca.org/cobit.

Collins, M., Dasgupta, S. & Schapire, R. E. (2001). A generalization of principal components analysis to the exponential family, in Dietterich, T., Becker, S., & Ghahramani, Z. (Eds.), *Advances in Neural Information Processing Systems,* MIT Press, Vol.14, 617–624.

Cohen, J., Krishnamoorthy, G., & Wright, A. (2017). Enterprise Risk Management and the Financial Reporting Process: The Experiences of Audit Committee Members, CFOs, and External Auditors. *Contemporary Accounting Research*, Vol.34 (2), 1178-1209.

Cooper, M.J., Gulen, H., & Schill, M.J. (2008). Asset Growth and the Cross-Section of Stock Returns. *The Journal of Finance*, VOL. LXIII, No. 4, August 2008.

Copeland, R. M. (1968). Income Smoothing. *Journal of Accounting Research*, Vol. 6, 101-116

Correia, M., Kang, J. & Richardson, S. (2018). Asset volatility. *Review of Account Studies*, Vol. 23, 37–94

COSO, (1993; 2013). Enterprise Risk Management--Integrated Framework: Application Techniques, Committee of sponsoring organizations of the Treadway commission.

Covaleski, M.A., Dirsmith, M.W. (1988). An Institutional Perspective on the Rise, Social Transformation, and Fall of a University Budget Category, Administrative Science Quarterly, Vol. 33, No. 4, pp. 562-587.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., & Suri, S. (2008). Feedback effects between similarity and social influence in online communities, In Proceedings of the 14th *ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 160–168, New York, NY, USA, ACM.

Crawford, K. (2016). Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics. *Science, Technology, & Human Values,* Vol. 41(1), 77-92

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, Vol. 5(7), 809–828.

Dilla, W., Janvrin, D.J. & Raschke, R. (2010). Interactive data visualization: New directions for accounting information systems research. *Journal of Information Systems,* Vol.24 (2), 1-37.

Dwoskin, E. (2015). How social bias creeps into web technology. *Wall Street Journal.* http://www.wsj.com/articles/computers-are-showing-their-biases-and-tech-firms-are-concerned-1440102894 ?mod=rss_Technology (Accessed 12 Aug 2016).

Emirbayer, M. (1997). Manifesto for a Relational Sociology, *American Journal of Sociology,* Vol.103, 281–317.

EY, (2014). Big risks require big data thinking- Global Forensic Data Analytics Survey 2014, Available at: http://www.ey.com/Publication/vwLUAssets/EY-Global-Forensic-Data-Analytics-Survey-2014/$FILE/EY-Global-Forensic-Data-Analytics-Survey-2014.pdf. (Accessed on July 2019)

Fekete, J.D., Wijk, J.V., Stasko, J., & North, C. (2008). The Value of Information Visualization, Springer. Information Visualization: Human-Centered Issues and Perspectives, Springer, pp.1-18, Lecture Notes in Computer Science, 978-3-540-70955-8.

Feldman, M.S. & Pentland, B.T. (2003). Reconceptualizing Organizational Routines as a Source of Flexibility and Change. *Administrative Science Quarterly,* Vol. 48 (1), pp. 94-118

Fisher, F. M. (1971). Aggregate Production Functions and the Explanation of Wages: A Simulation Experiment. *Review of Economics and Statistics,* Vol. 53 (4), 305-26.

Floridi, L. (1999). Information ethics: on the theoretical foundations of computer ethics. *Ethics and Information Technology*, Vol.1, 37–56.

Floridi, L. (2008). Information Ethics: Its Nature and Scope, in *Information Technology and Moral Philosophy* (Edited by Jeroen Van Den Hoven and John Weckert). Cambridge University Press.

Fountain, J.E. (2001). Building the Virtual State—Information Technology and Institutional Change. Brookings Institution Press.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transaction On Information System*,Vol.14 (3), 330-347.

Friedman, B., Smith, I. E., Kahn, P. H., Jr. Consolvo, S., & Selawski, J. (2006). Development of a privacy addendum for open source licenses: Value Sensitive Design in industry. In Proceedings of UbiComp 2006 (Orange County, CA, USA, Sept. 17-21, 2006), Springer-Verlag, (Berlin, 2006), 194-211.

Fujimura, J. (1996). Crafting science: a socio-history of the quest for the genetics of cancer. Cambridge, MA: Harvard University Press.

Gergen, K. (2010). Mobile communication and the new insularity. Qwerty, 5(1), 14–28.

Gillespie, T. (2016). Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem. *Information, Communication & Society*, Vol. 20, 1–18.

Gitelman, L. (2013). "Raw Data" Is an Oxymoron. *The MIT Press Cambridge*, Massachusetts London, England.

Greenacre, M. (2017). Correspondence Analysis in Practice. Third Edition, Taylor & Francis Group, LLC.

Greenleaf, G. (2014). Global Data Privacy Laws: 89 Countries, and Accelerating. Social Science Electronic Publishing, Inc. SSRN 2000034. Retrieved 16 February 2014.

Grimlund, R. A. (1982). An Integration of Internal Control System and Account Balance Evidence. *Journal of Accounting Research*, (Autumn), 316-42.

Grigoria, D., Casatib, F., Castellanosb, M., Dayalb, U., Sayalb, M., & Shanb, M.C. (2004). Business Process Intelligence. *Computers in Industry,* Vol. 53, 321–343.

GSMA, (2018). The Data Value Chain, GSM Association.

Halaweh, M., Fidler, C., McRobb, S. (2008). Integrating the Grounded Theory Method and Case Study Research Methodology Within IS Research: A Possible 'Road Map.' ICIS 2008 Proceedings, International Conference on Information Systems (ICIS)

Hajian, S., & Domingo-Ferrer, J. (2013). A Methodology for Direct and Indirect Discrimination Prevention in Data Mining. *IEEE Transactions on Knowledge and Data Engineering,* VOL. 25, NO. 7.

Haraway, D. S. (1991). Cyborgs and Women: The Reinvention of Nature. New York: Routledge.

Hill, R.K. (2015). What an algorithm is. *Philosophy & Technology*, 29(1), 35–59.

Hirschman, A. O. (1970). Exit, Voice and Loyalty. *Harvard University Press*, Cambridge, USA.

Husson, F., Lê, S., & Pagès, J. (2011). Exploratory Multivariate Analysis by Example Using R, Taylor and Francis Group, Boca Raton London New York.

ISACA: the Information Systems Audit and Control Association, www.isaca.org)

Issa, H., & Kogan, A. (2014) A Predictive Ordered Logistic Regression Model as a Tool for Quality Review of Control Risk Assessments. *Journal of Information System*, Vol. 28 (2), 209-229.

ITGI: IT Governance Institute, Available at: www.itgi.org.

Jackowski, Z. (2003). Business modeling with UML. A business process centered architecture. Available at: http://www.agilealliance.org/articles/articles/BPM.pdf (Accessed July 23, 2014).

Jans, M., Alles, M., &Vasarhelyi, M. (2013). The case for process mining in auditing: Sources of value added and areas of application. *International Journal of Accounting Information Systems,* Vol. 14, 1–20

Jans, M., Lybaert, N. & Vanhoof, K. (2009). A framework for internal fraud risk reduction at IT integrating business processes: The IFR framework. *The International Journal of Digital Accounting Research,* Vol. 9, 1–29.

Jans, M., Alles, M., Vasarhelyi, M. (2014). A Field Study on the Use of Process Mining of Event Logs as an Analytical Procedure in Auditing. *The Accounting Review*, Vol. 89 (5), pp. 1751–1773.

James H. Moor. (2008). Why I Need Better Ethics for Emerging Technologies, in *Information Technology and Moral Philosophy* (Edited by Jeroen Van Den Hoven and John Weckert). Cambridge University Press.

Jensen, M. C. & Murphy, K. J. (1990). Performance pay and top-management incentives. *Journal of Political Economy*, Vol. 98, 225–264.

Johnson, J.A. (2013). Ethics of data mining and predictive analytics in higher education, SSRN Scholarly Paper, Rochester, NY: Social Science Research Network. Available at: http://papers.ssrn.com/abstract=2156058 (accessed 22 July 2015).

Johnson, D.G., & Powers, T.M. (2008). Computers as Surrogate Agents, in *Information Technology and Moral Philosophy* (Edited by Jeroen Van Den Hoven and John Weckert), Cambridge University Press.

Jolliffe, I. T., (1986). Principal component analysis, Springer, Berlin.

Kaler, J. (1993). Responsibility, accountability and governance. *The Political Quarterly*, vol. 64 (3).

Kallinikos, J. & Mariátegui, J. (2011). Video as Digital Object: Production and Distribution of Video Content in the Internet Media Ecosystem. *The Information Society,* Vol.27 (5).

Kappel, G., Rausch-Schott, S. & Retschitzegger, W. (1998). Coordination in workflow management systems–A rule-based approach. In Coordination Technology for Collaborative Applications - Organizations, Processes, and Agents, (Eds, Conen, W. & Neumann, G.) Springer, pp. 99-119.

Karagiannis, D., Mylopoulos, J., & Schwab, M. (2007) Business process-based regulation compliance: the case of the Sarbanes-Oxley Act, Proceedings of the Requirements Engineering Conference, 2007, *RE '07. 15th IEEE International*, October 2007, pp. 315-321.

Katz, M., Ellahie, A., & Richardson, S. (2014). Risky Value, 2014 AAA annual conference.

Keeney, R.L. (1984). Ethics, Decision Analysis, and Public Risk. *Risk Analysis, Vol.* 4. No.2.

Keeney, R.L. (1996). Value-Focused Thing--A Path to Creative Decision making. Harvard University Press Cambridge, Massachusetts London, England.

Keeney, R. L. & Winterfeldt, D.V. (2010). Identifying and Structuring the Objectives of Terrorists. *Risk Analysis,* Vol. 30, No. 12, 2010.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society,* Vol.14 (April–June), 1–12.

Kim, Y. & Kogan, A. (2014). Development of an Anomaly Detection Model for a Bank's Transitory Account System. *Journal Of Information Systems,* Vol. 28, No. 1 (Spring), 145–165.

Kim, Y. & Vasarhelyi, M.A. (2012). A Model to Detect Potentially Fraudulent/ Abnormal Wires of an Insurance Company: An Unsupervised Rule-Based Approach. *Journal of Emerging Technologies in Accounting*, Vol. 9, 95–110.

Kinney, W. R. (2000). Research Opportunities in Internal Control Quality and Quality Assurance. *AUDITING: A Journal of Practice & Theory: Supplement 2000*, Vol. 19, pp. 83-90.

Klein, H. K., Kleinman, D. L. (2002) The Social Construction of Technology: Structural Considerations. *Science, Technology, & Human Values*, Vol. 27 No. 1, Winter 28-52.

Knuth, K.H. (2006). Optimal data-based binning for histograms. arXiv preprint physics/0605197. This method has been implemented in Matlab and Python. Available online: http://knuthlab.rit.albany.edu/index.php/Products/Code & http://www.astroml.org/ (accessed on 11 January 2015).

Kogan, A., Alles, M., Vasarhelyi, M., & Wu, J. (2014). Design and Evaluation of a Continuous Data Level Auditing System. *Auditing: A Journal of Practice & Theory,* Vol. 33 (4), 221–245

Kovacic, A. (2004). Business renovation: business rules (still) the missing link. *Business Process Management Journal,* Vol.10, pp.158.

KPMG (2015). Expectations of Risk Management Outpacing Capabilities –It's Time For Action, Available at:

https://assets.kpmg/content/dam/kpmg/pdf/2013/08/expectations-risk-management-survey-v3.pdf (accessed on January 2019)

Kujala, S., & Väänänen-Vainio-Mattila, K. (2009). Value of Information Systems and Products: Understanding the Users' Perspective and Values. *Journal of Information Technology Theory and Application* (JITTA), Vol. 9: 23-39.

Kunz, R.E., Siebert, J., MÜtterlein, J. (2016). Combining Value-focused Thinking and Balanced Scorecard to Improve Decision-making in Strategic Management. *Journal of Multi-criteria Criteria Decision Analysis,* Vol.23, 225–241.

Kraemer, F., van Overveld, K., & Peterson, M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology,* Vol. 13 (3), 251–260.

Krishnan, J. (2005). Audit Committee Quality and Internal Control: An Empirical Analysis. *The Accounting Review*, Vol. 80 (2), 649–675.

Krogstie, J. et al. (1991). Information Systems Development Using a Combination of Process and Rule-Based Approaches. In Third International Conference on Advanced Information Systems Engineering (CAiSE '91), (Eds, Andersen, R., Bubenko, J.A.j. & Solvberg, A.) Springer-Verlag, Trondheim, Norway, pp. 319-335.

Latham, R. & Sassen, S. (2007). Digital formations: IT and New Architectures in Global Realm. *Perspectives on Politics,* Vol. 5 (03)

Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford: Oxford University Press.

Lee, G. H. (2008). Rule-based and case-based reasoning approach for internal audit of bank. *Knowledge-Based Systems*, Vol. 21, 140–147.

Leonardi, P. M. (2003). The Emergence of Materiality within Formal Organizations, in Carlile, P. R., Nicolini, D., Langley, A. and Tsoukasmbridge, H. (eds.), How Matter Matters: Objects, Artifacts, and Materiality in Organization Studies, Oxford University Press, pp. 141-168.

Leonardi, PM. (2010). From Road to Lab to Math: The Co-evolution of Technological, Regulatory, and Organizational Innovations for Automotive Crash Testing. *Social Studies of Science*, Vol. 40(2), 243-274.

Leonardi, P. M. (2013). Theoretical foundations for the study of socio-materiality. *Information and Organization,* Vol. 23, 59–76.

Loshin, D. (2011). Evaluating the Business Impacts of Poor data quality. *Information Quality Journal*.

Luca, M., Kleinberg, J., & Mullainathan, S. (2016). Algorithms need managers, too. *Harvard Business Review,* Vol.94. No. 1, pp1-2.

Lupton, D. (2014). Self-tracking modes: reflexive self-monitoring and data practices. *Social Science Research Network*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2483549 or http://dx.doi.org/10.2139/ssrn.2483549

MacKenzie, D. (1996). Violent solutions: Revolutions, nationalism, and secret societies in Europe to 1918. Lanham, MD: University Press of America.

Marschinski, R., & Kantz, H. (2002). Analysing the Information Flow Between Financial Time Series: An Improved Estimator for Transfer Entropy. *European Physical Journal B*, Vol. 30 (2), 275–81.

Martin, K. (2018). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics,* Vol. 149 (2).

Massa, P. (2011). Social networks of Wikipedia, In Proceedings of the 22nd ACM conference on Hypertext and hypermedia, pages 221–230, New York, NY.

Mazmanian, M., Cohn, M., & Dourish, P. (2014). Dynamic Reconfiguration in Planetary Exploration: A Sociomaterial Ethnography. *MIS Quarte*rly, 38(3), 831-848.

McBrien, P. & Seltveit, A.H. (1995). Coupling Process Models and Business Rules. In: Sölvberg A., Krogstie J., Seltveit A.H. (eds) Information Systems Development for Decentralized Organizations. IFIP—The International Federation for Information Processing. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-34871-1_12

Mesle, C.R. (2008). Process-Relational Philosophy: An Introduction to Alfred North Whitehead. Templeton Press, West Conshohocken, pp. 49.

Miller, H.J. (2010). The data avalanche is here. Shouldn't I be digging? *Journal of Regional Science,* Vol.50 (1), 181–201.

Miranda, S. M., and Saunders, C. S. (2003). The Social Construction of Meaning: An Alternative Perspective on Information Sharing. *Information Systems Research*, Vol. 14, No. 1 (March), pp. 87-106.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society,* (July–December), 1–21.

Mock, T. J., Sun, L., Srivastava, R. P., Vasarhelyi, M. (2009). An evidential reasoning approach to Sarbanes-Oxley mandated internal control risk assessment. *International Journal of Accounting Information Systems,* Vol. 10, 65–78

Moor, J. H. (1985). What is computer ethics?, in T. W. Bynum (Ed.), *Computers and ethics,* Oxford, UK: Blackwell, pp. 263–275. Also published as the October 1985 special issue of Metaphilosophy.

Moor, J. H. (1996). Reason, relativity and responsibility in computer ethics', a keynote address at ETHICOMP96 in Madrid, Spain. Later published in *Computers and Society*, Vol. 28 (1).

Moor, J.H. (2008). Why We Need Better Ethics for Emerging Technologies, in *Information Technology and Moral Philosophy* (Edited by Jeroen Van Den Hoven and John Weckert), Cambridge University Press.

Morgan, T. (2002). Business Rules and Information Systems, Pearson Education, Indianapolis.

Munoko, I., Liburd-Brown, H. & Vasarhelyi, M.A. (forthcoming). The Ethical Implications of Using Artificial Intelligence in Auditing, Journal of Business Ethics, forthcoming.

Neuer, M. (2008). Context is King: A Practical Approach to Rule Mining. *Business Rules Journal,* Vol. 9 (5), 75-99.

Nigrini, M. J. & Mittermaier, L. J. (1997). The use of Benford's Law as an aid in analytical procedures. *Auditing: a Journal of Theory and Practice,* Vol. 16 (2), 52-67.

Orlikowski, W.J. & Scott, S.V. (2008). Socio-materiality: Challenging the Separation of Technology, Work, and Organization. *The Academy of management Annuals,* January 2008,

Orlikowski, W.J. (2010). The socio-materiality of organisational life: considering technology in management research, Cambridge Journal of Economics, Vol. 34, 125–141.

Pasquale, F. (2015). The black Box society: The secret algorithms that control money and information. Cambridge, MA: Harvard University Press.

PCAOB, (2004). IT Control Objectives for Sarbanes-Oxley: The Role of IT in the Design and Implementation of Internal Control Over Financial Reporting, 2nd Edition, ISBN 1-933284-76-5.

PCAOB, (2004). An Audit of Internal Control over Financial Reporting Performed in Conjunction with an Audit of Financial Statements, Auditing Standard No. 2. Washington, D.C.: PCA.

PCAOB, (2017). AS 2201: An Audit of Internal Control Over Financial Reporting That Is Integrated with An Audit of Financial Statements, PCAOB, November 2017.

PCAOB's Strategic Plan 2018-2022 (draft for comments), Public Company Accounting Oversight Board, August 2018.

PCAOB, (2019). Data and technology task force, accessed on December 2019: https://pcaobus.org/Standards/research-standard-setting-projects/Pages/Data-Technology-Task-Force.aspx

Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, 2 (11), 559–572.

Pentland, B.T., & Feldman, M. S. (2007). Narrative Networks: Patterns of Technology and Organization. *Organization Science,* Vol. 18, No. 5, (Sep. - Oct.), pp. 781-795

Pentland, B. T., Hærem, T. & Hillison, D. (2011). The (N)Ever-Changing World: Stability and Change in Organizational Routines. *Organization Science*, Vol. 22(6), pp. 1369-1383.

Pickering, A. (1995). The mangle of practice: Time, agency and science. Chicago, IL: University of Chicago Press.

Pinch, T.J. & Bijker, W.E. (1984). The social construction of facts and artifacts: or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science,* Vol. 14, 399-441

Pinch, T. (1996). The social construction of technology: A review. In R. Fox (Ed.), Technological change: Methods and themes in the history of technology (pp. 17–36). Amsterdam: Harwood.

Pinch, T., & Trocco, F. (2002). Analog days: The invention and impact of the moog synthesizer. Cambridge, MA: Harvard University Press.

Pinch, T. (2008). Technology and institutions: living in a material world. *Theory and Society,* Vol. 37, pp. 461–483.

Pipino, L. L., Lee, Y. W., & Richard, Y. W. (2002). Data Quality Assessment. *Communication of the ACM*, Vol. 45 (4).

Pfeffer, J. (1982). Organizations and Organization Theory. Marshfield, MA, Pitman.

Pfeffer, J. and G. R. Salancik (1978). The External Control of Organizations: A Resource Dependence Perspective, New York, NY, Harper and Row.

Prichard, D., & Theiler, J. (1995). Generalized redundancies for time series analysis. *Physica D: Nonlinear Phenomenon,* Vol. 84: 476–493.

PWC, (2019). Re-inventing internal controls in the digital age, extracted on 02/20/2020:https://www.pwc.com/sg/en/publications/reinventing-internal-controls-in-the-digital-age.html.

Rajendra, P. S. (1986). Auditing Functions for Internal Control Systems with Interdependent Documents and Channels. *Journal of Accounting Research*, Vol. 24, No. 2 (Autumn), 422-426

Renteln, P. (2002). Review of L. Smolin, Three roads to quantum gravity, *American Scientist*, Vol. 90 (1).

Rezaee, Z., Sharbatoghlie, A., Elam, R. & McMickle, P. L. (2002). Continuous auditing: Building automated auditing capability. *Auditing: A Journal of Practice & Theory,* Vol. 21(1), 147-163.

Ribes, D., & Jackson, S.J. (2013). Data bite man: The work of sustaining long-term study. In: Gitelman L (ed.) 'Raw Data' is an Oxymoron. Cambridge, MA: MIT Press, pp. 147–166.

Richardson, V. J., & Waegelein, J. F. (2002). The influence of long-term performance plans on earnings management and firm performance. *Review of Quantitative Finance and Accounting*, Vol.18, 161–183.

Rikhardsson, P., Best, P., Green, P., & Rosemann, M. (2006). Business Process Risk Management and Internal Control: A proposed Research Agenda in the context of Compliance and ERP Systems. In: Second Asia/Pacific Research Symposium on Accounting Information Systems, Melbourne.

Roberts, L. (1994). Process Reengineering: A Key to Achieving Breakthrough Success. Milwaukee, WI: ASQC Quality Press.

Robichaud G., Garrard KP., Barry JA. & Muddiman DC. (2013). MSiReader: an open-source interface to view and analyze high resolving power MS imaging files on Matlab platform. *Journal of the American Society for Mass Spectrometry,* Vol. 24, 718–721.

Roetzel, P. G. (2019). Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. *Business Research,* Vol. 12 (2), 479–522.

Roux, B.L., & Rouanet, H. (1998). Interpreting Axes in Multiple Correspondence Analysis: Methods of the Contributions of Points and Deviations, Chapter 16 in Vizualisation of Categorical Data, Blasius & greenacre (Ed), Academic Press, 1998, pp. 197-223.

Ross, R. G. (2003). Principles of the Business Rule Approach. Addison Wesley.

Scott, S. V. & Orlikowski, W. J. (2012). Reconfiguring relations of accountability: Materialization of social media in the travel sector. *Accounting, Organizations and Society,* Vol. 37 (1), 26-40.

Seibt, J. (2018). Process Philosophy, *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/process-philosophy/>.

Seo, MG. & Creed, W. E. D. (2002). Institutional Contradictions, Praxis, and Institutional Change: A Dialectical Perspective. *Academy of Management Review,* Vol. 27, No. 2.

Slife, B.D. (2005). Taking practice seriously: Toward a relational ontology. *Journal of Theoretical and Philosophical Psychology*, Vol. 24(2), 157–178.

Smieliauskas, W.J., & Robertson, J.C. (2004). Auditing: An International Approach (Third Canadian Edition, 2004), McGraw-Hill Ryerson: Toronto.

Smolin, L. (2001). Three roads to quantum gravity, New York, NY: Basic Books.

Soucek, R. & Moser, K. (2010). Coping with Information Overload in Email Communications: Evaluation of A Training Intervention. *Computers in Human Behavior*, 26 (6), 1458–1466.

Srivastava, R. P. (1986). Auditing Functions for Internal Control Systems with Interdependent Documents and Channels. *Journal of Accounting Research,* Vol. 24 (2): 422-426.

Stahl, B. C., Heersmink, R., Goujon, P., Flick, C., van den Hoven, J., Wakunuma, K., et al. (2010). Identifying the ethics of emerging information and communication technologies: An essay on issues, concepts and method. *International Journal of Technoethics*, 1(4), 20–38.

Steinke, G. & Nikolette, C. (2003). Business rules as the basis of an organization's information systems. *Industrial Management and Data Systems*, Vol. 103(1/2): 52.

Stinchcombe, A.L. (1965). Social structure and organization, in Handbook of Organization, March, J.G. (ed.), pp.142-193, Chicago: Kand McNally.

Tarantino, A. (2008). Governance, Risk, and Compliance Handbook. Wiley, New Jersey.

Taylor, J. (2012). Decision Management Systems: A Practical Guide to Using Business Rules and Predictive Analytics, Pearson Education, Inc. Boston, MA 02116, ISBN-13: 978-0-13-288438-9.

TeamMate, (2012). TeamMate Audit Solutions Annual Report, TeamMate Audit Solutions, NV.

The Arthur W. Page Center (2020). Ethical Principles of Responsibility and Accountability. https://pagecentertraining.psu.edu/public-relations-ethics/ethics-in-crisis-management/lesson-1-prominent-ethical-issues-in-crisis-situations/ethical-principles-of-responsibility-and-accountability/. Accessed on April 2, 2020.

The Institute of Internal Auditor. (2013b). The three lines of defense in effective risk management and control. Accessed November 27, 2014, from https://na.theiia.org/standards-guidance/Public%20Documents/PP%20The%20Three%20Lines%20of%20Defense%20in%20Effective%20Risk%20Management%20and%20Control. pdf.

Theodore, J.M., Sun, L., Srivastava, R. P., & Vasarhelyi, M. (2009). An evidential reasoning approach to Sarbanes-Oxley mandated internal control risk assessment. *International Journal of Accounting Information Systems,* Vol.10, 65-78.

Thomke, S. H. (2020). Experimentation Works: The surprising power of business experiments. Boston, MA: Harvard Business Press.

Tschaskert, N., Kokina, J., Kozlowski, S. & Vasarhelyi, M. (2016). The next frontier in data analytics. *Journal of Accountancy,* August.

Van de Ven, A.H. (1992). Suggestions for Studying Strategy Process: A Research Note. *Strategic Management Journal,* Vol. 13 (summer special issue), 169-188.

Vasarhelyi, M., & Halper. (1991). The Continuous Audit of Online Systems, *Auditing: A Journal of Practice & Theory,* Vol.10.

Vasarhelyi, M. A., Alles, M., Kuenkaikaew, S., & Littley, J. (2012). The acceptance and adoption of continuous auditing by internal auditors: A micro analysis. *International Journal of Accounting Information Systems*, 13(3), 267-281.

Vasarhelyi, M. A., & Alles, M. G. (2008). The "now" economy and the traditional accounting reporting model: Opportunities and challenges for AIS research. *International Journal of Accounting Information Systems,* 9(4), 227-239.

Vasarhelyi, M. A., Alles, M., & Williams, K.T. (2011). Continuous Assurance for the Now Economy, The Institute of Chartered Accountants in Australia, ABN 50 084 642 571

Ver Steeg, G., & Galstyan, A. (2012). Information transfer in social media, In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16–20 April 2012; ACM: New York, NY, USA, pp. 509–518.

Vicente, R., Wibral, M., Lindner, M., & Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience,* Vol. 30 (1), 45–67.

Wagner, G. (2005). Rule Modeling and Markup. In Reasoning Web, (Eds, Eisinger, N. & Maluszynski, J.) *Springer*, Msida, Malta, pp. 251-274.

Wagner, E. L., Newell, S. & Piccoli, G. (2010). Understanding Project Survival in an ES Environment: A Sociomaterial Practice Perspective, *Journal of association for information systems,* Vol. 11 (5), 276-297.

Weckert, J. & Van den Hoven, J. (2008). Information Technology and Moral philosophy, Cambridge University Press, ISBN 9780521855495.

Wiener, N. (1948). Cybernetics: or Control and communication in the animal and the machine, Cambridge, MA: Technology Press.

Wiener, N. (1950–1954). The Human use of human beings: Cybernetics and society, New York: Houghton Mifflin, 1950. (2nd ed. rev., Doubleday Anchor, 1954.)

Wiener, N. (1964). God & Golem, Inc.: A comment on certain points where cybernetics impinges on religion, Cambridge, MA: MIT Press.

Wigmore, J. (1913). The problem of proof. *Illinois Law Review*, Vol. 8, 77

Whittington, R. (2006). Learning More from Failure: Practice and Process. *Organization Studies,* Vol. 27 (12), 1903-1906.

Williamson, D.M., Bejar, I.I., & Mislevy, R.J. (2006). Automated Scoring of Complex Tasks in Computer-Based Testing, Edited by Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey 07430.

Winner, L. (1980). Do artifacts have politics? *Daedelus*, Vol. 109, pp.121-36.

Wright, D. (2011). A framework for the ethical impact assessment of information technology, Ethics Information Technology, Vol.13:199–226Wigmore, J. (1913). The problem of proof. *Illinois Law Review*, 8, 77

Appendix 1: The Application of Combined Audit Data Analytics Schema to Implement SOX 404

This appendix demonstrates how the combined analytical schema identifies and accumulates evidence to implement SOX 404 from the evaluation perspective of external auditors. The PCAOB issued Auditing Standard No. 2201(the successor of AU. 5) to guide the ICs assessment. This guide requires to accomplish the objectives of the following audits simultaneously: "(1) to obtain sufficient evidence to support the auditor's opinion on internal control over financial reporting as of year-end; and (2) to obtain sufficient evidence to support the auditor's control risk assessments for purposes of the audit of financial statements" (PCAOB, 2017). Because the portable domain-based database can be flexible enough to extract necessary data from the audited institutions, the analytics schema discussed in section 2.3 can provide a robust data-driven solution for the implementation of SOX 404. We consider the managers' internal control statement as the current existing internal controls to build the rule-based-scoring system. The specific roadmap is illustrated as follows.

Stage 1: The audited institution claims its IC statement.

This statement can be brokedown by a series of processes and business rules according to the guidance of the COSO's IC framework and the ERM framework.

Stage 2: Auditors prepare a master file to request necessary data.

Auditors transform the company's IC statement into control rules and related data attributes by utilizing the formalization module, and submit a master file to the client company. The master file includes, but not limited to, the statement related data attributes (auditors can add the initiative 164 explicit control objectives with the guidance of CobiT framework) if the IC statement does not include these control objectives. In Appendix 2, readers can find an exemplified master file of the 164 control objectives of CobiT5.

Stage 3: Auditors develop a domain-based database by using meta-data-driven approach.

In this stage, data extraction needs to solve some potential technical issues to combine data from a variety of different sources. Each source could have a different data type, and also the specialized database may need to structure Big Data for the data requirements. This study proposes to use meta-data-based architecture and type-level specification to assure the data combination and data privacy protection when combining different data sources.

(1) Meta-data-driven approach

Technically, the first step of the meta-data-driven data combination is to set up the identifier of the process. Take the situation of the follow-up case study as an example for this data combination technique. The principle is that the identifier has to be unique. For example, in the case study, the purchase order number can be the identifier of the Purchase-to-Payment process. After setting up the process identifier, one of two kinds of situations will appear. The one condition is that the process identifier is also a primary key in all of the data sources (e.g., Purchase Order Number is a primary key in Oracle Data Warehouse and also an attribute in SciQuest Purchase Order System). In this situation, it is easy to extract the necessary data attributes and combine them with the common identifier. The other situation is more complicated than the one just described. The process identifier (like Purchase Order Number) exists only in one data source (like SciQuest purchase order system). Still it is not present in other data sources in which the database will use some attributes. Under this condition, it is necessary to develop a meta-data relation. It needs first to find a common attribute between the one data source that has a process identifier and other data sources. The common attribute has to be *unique* in both sources. It needs to define a composite attribute if it fails to find a qualified common attribute. And then, it is well prepared to extract the necessary attributes and combine them to construct a specialized database.

(2) Utilizing type-level specification to hide sensitive data

Hiding sensitive data from external auditors (or a third party) is important when the companies need to provide transactional data. For example, it is a business secrecy violation to disclose the purchase price of a product from a specified vendor. This study offers solutions by following Geerts and McCarthy (2002) to utilize type-level specification to hide identity data. It is necessary to typify the vendor name and address as a group of numbers: e.g., it is good to utilize vendor numbers for vendors, and use "the initial of the state, city, street + room number + zip code” as the address index.

Stage 4: The auditors exert substantial test with a rule-based scoring system.

The rule-based scoring system is an automated rule enforcement system, and it has three features: (1) the logic basis is Boolean function based on the “IF-THEN” query; (2) it exerts an exhaustive investigation and assign a violation score for each transaction; (3) it has traceability to find the relevant managers, staffs, and departments. In this stage, the design performs walkthroughs to test the business rules’ compliance. The probing questions, combined with the IF-THEN queries, help auditors to gain a sufficient understanding of the process and to be able to identify the effectiveness and efficiency of ICs at critical nodes.

Stage 5: Application of the exploratory analysis.

Auditors conduct MCA with the extracted database (auditors can select another suitable exploratory analytics model to explore potential risks based on the specific task if necessary). In this process, communication with the internal auditors is encouraged, and this communication can improve the understanding of exploratory analysis results.

Stage 6: Application of Confirmatory Analytics.

Auditors then confirm the analytics result from Stage 5 by using an experimental comparison. If the newfound rules show a significantly higher risk than the ICs in the statement from the management, these rules are guided to be included in the internal control system.

Stage 7: Rerun Stage 3 with the newfound control rules.

Stage 8: Prepare the internal control evaluation report.

The mechanism of this analytics schema is technical guidance for the implementation of AS.2201. It can perform walkthroughs and operate exhaustive tests on comprehensive transactional data and decrease the potential risk that can come from sample selection. This mechanism captures the management behavior for the whole population not only within the current internal control system but also considers potential business risks. The internal control system report based on this analytic schema can offer more circumstantial information on control position coverage, business rule compliance, and the reliability level of the information systems used. This information, from the perspective of an independent party, can add value to the audited company and help the company improve the IC quality. The output from the second through the fourth module (see Figure 2.1) can be utilized to generalize the report, e.g., the output from the second module can demonstrate whether the company missed documenting some activities. If the company cannot provide necessary data, like products received, this data missing is strong evidence to show the quality of the related internal control system.

Reference:

Geerts, G.L. & McCarthy, W.E. (2002). An ontological analysis of the economic primitives of the extended-REA enterprise information architecture, *International Journal of Accounting Information Systems*, Vol.3 (1): 1-16.

Appendix 2: CobiT 5 Transformation Table with OUA Data Model

1. Application Control Objectives for the Financial Statement Close Cycle (Process Identity: Financial Account Code)

Control Objectives	Financial Assertions	Risks	Feature-engineering Conceptual Model	Users (stakeholders)	Control Attributes	Segregation of Duties
1. Entries booked in the close process are complete and accurate.	Completeness, Existence	Avoid missing transactions that happened in the accounting period, and avoid adding extra transactions that are not in the period.	The beginning of accounting period =< Date (transactions) =< the end of accounting period	Journal entries recorder and approver.	Transaction date, Journal entry time, Journal adjustment time; Adjust amount; Reason; Adjust account (debit); Adjust account (credit).	Approvers cannot be the same person as the recorder.
2. Automated amortization timing, periods and methods are appropriate and accurately entered.	Valuation Existence	Avoid using changing amortization methods to manipulate earnings.	Assets amortization_{ij} = Assets $\text{amortization}_{ij+1}$ (i is for asset item, j is period time)	Amortization recorder and approver.	Amortization recorded time; Amount; Method change; Method change reason.	Approvers cannot be the same person as the recorder.
3. Variance reports are generated for use to identify posting errors/out-of-balance conditions.	Completeness, Existence, Valuation.	Posting errors impacts the number in the statement.	Credit balance_i = Debit balance_i	Balance adjustment staff and approver	Balance adjustment time, Result.	NA
4. Standard, recurring period-end journal entries submitted from subsidiary ledger systems are automated, appropriately approved and entered accurately.	Completeness, Existence, Valuation.	The numbers in period-end journal entries are not accurate.	Amount of period-end journal = sum of the subsidiary ledgers	Period-end journal entries recording staff and approver.	Name of period-end journal, Input time, Amount of period-end journal, Name of subsidiary ledger, Amount of subsidiary ledger.	NA

5. Systems generate reports of all recurring and nonrecurring journal entries.	Completeness, Existence	Some manual changes cause error for the journal entries.	System input time existed	System supervisor	System input time; Name of the journal entry.	NA
6. All nonstandard journal entries are tracked and are appropriate.	Completeness, Existence	The record of nonstandard journal entries has errors.	Debit of nonstandard journal = Credit of nonstandard journal	nonstandard journal recorders and approvers.	Name of nonstandard journal, Input time, Amount of nonstandard journal.	Approvers cannot be the same person as the recorder.
7. Account codes and transaction amounts are accurate and complete, with exceptions reported.	Completeness, Existence	The mismatch account code cause error for the number of financial statement.	Account $code_i$ = Account $code_j$ .AND. $Amount_i$ = $Amount_j$	System supervisor	Account Code, Amount.	NA
8. General ledger balances reconcile to sub-ledger balances.	Completeness, Existence	Some errors happened in the reconciliation process.	Amount of General $Ledger_i$ = Sum (amount of sub-$ledger_j$)	General Ledger Recorder and Approver	System input time, Name of the GL, Code of GL, Amount of GL, Name of Sub-ledger, Code of Sub-ledger, Amount of sub-ledger.	NA
9. Recorded amounts undergo an automated comparison to predicted amounts.	Completeness, Existence	Necessary reconciliations for related accounts (e.g. bank statement and cash account).	Amount of $Account_i$ = Amount of $Account_j$ (e.g. amount of cash accounts = bank statement)	Account staffs for the two related accounts	Account Name, Account Code, Amount, System input time.	Two related accounts staff cannot be the same person.
10. Out-of-balance entries are prohibited.	Completeness, Existence	Out-of-balanc e account causes the error number in the financial statement.	Credit $Balance_i$ = Debit $Balance_i$	Balance reconciliation staff and approver	Account Name, Account Code, Debit balance, Credit balance, System time.	Balance reconciliation staff and approver cannot be the same person.

11. Enterprise wide consolidation, including standard intercompany eliminations, is automated/performed using a third-party software product.	Completeness, Existence, Valuation	The elimination of intercompany activities for consolidated report are not accurate.	Sum(Intercompany entity$_i$) = Sum(Intercompany entity$_j$)	Recorders and approvers of intercompany elimination activities.	Intercompany entity name, Transaction amount, Transaction time, Activity description.	Recorders and approvers cannot be the same person.
12. System functionality supports the segregation of the posting and approval functions.	Existence	Potential collusion for system functionality.	System Poster ≠ System Approver	System poster and System approver	System poster name, System approver name.	System poster cannot be the same person as system approver.
13. Access to general ledger records is appropriate and authorized.	Completeness, Existence, Valuation.	The system may have potential data quality issues if unauthorized people can access the system.	People accessed the GL system must be in the list of the authorized master file.	The supervisor of the master file, The system monitor.	The name of whom accessed the system, Time, The list of the authorized people.	The supervisor of the master file cannot be the same person as the system monitor.
14. Transactions cannot be recorded outside of financial close cutoff requirements.	Completeness, Existence, Valuation.	Cutoff time is to ensure the right information is reported within a specific period.	Transaction time > Cutoff time	Information recorder, Approver.	Transaction time, Cutoff time.	NA
15. Annually approved recurring accruals are accurately booked in the appropriate periods.	Completeness, Existence, Valuation.	It is to change accruals calculation methods to manipulate earnings.	Accruals$_{ij}$ = Accruals$_{ij+1}$ (i is for the accrual type and j is for year)	Information poster, Approver.	Accrual Account, Recorded time; Amount; Method change; Method change reason.	Approvers cannot be the same person as the poster.

16. System controls are in place for appropriate approval of write-offs.	Existence	The system may have potential data quality issues if unauthorized people can access the system.	People accessed the system have to be in the list of the authorized master file.	The supervisor of the master file, The system monitor.	The name of whom accessed the system, Time, The list of the authorized people.	The supervisor of the master file cannot be the same person as the system monitor.
17. Interrelated balance sheets and income statement accounts undergo automated reconciliation.	Completeness, Existence	Necessary reconciliations for related accounts.	Amount of Balance Sheet $Account_i$ = Amount of Income Statement $Account_j$	Account staffs for the two related accounts	Account Name, Account Code, Amount, System time.	NA
18. The sources of all entries are readily identifiable.	Existence	Avoid fabricated entries.	Sum($Account_i$) = Sum(Source $Account_j$)	Information recorders and approvers of entries.	Account Name, Account Code, Amount, System time.	Recorders and approvers of entries cannot be same.
19. Transactions are either rejected, or accepted and identified, on exception reports in the event of data exceptions.	Completeness, Existence	Avoid unchecked data issues for the transaction level information.	Transaction test action = "rejected" .OR. "accepted" .OR. "identified"	Posters and approvers of the test.	Transaction Number, Account Name, Account Code, Time, Test Action, Reason Description.	NA
20. Account mappings are up to date.	Existence	Avoid no map code in subsidiary ledger for the update of GL.	Amount of General $Ledger_i$ = Sum (amount of $subledger_j$)	General Ledger Recorder and Approver	System input time, GL name, GL Code, GL amount, Sub-ledger name, Sub-ledger code, sub-ledger amount, Map code of sub-ledger.	NA

2. Application Control Objectives for the General Ledge Cycle (Process Identity: GL Code)

Control Objectives	Financial Assertions	Risks	Conceptual Model and Criteria	Users (stakeholders)	Control Attributes	Segregation of Duties

21. Access to general ledger entries is appropriate and authorized.	Completeness, Existence, Valuation.	The system may have potential data quality issues if unauthorized people can access the GL system.	People accessed the GL system must be in the list of the authorized master file.	The supervisor of master file, System monitor.	The name of whom accessed the system, Time, The list of the authorized people.	The supervisor of the master file cannot be the same as the system monitor.
22. GL balances reconcile to sub-ledger balances and such reconciliations are reviewed for accuracy and approved by supervisory personnel.	Completeness, Existence.	Some errors happened in the reconciliation process.	Amount of General Ledger_i = Sum (amount of sub-ledger_j)	General Ledger Reviewer and Approver	System input time, GL name, GL code, GL Amount, Sub-ledger name, Sub-ledger code, sub-ledger amount.	General Ledger Reviewer and Approver cannot ben the same person.
23. Interrelated balance sheets and income statement accounts undergo automated reconciliations to confirm accuracy of such accounts.	Completeness, Existence.	Necessary reconciliations for related accounts.	Amount of Balance Sheet Account_i = Amount of Income Statement Account_j	Account staffs for the two related accounts	Account Name, Account Code, Amount, System input time.	NA
24. Systems generate reports of all recurring and nonrecurring journal entries for review by management for accuracy.	Completeness, Existence.	Journal entries have errors because of no reviewers' double check.	Journal Entry Poster ≠ Journal Entry Approver	Journal Entry Poster and Approver	Journal account name, Input time, Review time.	Journal entry recorder cannot be the same person as Approver.
25. System functionality exists to segregate the posting and approval functions.	Existence	Same as No.12 Control Objective				
26. All nonstandard journal entries are tracked and are appropriate.	Completeness, Existence.	Same as No.6 Control Objective				
27. Account codes and transaction amounts are accurate and complete, with exceptions reported.	Completeness, Existence.	Same as No.7 Control Objective				
28. Recorded amounts undergo automated comparison to predicted amounts to confirm accuracy of entries.	Completeness, Existence.	Same as No.9 Control Objective				

29. Out-of-balance entries are prohibited.	Completeness, Existence.	Same as No.10 Control Objective				
30. Enterprise wide consolidation, including standard intercompany eliminations, is automated / performed.	Completeness, Existence, Valuation.	Same as No.11 Control Objective				
31. Variance reports are generated for use to identify posting errors/out-of-balance conditions.	Completeness, Existence, Valuation.	Same as No.3 Control Objective				
32. System controls are in place for appropriate approval of write-offs.	Existence	Same as No.16 Control Objective				
33. Journal entries of exceptional amount that were posted to the general ledger during the month are flagged by the system and subsequently reviewed for accuracy and approved by the controller or CFO after month-end.	Completeness, Existence, Valuation.	Exceptional amount of GL entries causes the error of the financial statement.	Amount of General Ledger$_i$ $\neq$ Sum (amount of sub-ledger$_j$)	Posters and approvers of entries.	Account Name, Account Code, Amount, System input time.	Posters and approvers of entries cannot be the same person.
34. A report of all journal entries completed as part of the closing process is reviewed by management to confirm the completeness and appropriateness of all recorded entries.	Completeness, Existence.	The accuracy and completeness of the GL entries impacts the quality of financial reports	Amount of General Ledger$_i$ = Sum (amount of sub-ledger$_j$) .AND. Credit Balance$_i$ = Debit Balance$_i$	Posters and approvers of the entries.	Account name, Account code, Amount, System time.	Posters and approvers of entries cannot be the same person.
35. General ledger master file change reports are generated by the system and reviewed as necessary by an individual who does not input the changes.	Completeness, Existence.	People accessed the GL system must be in the list of the authorized master file.	The poster and the approver of the authorized master file, The system monitor.	The name of whom accessed the system, Time, The list of the authorized people.	The supervisor of the master file cannot be the same person as the system monitor.	Posters and approvers of the GL master file cannot be same.

36. Actual-to-actual, actual-to-budget and yield reports are produced from the GL system on a monthly basis prior to the final close of the GL. Reports are distributed to and reviewed by the controller and CFO. Unusual amounts are investigated and reclassified when applicable.	Completeness, Existence, Valuation.	Avoid unauthorized purchase orders.	Transaction approval date > Invoice date	Transaction requestor; Purchase approver	Request date; Approval date; Requestor ID; Approver ID; Requestor department;	Requestor cannot be the same person as approver.
37. A standard chart of accounts has been approved by management and is utilized within all entities of the corporation. Adding to or deleting from the GL is limited to authorized accounting department personnel.	Completeness, Existence.	Avoid no standard map code in subsidiary ledger for the update of GL.	Amount of General $Ledger_i \neq$ Sum (amount of sub-$ledger_i$)	Posters and approvers of standard map code.	Account Name, Account Code, Amount, System input time.	Posters and approvers of map code cannot be the same person.
38. A stale items report is generated by the system to monitor timely follow-up and resolution of outstanding items.	Completeness, Existence.	Avoid delayed actions to solve the stale item issue.	Date() - Date(report) > 90	Report poster and the manager who is responsible for the report.	Account name, Account code, Date, Time, Accident description.	NA
39. Entries booked in the close process are complete and accurate.	Completeness, Existence.	Avoid missing transactions happened within the accounting period, and avoid adding extra transactions that are not in the period.	The beginning of accounting period =< Data (transactions) =< the end of accounting period	Journal entries poster and approver.	Transaction Date, Journal entry time, Journal adjust time; Adjust amount; Reason; Adjust account (debit); Adjust account (credit).	Approvers cannot be the same person as the record staff.
40. Automated amortization timing, periods and methods are appropriate and accurately entered.	Valuation, Existence.	Same as No.2 Control Objective				

41. Standard, recurring period-end journal entries submitted from subsidiary ledger systems are automated, appropriately approved and entered accurately.	Completeness, Existence, Valuation.	Same as No.4 Control Objective				
42. Transactions cannot be recorded outside of financial close cutoff requirements.	Completeness, Existence, Valuation.	Same as No.14 Control Objective				
43. Annually approved recurring accruals are accurately booked in the appropriate periods.	Completeness, Existence, Valuation.	Same as No.5 Control Objective				
44. The sources of all entries are readily identifiable.	Existence	Avoid fabricated entries.	Sum (Geneal Ledger$_i$) = Sum(Subsidiary Ledger$_j$)	Posters and approvers of entries.	Account Name, Account Code, Amount, System time.	Posters and approvers of entries cannot be the same person.
45. Transactions are rejected, or accepted and identified, on exception reports in the event of data exceptions.	Completeness Existence	Same as No.19 Control Objective				
46. Account mappings are up to date.	Existence	Same as No.20 Control Objective				
3. Application Control Objectives for the Sales Cycle (Process Identity: Sales Order Number)						
Control Objectives	Financial Assertions	Risks	Conceptual Model and Criteria	Users (stakeholders)	Control Attributes	Segregation of Duties
47. Orders are processed only within approved customer credit limits.	Valuation	The credit sales exceeds customer credit limits can cause potential sales loss.	Accounts receivables (customer$_i$) <= Credit limit (customer$_i$)	Sales requestor, Sales approver.	Invoice number, Time, Amount, Customer credit limit, Accounts receivables, Customer number.	Requestor cannot be the same person as approver.
48. Orders are approved by management as to prices and terms of sale.	Existence	Unauthorized sales terms cause potential sales loss.	Sales terms <= Sales term criterion .AND. Discount <= discount criterion.	Sales discount requestor, Sales discount approver.	Transaction time, Amount, Sales discount criterion, Sales term criterion, Unit	Requestor cannot be the same person as approver.

					price, Customer number, Invoice number.	
49. Orders and cancellations of orders are input accurately.	Valuation	The error of the orders and cancellation can cause potential loss of accounts receivables or duplicate accounts receivables.	Invoice date = Invoice date .AND. Invoice amount = invoice amount .AND. Customer number = Customer number .AND. Invoice number = Invoice number	The requestors of orders or cancellations, the approvers of orders or cancellations.	Transaction Number, Invoice number, Transaction time (cancellation time), Invoice amount, Customer number.	Requestor cannot be the same person as approver.
50. Order entry data are transferred completely and accurately to the shipping and invoicing activities.	Valuation, Completeness.	Avoid delayed actions to processing invoice and shipping activities.	Date (system inputed) < Invoice Date .AND. Invoice Date < Date (shipping)	Staffs are responsible for processing order activities.	Transaction number, Invoice number, Transaction system input time, Invoice date, Shipping date.	NA
51. All orders received from customers are input and processed.	Completeness	Avoid delayed actions to processing invoice activities.	Date (system input) < Invoice Date	Staffs are responsible for processing order activities.	Transaction number, Invoice number, Transaction system input time, Invoice date.	NA
52. Only valid orders are input and processed.	Existence	The error of the orders can cause potential loss.	Approved time < System input time .AND. Approver ≠ Poster	The system poster, the approvers of orders.	Transaction number, Invoice number, Transaction system input time, Approved date.	The system poster cannot be the same person as the approvers of orders.
53. Invoices are generated using authorized terms and prices.	Valuation	Same as No.48 Control Objective				

54. Invoices are accurately calculated and recorded.	Valuation	The error of invoices can cause potential loss.	Transaction amount = Invoice amount .AND. Accounts receivables = Invoice amount .AND. Invoice terms = Sales order terms.	Invoice poster, Invoice approvers.	Transaction amount, Invoice amount, Accounts receivables, Invoice terms, Sales order terms.	Invoice poster cannot be the same person as invoice approvers.
55. Credit notes and adjustments to accounts receivable are accurately calculated and recorded.	Valuation	Same as No.54 Control Objective				
56. All goods shipped are invoiced.	Completeness	The error of invoices can cause potential loss.	Transaction amount = Invoice amount .AND. Shipped sales amount = Invoice amount .AND. Invoice terms = Sales order terms.	Invoice poster, Invoice approvers.	Transaction amount, Invoice amount, Accounts receivables, Invoice terms, Sales order terms, Shipped amount, Sales shipped time.	Invoice poster cannot be the same person as invoice approvers.
57. Credit notes for all goods returned and adjustments to accounts receivable are issued in accordance with organization policy.	Existence	Same as No.48 Control Objective				
58. Invoices relate to valid shipments.	Existence	Same as No.56 Control Objective				
59. All credit notes relate to a return of goods or other valid adjustments.	Completeness	Products returns and related sales adjustment should be recorded timely and accurately.	Debit (Products returned) = Credit (Accounts receivable) .AND. (Date (Products returned) - Date (Products shipping)) < Credit return term days	Products return received staff, Products return approver, Customer number, Sales represent.	Products returned amount, Product unit price, Return reason, Products return received time, Products shipping time, Products received quality description.	Products return received staff cannot be the same person as products return approver.

60. All invoices issued are recorded.	Completeness	Missing invoices causes errors of financial report.	Use Gap detection analysis to find missing numbers in the sequence number of invoices.	Invoice preparer, Invoice requestor, Invoice approver, Invoice shipping staff.	Invoice sequence number; Invoice date, Invoice shipping date.	Invoice requestor cannot be the same person as the invoice approver
61. All credit notes issued are recorded.	Existence	The errors in the record of credit notes can cause errors of sales and accounts receivables.	Debit (credit notes) = Credit (Accounts receivable)	Credit notes issuer, Accounts receivables staff.	Credit notes amount, Credit notes issued date, Customer number, Credit notes reason description.	Credit notes issuer cannot be the same person as accounts receivables staff.
62. Invoices are recorded in the appropriate period.	Valuation	Avoid potential earnings manipulation to assure invoice to be recorded in the right period.	Month (Invoice Date) = Month (Accounting Period)	The poster and the reviewer of the invoice activities.	Transaction number, Invoice number, Invoice system input time, Invoice date, Shipping date.	Invoice poster cannot be the same person as invoice reviewer.
63. Credit notes issued are recorded in the appropriate period.	Valuation	Avoid potential earnings manipulation to assure credit notes to be recorded in the right period.	Month (Credit note) = Month (Accounting Period)	The poster and the approver of the credit notes activities.	Transaction number, Invoice number, Credit note system input time, Credit note issued date.	The poster and the approver of the credit notes activities cannot be the same person.
64. Cash receipts are recorded in the period in which they are received.	Valuation	Avoid potential earnings manipulation to assure cash to be recorded in the right period.	Month (cash receipts) = Month (Accounting Period)	Cashier and accounts receivables staff.	Transaction number, Invoice number, Invoice system input time, Invoice date, Shipped date.	Cashier and accounts receivables staff cannot be the same person.
65. Cash receipts data are entered for processing accurately.	Valuation	Assure the safe of cash.	Amount (cash) = Periodic incremental credit balance (Accounts receivables)	Cashier and accounts receivables staff.	Transaction number, Invoice number, Invoice system input time, Invoice date, Shipped date.	Cashier and accounts receivables staff cannot be the same person.

66. All cash receipts data are entered for processing.	Existence	Same as No.65 Control Objective				
67. Cash receipts data are valid and are entered for processing only once.	Completeness	Same as No.65 Control Objective				
68. Changes to the customer master file are accurate.	Valuation	The error in the customer master file provides potential fraud opportunity.	Customer Number (master file) = Customer Number (invoice)	Requestor and Reviewer of the customer master file.	Customer number, Customer name, Customer address, Contact name, Master file updated time.	Requestor and Reviewer of the customer master file cannot be the same person.
69. Changes to the customer master file are processed in a timely manner.	Completeness, Existence.	Same as No.68 Control Objective				
70. Customer master file data remain up to date.	Completeness, Existence.	Same as No.68 Control Objective				

4. Application Control Objectives for the Purchasing Cycle (Process Identity: Purchase Order Number)

Control Objectives	Financial Assertions	Risks	Conceptual Model and Criteria	Users (stakeholders)	Control Attributes	Segregation of Duties
71. Purchase orders are placed only for approved requisitions.	Existence	Unauthorized purchase orders can cause potential fraud.	Purchase Approval date > Invoice Date	Purchase requestor; Purchase approver	Request Date; Approval date; Requestor ID; Approver ID; Requestor department; Vendors number	Requestor cannot be the same person as approver.
72. Purchase orders are accurately entered.	Valuation	Assure the order amount is the same as the amount in the invoice.	Order Amount = Invoice Amount .AND. Order unit price = Invoice unit price	Order information Input staff	Order Amount; Invoice Amount; Order Information Input Date; Vendors Number	NA
73. All purchase orders issued are input and processed.	Completeness	Missing purchase orders causes errors in financial report .	Use Gap detection analysis to find missing numbers in the purchaser order	Purchase requestor; Purchase approver; Order information input staff	Purchase order number; purchase order date	Requestor cannot be the same person as the staff to input

			sequence number.			information.
74. Amounts posted to accounts payable represent goods or services received.	Existence	Non-existed accounts payable causes potential fraud.	Products received = order quantity = Product quantity in invoice	Products receiver	Product received date, Product received quantity, Invoice quantity, Product quality review.	Products receiver cannot be the same person as purchase approver.
75. Accounts payable amounts are accurately calculated and recorded.	Valuation	Incorrect accounts payable causes potential losses.	Accounts payable = invoice amount = purchase order amount	Accounts payable clerk	Accounts payable amount, Invoice amount, Purchase order amount, Vendor number.	Accounts payable clerk cannot be the same person as cashier.
76. All amounts for goods or services received are input and processed to accounts payable.	Completeness	Duplicate accounts payable or missed accounts payable causes potential losses.	Invoice date=Invoice date .AND. Invoice amount=invoice amount .AND. Vendor number=vendor number .AND. Invoice number= Invoice number	Accounts payable clerk	Accounts payable amount, Invoice amount, Purchase order amount, Vendor number, Invoice number	Accounts payable clerk cannot be the same person as cashier.
77. Amounts for goods or services received are recorded in the appropriate period.	Valuation	Purchase activities are wrongly moved into other periods causes potential earnings manipulation.	Month (Product received date)=Month (Account Payable date)	Accounts payable clerk; Products receiver	Product received date; Account Payable date	Products receiver cannot be the same person as accounts payable clerk.
78. Accounts payable are adjusted only for valid reasons.	Completeness, Existence.	Avoid potential fraudulent change.	Accounts payable > Invoice amount	Accounts payable clerk; Payment approver	Accounts payable; Invoice amount; the approver of accounts payable change; reasons description of the amount change	The approver of the accounts payable cannot be the same person as accounts payable clerk.

79. Credit notes and other adjustments are accurately calculated and recorded.	Valuation	Missing available discount benefit.	(Discount due date - invoice date) > discount term days	Accounts payable clerk; Payment approver	Discount ratio; Discount terms; invoice date; payment date	NA
80. All valid credit notes and other adjustments related to accounts payable are input and processed.	Completeness Existence	Missing available discount benefit	Discount benefit conditions	Accounts payable clerk; Payment approver	Discount ratio; Discount terms; invoice date; payment date	NA
81. Credit notes and other adjustments are recorded in the appropriate period.	Valuation	Purchase credit notes are wrongly moved into other periods	Month (Purchase Order date)=Month(Credit notes date)	Accounts payable clerk; Purchase order requestor	Purchase Order date; Credit notes date	Payment approver cannot be the same person as accounts payable clerk.
82. Disbursements are made only for goods and services received.	Existence	Non-existed transactions can cause potential fraud.	Products received = Order quantity = Product quantity in invoice	Payment approver; products receiver	Product received date; Product received quantity; Invoice quantity; payment approved date; payment date	Products receiver cannot be the same person as payment approver.
83. Disbursements are distributed to the appropriate suppliers.	Existence	Payment to wrong vendors	Invoice number=Invoice number .AND. Vendor number=Vendor number .AND. Purchase order number=purchase order number	Purchase order requestor; payment approver; products receiver	Purchase order number, Purchase order date, Invoice number, vendor number, Invoice amount, Payment date	Purchase requestor cannot be the same person as payment approver; and payment approver cannot be the same person as cashier
84. Disbursements are accurately calculated and recorded.	Valuation	Incorrect payment causes potential loss.	Payment amount = invoice amount = purchase order amount	Cashier; Payment approver	Payment amount, Invoice amount, Purchase Order Amount, Vendor Number, payment date, purchase order date, products received date.	Payment approver cannot be the same person as cashier.

85. All disbursements are recorded.	Completeness	Duplicate payments causes potential loss.	Invoice date=Invoice date .AND. Invoice amount=invoice amount .AND. Vendor number=vendor number .AND. Invoice number= Invoice number .AND. Payment date <> Payment date	Cashier; Payment approver	Invoice date, Invoice amount; Vendor number, Invoice number; Payment date	Payment approver cannot be the same person as cashier.
86. Disbursements are recorded in the period in which they are issued.	Valuation	Payments are wrongly moved into other periods.	Month (Payment date =Month (Check date)	Cashier; Payment approver	Payment date; Check date	Payment approver cannot be the same person as Cashier.
87. Only valid changes are made to the supplier master file.	Completeness Existence	Accepting invalidated vendors can cause potential fraud.	Supplier master file change date > supplier master file change request date	Supplier master file change approver; supplier master file change requestor	Supplier master file change date, supplier master file change request date.	Supplier master file change approver cannot be the same person as requestor
88. All valid changes to the supplier master file are input and processed.	Completeness Existence	The error in the vendor master file provides potential fraud opportunity.	Supplier Number (master file) = Supplier Number (invoice)	Requestor and Reviewer of the vendor master file.	Vendor number, Vendor name, Vendor address, Vendor contact name, Master file updated time.	Requestor and Reviewer of the vendor master file cannot be the same person.
89. Changes to the supplier master file are accurate.	Valuation	Same as No.88 Control Objective				
90. Changes to the supplier master file are processed in a timely manner.	Completeness Existence	Same as No.88 Control Objective				
91. Supplier master file data remain up to date.	Completeness Existence	Same as No.88 Control Objective				

5. Application Control Objectives for the Inventory Cycle (Process Identity: Inventory Number)

Control Objectives	Financial Assertions	Risks	Conceptual Model and Criteria	Users (stakeholders)	Control Attributes	Segregation of Duties
92. Adjustments to inventory prices or quantities are recorded promptly and in the appropriate period.	Existence Completeness Valuation	Avoid the inventory account become an earnings manipulation pool.	Month (Products received) = Month (Accounting Period) .AND. Month (Finished Goods) = Month (Accounting Period) .AND. Month (Sales) = Month (Accounting Period)	Products receiver, Inventory manager, Sales representative, Purchase order staff.	Products received date, Finished goods date, products shipping date, Inventory accountant.	Inventory accountant cannot be the same person as products receiver or inventory manager.
93. Adjustments to inventory prices or quantities are recorded accurately.	Valuation	Avoid the inventory account become a earnings manipulation pool.	Quantity (Products Received) = Quantity (Purchase order) .AND. Unit Price (Products Received) = Unit Price (Purchase order)	Products receiver, Inventory manager, Sales representative, Purchase order staff.	Quantity of products received, Quantity of purchase order, Unit Price of products received, Unit Price of purchase order.	Inventory accountant cannot be the same person as products receiver or inventory manager.
94. Raw materials are received and accepted only if they have valid purchase orders.	Existence	Unauthorized raw materials purchase can cause potential fraud.	Date (Purchase Approved) > Date (Products Received)	Purchase Requestor; Purchase Approver; Products Receiver.	Request Date, Approval date, Requestor ID, Approver ID, Requestor Department, Vendors Number, Products received date, Product Quality Description.	Requestor cannot be the same person as approver.
95. Raw materials received are recorded accurately.	Valuation	Assure that the order amount is the same as the amount in the invoice.	Order Amount = Invoice Amount = Product received amount .AND. Order unit price = Invoice unit price	Purchase order accounting staff, products receiver.	Order Amount, Invoice Amount, Order Information Input Date, Product received amount, Vendors Number.	NA

96. All raw materials received are recorded.	Completeness	Assure that product received amount is the same as the amount in the invoice.	Invoice Amount = Product received amount .AND. Invoice quantity = Products received quantity	Purchase order accounting staff, products receiver.	Order Amount, Invoice Amount, Order Information Input Date, Product received amount, Vendors Number, Invoice quantity, Products received quantity.	Purchase order accounting staff and products receiver cannot be the same person.
97. Receipts of raw materials are recorded promptly and in the appropriate period.	Valuation	Avoid delayed actions to process raw materials purchase accounting activities.	(Date (Products received) +5) > Date (Product received system input)	Purchase order information recorder, products receiver.	Request Date, Approval date, Products received date, Product Quality Description.	Purchase order accounting staff and products receiver cannot be the same person.
98. Defective raw materials are returned promptly to suppliers.	Existence	Avoid delayed actions to return defective raw materials to suppliers.	(Date (Products received) +5) > Date (Shipping defective products)	Purchase order return staff, Products receiver.	Products received date, Defective products return request date, Return reason, Return approval date, Return shipping date.	NA
99. All transfers of raw materials to production are recorded accurately and in the appropriate period.	Valuation Completeness	The errors in the production process can cause wrong product cost and errors in financial report.	Amount (Finished goods) = Sum (Amount (raw materials cost) + Amount (manufacture overhead) + Amount (labor cost))	Production accountant, Inventory manager	Amount (Finished goods), Amount (raw materials cost), Amount (manufacture overhead), Amount (labor cost).	Production accountant and Inventory manager cannot be the same person.
100. All direct and indirect expenses associated with production are recorded accurately and in the appropriate period.	Valuation	Avoid using inventory record period to manipulate earnings.	Month (Accounting period) = Month(raw materials) = Month (manufacture overhead) = Month (labor cost)	Production accountant, Accounting information reviewer.	Date (raw materials cost), Date (manufacture overhead), Date (labor cost).	Production accountant cannot be the same person as the reviewer.

101. All transfers of completed units of production to finished goods inventory are recorded completely and accurately in the appropriate period.	Valuation Completeness	Avoid using inventory record period to manipulate earnings.	Month (Accounting period) = Month (Finished goods)	Production accountant, Accounting information reviewer.	Date (Finished goods), Date (raw materials cost), Date (manufacture overhead), Date (labor cost).	Production accountant cannot be the same person as the reviewer.
102. Finished goods returned by customers are recorded completely and accurately in the appropriate period.	Valuation Completeness	Avoid using the account of finished goods return to manipulate earnings.	Month (Accounting period) = Month (Finished goods returned) =	Production accountant, Accounting information reviewer.	Date (Finished goods returned), Return reason.	Production accountant cannot be the receiver of finished goods.
103. Finished goods received from production are recorded completely and accurately in the appropriate period.	Completeness Valuation	Avoid using inventory record period to manipulate earnings.	Month (Accounting period) = Month (Finished goods received) =	Production accountant, Accounting information reviewer.	Date (Finished goods received), Quality description, Finished goods received approval.	Production accountant cannot be the same person as the reviewer.
104. All shipments are recorded.	Existence	Missing Purchase Orders causes errors in financial report.	Use Gap detection analysis to find missing numbers in the purchaser order sequence number.	Purchase Requestor; Purchase Approver; Order information input staff	Purchase order number; purchase order date	Requestor cannot be the same person as the staff to input information.
105. Shipments are recorded accurately.	Valuation	Missing shipments may cause potential loss.	Use Gap detection analysis to find missing numbers in the shipments sequence number.	Shipments Requestor; Shipments Approver; Carrier of the shipments.	Shipments number; Invoice date, Invoice number, Customer number.	NA
106. Shipments are recorded promptly and in the appropriate period.	Valuation	The error in shipments record can cause mistakes to determine sales.	Month (Accounting period) = Month(Finished goods shipment) =	Production accountant, Shipment staff.	Date (finished goods shipment)	NA
107. Inventory is reduced only when goods are shipped with approved customer orders.	Completeness Existence	Unauthorized inventory reduced can cause potential fraud.	Date (Sales Approved) > Date (Finished goods shipment)	Sales Requestor; Sales Approver; Inventory manager.	Sales Request Date, Approval date, Requestor ID, Approver ID,	Requestor cannot be the same person as approver.

					Customer Number, Finished goods shipment date.	
108. Costs of shipped inventory are transferred from inventory to cost of sales.	Existence Valuation	The error in inventory record can cause mistakes in financial report.	Sum (Debit (Finished goods shipped)) = Sum(Credit(Finished goods)	Sales Approver; Inventory manager.	Debit amount of Finished goods shipped, Credit amount of finished goods.	Inventory accountant cannot be the same person as inventory manager.
109. Costs of shipped inventory are accurately recorded.	Valuation	The error in inventory record can cause mistakes in financial report.	Sum (Credit (Cost of goods sold)) = Sum (Debit(Finished goods)	Sales Approver; Inventory manager.	Credit amount of Finished goods, Debit amount of Cost of goods sold.	Inventory accountant cannot be the same person as inventory manager.
110. Amounts posted to cost of sales represent those associated with shipped inventory.	Completeness Existence	The error in inventory record can cause mistakes in financial report.	Sum (Amount (Finished goods shipped)) = Sum(Amount (Finished goods))	Sales Approver; Inventory manager.	Amount of Finished goods shipped, Amount of finished goods.	Inventory accountant cannot be the same person as inventory manager.
111. Costs of shipped inventory are transferred from inventory to cost of sales promptly and in the appropriate period.	Valuation	The error in inventory record can cause mistakes in financial report.	Sum(Amount (Finished goods)) = Sum(Amount (Cost of goods sold))	Sales Approver; Inventory manager.	Amount of Finished goods shipped, Amount of finished goods.	Inventory accountant cannot be the same person as inventory manager.
112. Only valid changes are made to the inventory management master file.	Existence Completeness	Accepting invalidated inventory master file can cause potential fraud.	Inventory master file change date > Inventory master file change request date	Inventory master file change approver; Inventory master file change requestor	Inventory master file change date, Inventory master file change request date	Inventory master file change approver cannot be the same person as requestor

113. All valid changes to the inventory management master file are input and processed.	Existence Completeness	The error in the Inventory master file provides potential fraud opportunity.	Inventory Catalog Number (master file) = Inventory Catalog Number (invoice)	Requestor and Reviewer of the Inventory master file.	Inventory number, Inventory name, Inventory manager name, Master file updated time.	Requestor and reviewer of the inventory manager master file cannot be same.
114. Changes to the inventory management master file are accurate.	Valuation	Same as No. 114 Control Objective				
115. Changes to the inventory management master file are promptly processed.	Existence Completeness	Same as No.114 Control Objective				
116. Inventory management master file data remain up to date.	Completeness Existence	Same as No.114 Control Objective				

6. Application Control Objectives for the Fixed Asset Cycle (Process Identity: Fixed Asset Number)

Control Objectives	Financial Assertions	Risks	Conceptual Model and Criteria	Users (stakeholders)	Control Attributes	Segregation of Duties
117. Fixed asset acquisitions are accurately recorded.	Valuation	An accurate fixed asset record need a well-defined criterion of categorization.	Unit price >= Defined criterion .AND. Usage period > 1 year .AND. Fixed asset Value = Payment of fixed asset	Fixed asset requestor; Fixed asset approver	Request date, Approval date, Requestor ID, Approver ID, Requestor Department, Fixed asset description, Fixed asset categorization criterion.	Requestor cannot be the same person as approver.
118. Fixed asset acquisitions are recorded in the appropriate period.	Valuation	The error in fixed asset record can cause mistakes in financial report.	Month (Accounting period) = Month (Fixed asset acquisition)	Fixed asset accountant, Fixed asset account reviewer	Date (Fixed asset acquisition)	NA
119. All fixed asset acquisitions are recorded.	Completeness	Missing fixed asset acquisition causes errors in financial report .	Sum (Amount (Fixed asset order)) = Sum(Amount (Fixed asset received)) = Sum(Amount(Fixed asset))	Fixed asset accountant, Fixed asset account reviewer.	Amount (Fixed asset order), Amount (Fixed asset received), Amount(Fixed asset)	Fixed asset accountant, cannot be the same person as fixed asset

						account reviewer.
120. Depreciation charges are accurately calculated and recorded.	Valuation	Avoid using changing depreciation methods to manipulate earnings.	Fixed assets depreciation$_{ij}$ = Fixed assets depreciation$_{ij+1}$ (i is for asset item, j is period time)	Depreciation recorder and approver.	Depreciation recorded time; Amount; Method change; Method change reason.	Approvers cannot be the same person as the record staff.
121. All depreciation charges are recorded in the appropriate period.	Existence Valuation Completeness	The error in depreciation record can cause mistakes to of the sales amount.	Month (Accounting period) = Month(Depreciation) =	Fixed asset accountant, Fixed asset depreciation ratio approval.	Date (Depreciation)	NA
122. All fixed asset disposals are recorded.	Existence	Any errors or event missed in the fixed asset disposals can cause mistake in financial report.	Sum (amount (Fixed asset disposed)) = Sum (Amount (sale of fixed assets) + Amount (gain from fixed assets sales)) = Amount (Proceeds from fixed assets disposals)	Fixed asset disposal accountant, Fixed asset disposal approver.	Amount (Fixed asset disposed), Amount (sale of fixed assets), Amount (gain from fixed assets sales), Amount (Proceeds from fixed assets disposals), Date (Fixed asset disposal), Fixed asset disposal reason.	Fixed asset disposal accountant and fixed asset disposal approver cannot be the same person.
123. Fixed asset disposals are accurately calculated and recorded.	Valuation	The errors in fixed asset disposals can cause mistake in financial report.	Sum (amount (Fixed asset disposed)) = Sum (Amount (sale of fixed assets) + Amount (gain from fixed assets sales)) = Amount (Proceeds from fixed assets disposals)	Fixed asset disposal accountant, Fixed asset disposal approver.	Date (Fixed asset disposal)	Fixed asset disposal accountant and fixed asset disposal approver cannot be the same person.
124. Fixed asset disposals are recorded in the appropriate period.	Valuation	The error in fixed asset record can cause mistakes in	Month (Accounting period) = Month(Fixed asset disposal)	Fixed asset accountant, Fixed asset account reviewer	Date (Fixed asset acquisition)	NA

		financial report.				
125.Records of fixed asset maintenance activity are accurately maintained.	Completeness	The fixed asset maintenance information is strong evidence for the value update of fixed assets.	Date (Fixed asset maintenance requestor) > Date (Fixed asset maintained)	Fixed asset maintenance requestor and approver.	Date (Fixed asset maintenance requestor), Date (Fixed asset maintained), Maintenance Reason description.	Fixed asset maintenance requestor and approver cannot be the same person.
126. Fixed asset maintenance activity records are updated in a timely manner.	Completeness	The fixed asset maintenance information is recorded timely.	Date (Fixed asset maintained) < Date (Fixed asset maintenance information input) + 5	Fixed asset maintenance requestor and approver.	Date (Fixed asset maintenance requestor), Date (Fixed asset maintained), Date (Fixed asset maintenance approved), Maintenance reason description.	NA
127. Only valid changes are made to the fixed asset register and/or master file.	Completeness Existence	The error in the fixed asset master file provides potential fraud opportunity.	Fixed asset Number (master file) = Fixed asset Number (General Ledger)	Requestor and Reviewer of the fixed asset master file.	fixed asset number, fixed asset name, fixed asset manager, fixed asset updated time.	Requestor and Reviewer of the fixed asset master file cannot be the same person.
128. All valid changes to the fixed asset register and/or master file are input and processed.	Completeness Existence	Same as No.129 Control Objective				
129. Changes to the fixed asset register and/or master file are accurate.	Valuation	Same as No.129 Control Objective				
130. Changes to the fixed asset register and/or master file are promptly processed.	Completeness Existence	Same as No.129 Control Objective				
131. Fixed asset register and/or master file data remain up to date.	Completeness Existence	Same as No.129 Control Objective				

7. Application Control Objectives for the Human Resources Cycle

Control Objectives	Financial Assertions	Risks	Conceptual Model and Criteria	Users (stakeholders)	Control Attributes	Segregation of Duties
132. Additions to the payroll master files represent valid employees.	Existence	The error in payroll master file can cause potential fraud and economic loss.	Date (Employee addition requestor) < Date (Employee addition approver)	Payroll master file change requestor and approver.	Date (Employee addition request), Date (Employee addition approval), Department, Home address, Employee type, Employee hired date.	Payroll master file change requestor and approver cannot be the same person.
133. All new employees are added to the payroll master files.	Completeness	Missing new employees can cause missing employees' payments.	Employee ID (master file) = Employee ID (Payroll list)	Payroll master file staff, Payroll payment staff.	Employee ID, Employee name, Employee home address, Employee hired date, Employee master file updated time.	Payroll master file staff and Payroll payment staff cannot be the same person.
134. Terminated employees are removed from the payroll master files.	Existence	Missing a terminated employee from the payroll master file can cause loss.	Date (Employee termination requestor) < Date (Employee termination approver)	Payroll master file change requestor and approver.	Date (Employee termination request), Date (Employee termination approval), terminated employee department, Home address, Employee type, Employee hired date.	Payroll master file change requestor and approver cannot be the same person.
135. Employees are terminated only within statutory and union requirements.	Completeness	Missing new employees can cause missing employees' payments.	Employee ID (master file) = Employee ID (Payroll list)	Payroll master file staff, Payroll payment staff.	Employee ID, Employee name, Employee home address, Employee master file updated time.	Payroll master file staff and Payroll payment staff cannot be the same person.

136. Deletions from the payroll master files represent valid terminations.	Completeness	The error in the updated payroll master file can cause potential fraud.	((Employee ID (master $file_i$) - Employee ID (master $file_{i-1}$)) = Employee Type (terminated) "i" is accounting period.	Payroll master file staff, Payment staff.	Employee ID, Employee name, Employee type, Employee terminated date.	Payroll master file staff and Payroll payment staff cannot be the same person.
137. All time worked is input.	Completeness	The time worked in the payroll account should match to the number input by the related department	Time worked (payroll system) = Time worked (Work assignment of departments) .AND. Employee ID (payroll system) = Employee ID (Work assignment of departments)	Payroll payment staff, Time worked record staff.	Employee ID, Employee name, Employee type, Time worked.	Payroll accountant cannot be the same person as the staff to record time worked.
138. Time worked is accurately input and processed.	Valuation	Same as No.138 Control Objective				
139. Payroll is recorded in the appropriate period.	Valuation	The mismatch of the payroll and the accounting period can cause mistakes to of the sales amount.	Month (Accounting period) = Month(Payroll time)	Payroll accountant, Payroll information reviewer	Date (Payroll time)	NA
140. Payroll (including compensation and withholdings) is accurately calculated and recorded.	Valuation	The payroll should match to the number that was approved.	Payroll Amount (payroll system) = Payroll Amount (Approved payroll system) .AND. Employee ID (payroll system) = Employee ID (Approved payroll system)	Payroll payment staff, Approved payroll system.	Employee ID, Employee name, Employee type, Payroll amount approved.	Payment staff cannot be the same person as the staff to approve payrolls.
141. Payroll is disbursed to appropriate employees.	Existence	Same as No.141 Control Objective				

142. Only valid changes are made to the payroll master files.	Existence Completeness	The error in the payroll master file provides potential fraud opportunity.	Employee ID (payroll system) = Employee ID (Approved payroll system)	Requestor and Reviewer of the payroll master file.	Employee ID, Employee name, Employee home address, Employee hired date, Employee master file updated time.	Requestor and Reviewer of the payroll master file cannot be the same person.
143. All valid changes to the payroll master files are input and processed.	Existence Completeness	Same as No.133 Control Objective				
144. Changes to the payroll master files are accurate.	Valuation	Same as No.143 Control Objective				
145. Changes to the payroll master files are processed in a timely manner.	Existence Completeness	Same as No.133 Control Objective				
146. Payroll master file data remain up to date.	Existence Completeness	Same as No.133 Control Objective				
147. Only valid changes are made to the payroll withholding tables.	Existence Completeness	Same as No.133 Control Objective				
148. All valid changes to the payroll withholding tables are input and processed.	Completeness Existence	The error from the payroll withholding tables can cause loss.	Date (Payroll withholding table change request) < Date (Payroll withholding table change request approval)	Payroll withholding table change requestor and approver.	Date (Payroll withholding table change request), Date (Payroll withholding table change request approval), Change reason.	Payroll withholding table change requestor and approver cannot be the same person.

149. Changes to the payroll withholding tables are accurate.	Valuation	The payroll withholding table change should match to the number that was approved.	Payroll withholding amount (payroll system) = Payroll withholding amount (Approved withholding ratio system) .AND. Employee ID (payroll system) = Employee ID (Approved withholding ratio system)	Payroll payment staff, Approved withholding ratio system.	Employee ID, Employee name, Employee type, Approved payroll withholding ratio system.	Payroll payment staff cannot be the same person as the staff to approve payroll withholding ratio system.
150. Changes to the payroll withholding tables are promptly processed.	Existence Completeness	Same as No.149 Control Objective				
151. Payroll withholding table data remain up to date.	Existence Completeness	Same as No.149 Control Objective				

Application Control Objectives for the Tax Cycle

Control Objectives	Financial Assertions	Risks	Conceptual Model and Criteria	Users (stakeholders)	Control Attributes	Segregation of Duties
152. Automated workflows are used for timely filing of returns.	Completeness	Avoid delayed actions to process tax return.	(Date (Tax return submit) + 2) > Date (Deadline of tax return)	Staff to submit tax return, and the approver of tax return.	Tax return submitted date, Tax return approval date, Tax return deadline.	NA
153. Tax payments are correctly calculated and recorded to the general ledger.	Completeness Valuation Existence	Errors in the calculation of tax payment can cause potential loss.	Sum (Amount (Tax payable)) = Tax Payment .AND. Amount (GL Tax payable) = Sum (Amount(Sub-ledger Tax payable))	Tax payable accountant, Tax payment approver, Cashier.	Tax payable accountant, Tax payment approver,	Tax payment approver cannot be the same person as cashier.
154. Tax exposures and valuation allowances are correctly calculated and recorded.	Completeness Existence Valuation	Same as No.153 Control Objective				

155. Tax expenses are recorded in the correct periods.	Completeness Existence Valuation	The mismatch of the tax expense and the accounting period can cause mistakes to of the expense amount.	Month (Accounting period) = Month (Tax Expense)	Tax accountant, Tax information reviewer	Date (Tax payable), Date (Tax expense)	NA
156. Permanent and temporary differences are identified and recorded accurately.	Completeness Existence Valuation	No criterion or approval process used to identify permanent and temporary differences.	Date (request of tax difference identification) < Date (review of tax difference identification)	Tax accountant, Tax information reviewer.	Date (request of tax difference identification), Date (review of tax difference identification), Criterion for tax difference identification.	NA
157.Correct book income is used in the tax accrual.	Completeness Existence	A wrong reconciliation of book income and tax income causes errors in financial report.	Sum (Net income – Accrued expenses+ Accrued Income) = Amount (Tax payable)	Tax accountant, Tax payment staff.	Amount (Net income), Amount (Accrued expenses), Amount (Accrued Income), Amount (Tax payable)	NA
158.Tax assets, liabilities and expenses are complete and correctly calculated and reported.	Completeness Existence	The errors in deferred tax liability and deferred tax asset cause error in financial report.	Sum (Revenue – Taxable revenue) = Amount (Deferred tax liability) .AND. Sum (Expense – Taxable expense) = Amount (Deferred tax asset)	Tax accountant.	Amount (Revenue), Amount (Taxable revenue), Amount (Deferred tax liability), Amount (Expense), Amount (Taxable expense), Amount (Deferred tax asset)	NA
159. Depreciation is calculated using appropriate bases, resulting in correct charges and tax ramifications.	Completeness Existence	The error of taxable depreciation deduction causes wrong tax payment.	Amount (Deferred tax) = Amount (Depreciation) – Amount (Taxable depreciation deduction)	Tax accountant, Fixed asset accountant.	Amount(Deferred tax), Amount(Depreciation), Amount(Taxable depreciation deduction)	
160. Sales and use tax is calculated appropriately, correctly and in a timely manner.	Completeness Existence	The mismatch of the tax sales and the accounting period can cause mistakes to of the tax payable amount.	Month (Tax period) = Month (sales) = Month(Products shipped)	Tax accountant, Sale accountant	Date (sales), Date (Products shipped).	NA
161. Value-added tax is correctly accounted for and	Completeness Existence	The wrong record of value-added tax can	Amount (Value-added asset gain) = Amount	Tax accountant, Capital gain	Amount (Capital gain), Name of the	NA

filed appropriately.		cause tax fraudulent result.	(Tax payable)	accountant	value-added assets, Approval of the gain recorded.	
162. Transfer pricing policies are up to date and accurately represented in the systems.	Completeness Existence	The error of the transfer pricing policy in intercompany activities can cause tax fraudulent result.	Sum (Intercompany $entity_i$) = Sum (Intercompany $entity_j$)	Recorders and approvers of intercompany elimination activities.	Intercompany entity name, Transaction amount, Transaction time, Activity description.	Recorders and approvers cannot be the same person.
163. All tax payments are accurately reflected in the general ledger.	Valuation	The error of tax payment record can cause potential loss.	Sum (Tax payable) = Sum (Tax payment)	Tax accountant, Cashier.	Amount (Tax payable), Amount (Tax payment)	Tax accountant and cashier cannot be the same person.
164. Property tax filings are timely and accurate.	Completeness Existence Valuation	The error of property tax filing can cause potential loss.	Month (property tax payment) = Month (Accounting period) .AND. Amount (property tax payable) = Amount(property current value)	Tax accountant, Fixed asset accountant.	Date (property tax payment), Amount (property tax payable), Amount(property current value)	

Appendix 3: Description of the Data of the Purchase-to-Payment Process

Attribute	Description	Source	Source		
D_Invoice_ID	The invoice ID in the Oracle Data Warehouse	Oracle DW	Jaggaer_Supplier_ID	the supplier ID in the SciQuest system	SciQuest
D_Invoice_NUM	The invoice number from the Oracle Data Warehouse	Oracle DW	Fulfillment_Center_ID	The ID of the fulfillment center	SciQuest
D_Invoice_Amount	The amount in the invoice	Oracle DW	Fulfillment_Center_Name	The name of the fulfillment center	SciQuest
D_Amount_Paid	The amount paid by the university	Oracle DW	Use_Payment_Terms	The terms of the payment conditions	SciQuest
D_Payment_Status_Flag	It shows the status of the invoice:	Oracle DW	Payment_Terms_Discount	The discount that the vendor	SciQuest

	paid, payable, concealed			offers for the purchases	
D_PO_Header_ID	The header ID of the purchase order	Oracle DW	Payment_Terms_Days	The payable days that the vendor offers for the purchases	SciQuest
D_Voucher_NUM	The identifier in this database; the invoice number from the vendor	Oracle DW	Payment_Terms_Net	The payable net days that the vendor offers for the purchase	SciQuest
D_Approval_Status	The approval of the invoice: approved, canceled,	Oracle DW	Threshold_Vendor_Terms	The threshold that the vendor offers for the purchases	SciQuest
D_Check_ID	The ID of the check paid	Oracle DW	P_PO	The Purchase Order Number	SciQuest
D_Period_Name	The name of the invoice period	Oracle DW	P_Creation_Date	The creation date of the purchase order	SciQuest
D_Created_By	Who created the record in the data warehouse	Oracle DW	P_Original_Revision_Date	The original revision date of the purchase order	SciQuest
D_Creation_Date	The date of the record creation	Oracle DW	P_Lst_Revision_Date	The latest revision date of the purchase order	SciQuest
D_Payment_Date	The date of the payment	Oracle DW	P_Last_Distribution_Date	The latest distribution date of the purchase order	SciQuest
D_Accounting_Date	The accounting date of the invoice	Oracle DW	P_Workflow_Completion_Date	The creation completion date of the purchase order	SciQuest
D_Distribution_Line_Number	The line number of the invoice distribution	Oracle DW	P_PO_Closed_Date	The closing date of the purchase order	SciQuest
D_Dist_Code_Combination	The combination with other datasets from other tables in the data warehouse	Oracle DW	P_Buyer_Username	The buyer user name of the purchase order	SciQuest
D_Oracle_Import_Date	The date to input the information into the Oracle data warehouse	Oracle DW	P_Buyer_First_Name	The buyer's first name of the purchase order	SciQuest
D_Code_Combination_ID	The ID of the record to combine with other datasets in the data warehouse	Oracle DW	P_Buyer_Last_Name	The buyer's last name of the purchase order	SciQuest
D_Segment1	The department by the first level hierarchy	Oracle DW	P_Buyer_Email	The buyer's email of the purchase order	SciQuest
D_Segment2	The department by the second level hierarchy	Oracle DW	P_Buyer_Department	The buyer's Department of the purchase order	SciQuest
D_Segment3	The department by the third level hierarchy	Oracle DW	P_Workflow_Step_Name	The name of the workflow of the purchase order	SciQuest
D_Segment4	The department by the fourth level hierarchy	Oracle DW	P_Workflow_Step_Date	The date of the workflow of the purchase order	SciQuest
D_Segment5	The department by the fifth level hierarchy	Oracle DW	P_Workflow_Step_Action	The Approval of the workflow of the purchase order	SciQuest

D_Segment6	The department by the sixth level hierarchy	Oracle DW	P_Approver_UserName	The username of the approver of the purchase order	SciQuest
D_Segment7	The department by the seventh level hierarchy	Oracle DW	P_Approver_First_Name	The first name of the approver of the purchase order	SciQuest
Invoice_ID	The invoice ID in the SciQuest system	SciQuest	P_Approver_Last_Name	The last name of the approver of the purchase order	SciQuest
Invoice_NO	The invoice number from the vendor (same as D_Voucher_NUM)	SciQuest	P_Approver_Email	The email of the approver of the purchase order	SciQuest
PO_Line_NO	How many lines in the same Purchase Order	SciQuest	P_Approver_Department	The department of the approver of the purchase order	SciQuest
Invoice_Owner	Who is the owner of the related purchase transaction	SciQuest	H_NetID	The netID of the employee	HR People Software
Supplier_Name	The name of the vendor	SciQuest	H_Name	The name of the employee	HR People Software
Supplier_Number	The identifier of the vendor from the university vendor master file	SciQuest	H_Last_Name	The last name of the employee	HR People Software
Supplier_Invoice_NO	The invoice number from the vendor (same as D_Voucher_NUM)	SciQuest	H_First_Name	The first name of the employee	HR People Software
PO_NO	The ID of the purchase order	SciQuest	H_Middle_Name	The middle name of the employee	HR People Software
Invoice_System_Created_Date	The invoice creation date in the SciQuest system	SciQuest	H_Country	The country of the employee	HR People Software
Invoice_System_Created_Time	The invoice creation time in the SciQuest system	SciQuest	H_Address1	The address of the employee	HR People Software
Invoice_Source	The source of the invoice: manual or electronic invoice input	SciQuest	H_Address2	The extended address of the employee	HR People Software
Invoice_Type	The type of the invoice	SciQuest	H_City	The city of the employee	HR People Software
Invoice_Total	The total amount of the invoice	SciQuest	H_State	The state of the employee	HR People Software
Invoice_Status	In the process, payable, paid	SciQuest	H_Postal	The postal code of the employees	HR People Software
Catalog_NO	The type of the catalog (Hosted or PunchOut Catalog)	SciQuest	H_Empl_Status_Descr	The status of the employee: active, tired, expired	HR People Software
Product_Nmae	The name of the purchased items	SciQuest	H_Empl_class	Seven levels of the employees' class	HR People Software
Invoice_Line_Unit_Price	The unit price for each line of the same invoice	SciQuest	H_Empl_class_Descr	1: salaried; 2: tradespeople, 3: Term assignment, 4: causal, 5:	HR People Software

				student, 6: TA/GA, 7: parttime lecture, 8: coadjutants, 9: nonemployees	
Invoice_Line_Currency	The currency used for each line of the same invoice	SciQuest	H_full_part_time_descr	The description of the employees' status: full time or part-time	HR People Software
Quantity	The quantity of the purchase in the invoice	SciQuest	H_Orig_Hire_DT	The original hire date of the employee	HR People Software
Invoice_Line_Extended_Price	The extended price for each line of the same invoice	SciQuest	H_Last_Hire_DT	The latest hire date of the employee	HR People Software
Previously_Sent_To_ERP	Is the invoice previously sent to the ERP system; True or False	SciQuest	H_Termination_DT	The termination date of the employee	HR People Software
Date_of_Export	The date of the invoice export	SciQuest	H_Address_All	The combination of the address of the employee	HR People Software
Time_of_Export	The time of the invoice export	SciQuest	HV_Jaggaer_Supplier_ID	The supplier ID in the SciQuest system	Vendor Master File
PO_Business_Unit_Vendor_ID	The vendor ID for the business unit of the purchase order	SciQuest	HV_Supplier_Name	The name of the supplier	Vendor Master File
Invoice_Discount	The discount that the vendor offers for the invoice	SciQuest	HV_Supplier_Number	The identifier number of the supplier	Vendor Master File
Paid_Date	The payment date of the invoice	SciQuest	HV_Vendor_Name	The name of the supplier	Vendor Master File
Payble_Date	The payable date based on the vendor's terms	SciQuest	HV_Address_Type	The address type of the supplier	Vendor Master File
Accounting_Date	The accounting date of the invoice	SciQuest	HV_Address_ID	The address ID of the supplier	Vendor Master File
Record_Date	The date to record the invoice in the system	SciQuest	HV_Col_3rd_Party_ID	The third party ID of the supplier	Vendor Master File
Record_NO	The record number of the purchased invoice	SciQuest	HV_Active	The supplier is active or not	Vendor Master File
Bill_To_Address_Internal_Name	The internal name of the purchase bill	SciQuest	HV_Delivery_Target	The target delivery location of the supplier	Vendor Master File
Bill_To_Address1	The address of the purchase bill	SciQuest	HV_Country	The country of the supplier	Vendor Master File
Bill_To_Address2	The extended address of the purchase bill	SciQuest	HV_Street_Line1	The street line of the supplier	Vendor Master File
Bill_To_Contact1	The contact name of the purchase bill	SciQuest	HV_Street_Line2	The extended street line of the supplier	Vendor Master File
Bill_To_Contact2	The extended contact name of the	SciQuest	HV_Street_Line3	The extended street line of the	Vendor Master

	purchase bill			supplier	File
Bill_To_City	The city of the purchase bill	SciQuest	HV_City/Town	The city/town of the supplier	Vendor Master File
Bill_To_State	The state of the purchase bill	SciQuest	HV_State	The state of the supplier	Vendor Master File
Bill_To_Postal_Code	The postal code of the purchase bill	SciQuest	HV_Postal_Code	The postal code of the supplier	Vendor Master File
Bill_To_Country	The country of the purchase bill	SciQuest	HV_Phone	The phone number of the supplier	Vendor Master File
Supplier_ID	the supplier ID in the SciQuest system	SciQuest	HV_Toll_Free_Phone	The toll free phone number of the supplier	Vendor Master File
Item_Type	It is NonCatalog product or punchout product	SciQuest	HV_Fax	The fax number of the supplier	Vendor Master File
Product_Description	the textual description of the purchased product	SciQuest	HV_Notes	The notes is for the supplier	Vendor Master File
Manufacturer	The manufacturer of the purchased product	SciQuest			
Buyer_Username	The username of the buyer of the transaction	SciQuest			
Buyer_First_Name	The first name of the buyer of the transaction	SciQuest			
Buyer_Last_Name	The last name of the buyer of the transaction	SciQuest			
Buyer_email	The email of the buyer of the transaction	SciQuest			
Workflow_Step_Name	The workflow step: matching exceptions or closed PO exceptions	SciQuest			
Workflow_Step_Date	The date of the workflow step (matching exceptions or closed PO exceptions)	SciQuest			
Workflow_Step_Action	The approval of the invoice	SciQuest			
Approver_username	The username of the approver	SciQuest			
Approver_First_name	The first name of the approver	SciQuest			
Approver_Last_name	The last name of the approver	SciQuest			
Approver_email	The email of the approver	SciQuest			

Appendix 4: The effectiveness score of the 35 most risky vendors

Vendors	NO_OF_RECS	V_score1	V_score2	V_score3	V_score4	V_score5	V_score8	V_score9	V_score10	V_score7	V_score6	Effectiveness_S
P19987632	4	0	0	0	7.5000	5.0000	0.0	0.0	0	112.75	0.00	-25.25
P198704691	2	0	0	0	7.5000	5.0000	0.0	0.0	0	18.90	0.00	68.60
P910132527	25	0	0	1	7.5000	5.0000	0.0	10.0	0	0.00	0.00	76.54
P910103248	1	5	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	77.50
P910260039	1	5	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	77.50
P910273143	1	5	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	77.50
P910273751	1	0	0	0	7.5000	5.0000	0.0	10.0	0	0.00	0.00	77.50
P920167672	2	0	0	0	7.5000	5.0000	0.0	10.0	0	0.00	0.00	77.50
P19958534	1	5	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	77.50
P910133000	8	0	0	0	6.8750	4.3750	0.0	10.0	0	0.00	0.00	78.75
P912732741	2	0	0	0	7.5000	2.5000	0.0	10.0	0	0.00	0.00	80.00
P910263727	5	0	0	0	6.5000	3.0000	0.0	10.0	0	0.00	0.00	80.50
P910268426	9	0	0	0	7.5000	1.6667	0.0	10.0	0	0.00	0.00	80.83
P915755807	6	0	0	0	5.0000	4.1667	0.0	10.0	0	0.00	0.00	80.83
P19987168	35	0	0	0	5.7143	2.0000	0.0	10.0	0	0.00	0.00	82.29
P910133827	1	5	0	0	7.5000	5.0000	0.0	0.0	0	0.00	0.00	82.50
P910167256	1	5	0	0	7.5000	5.0000	0.0	0.0	0	0.00	0.00	82.50
P910181982	2	5	0	0	7.5000	5.0000	0.0	0.0	0	0.00	0.00	82.50
P910183125	5	0	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	82.50
P910220466	1	5	0	0	7.5000	5.0000	0.0	0.0	0	0.00	0.00	82.50
P910276026	1	0	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	82.50
P910280680	1	0	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	82.50
P910280721	1	0	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	82.50
P915620207	1	0	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	82.50

P917162351	1	0	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	82.50
P918515704	1	0	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	82.50
P920550363	1	0	0	0	7.5000	0.0000	0.0	10.0	0	0.00	0.00	82.50
P19904429	1	5	0	0	7.5000	5.0000	0.0	0.0	0	0.00	0.00	82.50
P19939345	1	5	0	0	7.5000	5.0000	0.0	0.0	0	0.00	0.00	82.50
P19956529	1	5	0	0	2.5000	0.0000	0.0	10.0	0	0.00	0.00	82.50
P19984938	1	0	0	0	2.5000	5.0000	0.0	0.0	10	0.00	0.00	82.50
P19989227	3	5	0	0	7.5000	5.0000	0.0	0.0	0	0.00	0.00	82.50
P917674301	12	0	0	0	7.5000	3.3333	0.0	0.0	0	6.24	0.00	82.93
P910216029	4	0	0	0	7.5000	5.0000	0.0	0.0	0	1.79	2.50	83.21
P915451590	7	0	0	0	6.7857	0.0000	0.0	10.0	0	0.00	0.00	83.21
P910168457	10	0	0	0	0.0000	0.0000	0.0	0.0	0	16.66	0.00	83.34
P910281836	11	0	0	0	6.5909	0.0000	0.0	10.0	0	0.00	0.00	83.41
P910181882	18	0	0	0	2.5000	3.3333	0.0	0.0	10	0.00	0.00	84.17

Note: The V_scoreN means the violation score for the No. N business rule, which is defined in Table 1, e.g., the V_score1 is the violation score that the transaction violated the rule- “Invoice date has to be later than purchase order creation date

Appendix 5: Mathematics Basis for Multiple Correspondence Analyses

MCA is applied to tables with observations in the rows and categorical variables in the columns. The principal component analysis (PCA) is to summarize the relationships between the variables (columns). The aim of MCA depicts the relationships among observations (rows) and variables (columns). The information carried by a synthetic variable can be studied in terms of its categories, which represent both variables and a group of observations (all the observations who select this category).

The denotations are as follows: x_{ij} is the category chosen by the individual i for variable j; i varies from 1 to i and j from 1 to j. We consider categorical variable j to have K_j categories.

The data is analyzed according to the observations, the variables, and the categories. The statistical basis for each type of study is illustrated as follows.

Studying Observations

Studying the observations means understanding the similarities between the individuals in terms of all the variables. It is similar to cluster analysis to provide a typology of the observations: which are the most similar transactions? This similarity can uncover the knowledge of the implementation of internal controls. Observations are compared on a basis of presence–absence of the features that they presented in the operations. From this perspective alone, the distance between two transactions depends entirely on their characteristics and not on those of the other transactions.

(1) Distance between observations

The distances between transactions can be calculated by adding the differences between categories, $(x_{ik} - x_{i'k})^2$, and counterbalanced using a function inversely proportional to I_k (with I_k the number of individuals carrying category k). This distance (squared) is expressed as:

$$d_{i,i'}^2 = C\sum_{k=1}^{k} \frac{(X_{ik} - X_{ik'})^2}{I_k} \quad \text{(where C is a constant)} \qquad \text{(Equation 1.1)}$$

(2) Distance between categories

The distance between two categories k and k is calculated by counting the individuals which carry either category k or category k′, and counterbalancing using a function inversely proportional to I_k and $I_{k'}$. The category distance can be expressed as:

$$d_{k,k'}^2 = C' \frac{I_{K \neq K'}}{I_K I_{k'}} \quad \text{(where C' is a constant)} \qquad \text{(Equation 1.2)}$$

Using $I_{K \neq K'} = \sum_{i=1}^{I} (X_{ik} - X_{ik'})^2$, Equation 1.2 can be developed into:

$d^2_{k,k'} = C'(\frac{X_{ik}}{I_k} - \frac{X_{ik'}}{I_{k'}})^2$ (where C' is a constant) (Equation 1.3)

(3) Inertia of categories

The inertia for a category K can be calculated from the distance from K to the center of gravity of the cloud of categories with coordinates of 1/I, which is the average vector of all of the categories.

$d^2_{k,G_K} = I \sum_{I=1}^{I}(\frac{X_{ik}}{I_k} - \frac{1}{I})^2$ (Equation 1.4)

(4) The transition relation

The variables can be represented by calculating the correlation between the transaction' coordinates on one component and each of the categorical variables. As in PCA, transition relationship links the cloud of observations N_I to the cloud of categories N_K. The transition formulae are as follows.

$F_s(i)$ and ($G_s(k)$ designate the coordinate for individual i (and category k, respectively) on the component of rank s.

$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j=1}^{J} \sum_{k=1}^{K_j} \frac{x_{ik}}{J} G_s(k)$ (Equation 1.5)

$G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_{1=1}^{I} \sum_{k=1}^{K_j} \frac{x_{ik}}{i_k} F_s(i)$ (Equation 1.6)

On the component of rank s, up to the multiplicative factor $1/\sqrt{\lambda}$, the first relationship expresses that individual i is at the center of gravity of the categories that it carries (as $x_{ik} = 0$ for the categories that it does not carry). The second relationship expresses that category k is at the center of gravity of the individuals that carry it.

(5) Interaction measurement

MCA is often based upon proximities between points in a low-dimensional map, i.e., two or three dimensions. The analysis has two measurements to express the interaction. The squared cosine helps locating the factors important for a given observation or variable; the contribution helps locating the observations important for a given factor.

The squared cosine between row i and factor l and column j and factor l are obtained respectively as:

$O_{i,l} = \frac{f^2_{i,l}}{d^2_{r,i}}$ and $O_{j,l} = \frac{g^2_{i,l}}{d^2_{c,j}}$ (Equation 1.7)

(with $d^2_{r,i}, and\ d^2_{c,j}$, being respectively the squared distance from the i-th element to respective barycenter, and the square distance from the j-th element to the respective barycenter).

The contribution of row i to factor l and of column j to factor l are obtained respectively as:

$$t_{i,l} = \frac{f_{i,l}^2}{\lambda_l} \quad and \quad t_{j,l} = \frac{g_{i,l}^2}{\lambda_l} \qquad \text{(Equation 1.8)}$$

where λ_l is the eigenvalue of factor l.

References:

Greenacre, M. (2017). Correspondence Analysis in Practice, Third Edition, Taylor & Francis Group, LLC.

Husson, F., Lê, S., Pagès, J. (2017). Exploratory Multivariate Analysis by Example Using R, *Taylor and Francis Group*. Boca Raton London New York

Abdi, H. & Valentin, D. (2007). Multiple Correspondence Analysis, In Neil Salkind (Ed.) Encyclopedia of Measurement and Statistics.