Knowledge Discovery for Continuous Financial Assurance Using Multiple Types of Digital Information

Daniel E. O'Leary

University of Southern California

oleary@usc.edu

Comments are Solicited

June 2011, Revised November 2012, Revised February 2012

Keywords: Knowledge discovery, social media, Continuous financial assurance, Business Performance Management, Reputation Monitoring

Acknowledgement: An earlier version of this paper was presented at Rutgers University, on November 4, 2011, at the "World Conference on Continuous Auditing and Reporting." The author would like to thank the participants from the conference for their comments.

Knowledge Discovery for Continuous Financial Assurance Using Multiple Types of Digital Information

Abstract

This paper investigates some extensions to what has emerged as classic continuous (financial) assurance (CFA). First, this paper suggests broadening the base of information that is used in such systems to include emerging social media. Second, because of the new source of data, the methodologies also need to be expanded to include knowledge discovery approaches to facilitate analysis of new and existing information. Third, I suggest that continuous financial assurance begin to integrate with other business efforts at monitoring and knowledge discovery, for example, "business process management." Ultimately, an integrated system could gain from information and discoveries in these other monitoring and knowledge discovery efforts. Finally, this paper suggests that additional proactive knowledge discovery be used in addition to classic continuous financial assurance data to facilitate additional discovery.

Introduction

In an important paper, Vasarhelyi et al. (2004) established the foundations of continuous financial assurance. In particular, Vasarhelyi et al. (2004) have defined continuous assurance as "… a methodology for the analytic monitoring of corporate business processes taking advantage of the automation and integration of business processes brought about by information technologies."

As seen in this classic description, continuous financial assurance is largely aimed at "monitoring" without a specific focus on knowledge "discovery." In particular, although there may be analytic approaches used in continuous assurance, there has been limited use of knowledge discovery approaches. Further, although monitoring suggests analysis in real time, there is limited depth associated with simply "monitoring" in continuous financial assurance. As defined by Wikipedia, "monitoring" suggests being aware of the state of a system. Knowledge "discovery" suggests a greater depth of analysis beyond monitoring, by searching for knowledge. As a result, this paper applies notions of knowledge discovery to accounting and auditing, extending the notion of what is referred to as "continuous (financial) assurance" focusing on "knowledge discovery." In particular, I investigate the use of knowledge discovery across new information sources, e.g., social media and classic existing information sources, extending the methodologies of continuous financial assurance.

Further, researchers in continuous financial assurance (CFA) have treated CFA as an independent investigation, one that does not relate to other corporate efforts or other academic disciplines, and an effort that stands alone. However, there are a number of other academic and corporate efforts going on at the same time that CFA has been developed, e.g., business process management. As a result, it is important to suggest that financial monitoring, discovery and assurance should be integrated with other forms of monitoring and discovery, such as business process management, leveraging data and capabilities from those areas and ultimately hoping for symbiotic results.

Purposes

Accordingly, the purpose of this paper is five-fold:

- Briefly review classic continuous financial assurance,

- Relate continuous financial assurance to other types of continuous monitoring and analysis, such as business process management,
- Examine some extensions to classic continuous assurance of financial information by extending the information sources to include other digital information (e.g., social media),
- Apply knowledge discovery to continuous financial assurance,
- Examine some emerging "proactive" uses of social media applications (e.g., Facebook and LinkedIn) for continuous monitoring and assurance.

This Paper

This paper proceeds as follows. The next section briefly reviews continuous financial monitoring and assurance, including discussing its focus on "verification." The following section discusses a number of related approaches used in business that use monitoring and discovery, including gathering data from social media. The fourth investigates IBM's use of knowledge discovery in social media, corporate brand and reputation analysis (COBRA). The fifth section then applies that basic model and the use of knowledge discovery to continuous financial assurance. The following section analyzes the application of knowledge discovery to classic financial information in CFA. The penultimate section discusses some additional proactive approaches to potentially generate additional data for monitoring and discovery. Finally, the last section summarizes the paper, discusses its contributions and reviews some potential extensions.

Continuous Financial Assurance (CFA): Classic and Emerging Issues

Attaining classic continuous (financial) assurance ultimately requires continuous monitoring using different analytic approaches and a range of different types of information sources in order to accomplish different sets of verification activities, distinguished by levels that focus on different types of verification. Vasarhelyi et al. (2004, exhibit 1, p. 6) suggest four different levels of analytical monitoring: Level 1 is for transaction verification, Level 2 is for compliance verification, Level 3 is for estimate verification, and Level 4 is for judgment verification.

According to Dictionary.com, verification is "evidence that establishes or confirms the accuracy or truth of something." Thus, as defined, CFA is searching to confirm the accuracy or truth of transactions, compliance, estimates and judgment. Unfortunately, there are other concerns potentially associated with CFA. For example, O'Leary (1991) was concerned about how knowledge discovery could be used to undermine security and quality of data, while O'Leary (1995) was concerned about privacy issues arising from knowledge discovery. Further, verification does not explicitly search for process issues such as potential fraud and inefficiencies. As a result, the apparent focus on "verification" limits the view of what CFA is, the data that it uses and how that data is used. Accordingly, there are other concerns associated with CFA beyond verification issues. As seen in table 1, there has been substantial effort devoted to continuous monitoring of auditing, but other issues, such as continuous monitoring of blogs and other social media, have received scant attention (table 1).

Classic Data being Monitored

The focus on "verification" probably limits CFA to analyzing existing sources of data. As noted in Vasarhelyi et al. (2004), CFA has four different "Levels." Levels 1-3 are focused on

transaction data. Level 4: Judgment verification that " … searches in litigation databases and searches in major news sources" can be used. Thus, historically, CFA systems have focused on monitoring financial information.

As a result, important sources of potential information that are not part of Levels 1-4 "verification" are not included. For example, over the last five years or so there has been a substantial increase in the use of social media and other Internet-based information, however, social media is not likely a source for verification of transaction data. Social media capture conversations about a range of issues in text. As a result, that text could provide insight into different issues and could signal discovery of other events that may be interesting and worth further analysis in finance and corporate-wide.

Emerging New Data Sources

Accordingly, this paper extends the scope of information used in continuous financial monitoring by focusing on those new data sources. Recently, a number of types of available digital information have emerged as part of social media, including Blogs, Wikis, Micro – Blogs (Twitter), Message boards (RagingBull.com, TheLion.com, Yahoo.com), LinkedIn, Facebook, etc.

In general, social media work to remove information asymmetries by increasing the "scope" or "reach" of information dissemination, and making information that is private or available only in "limited scope," more broadly available. Blogs create and reuse information and make others aware of that information. Information that appears in a blog in a small corner of the world can be captured and repeated by others, further dispersing the information to larger audiences.

In addition, in many cases information that shows up in social media, such as blogs and micro blogs precedes disclosure of that information in news media, such as television and radio. In particular, there is some evidence that information shows up in …

- Blogs and micro blogs and then in news media
- Blogs also reuse information getting it into the public view
- Blogs and micro blogs and then shows up in the price of stocks
- Message boards and then in the price of stocks.

As examples, substantial research related to financial message boards has been developed, including the following. Tumarkin (2002) investigated messages on RagingBull.com, and found that abnormal positive returns preceded the days where there were strong positive reviews on the message boards. Sabherwal et al. (2008) studied messages on "TheLion.com" and found that there were abnormal returns if a stock was one of the 10 most talked about. Lerman (2010) analyzed Yahoo financial message boards for their use of accounting terms and other issues. There is also evidence that blog information is correlated with message board information (O'Leary 2011).

Accordingly, social media allow us to identify and monitor a wide range of risks and other kinds of information. Also, social media can provide data that will allow us to focus on issues other than verification. I have listed some potential phrases of interest and their corresponding number of Google pages, in table 2.

Classic Approaches to CFA

Vasarhelyi et al. (2004) summarized a number of innovative approaches used to analyze the available data using four different verification levels. Level 1 includes a review of data, reconciliations, automated confirmations and rule-based transaction evaluation. Level 2 employs XML data, with continuity equations, structural knowledge and time series analysis. Level 3 uses continuity equations and time series analysis. At level 3 Vasarhelyi et al. (2004) note that the intuition of experts can be captured in model parameters. Finally, Level 4 employs expert systems, continuity equations and time series analysis. As Vasarhelyi et al. (2004) note, expert systems are used to formalize measurement rules, used to capture accounting standards.

Emerging Approaches to CFA

Below I will extend the set of approaches available for use in CFA. In particular, I will discuss the potential integration of classic knowledge discovery approaches to analyze new and classic CFA data.

Other Forms of Continuous Monitoring and Analysis

Continuous monitoring occurs in many different kinds of applications that have been developed for business and other settings. Further, along with each monitoring task, knowledge discovery has emerged as an important component. This section examines some of those other types of continuous monitoring that occur with other applications.

Intrusion-Detection Systems

Intrusion-detection systems are aimed at being able to identify when an intruder has control or is attempting to take control of a system. Typically, such systems establish a set of expectations and then continuously monitor the system to determine if and when those expectations are violated. Notions of "intrusion-detection" (e.g., O'Leary 1992) have been built into many systems, including virus protection systems. Further, over the years there have been a number of proposals aimed at using knowledge discovery as part intrusion-detection systems (e.g., Lee and Stalfo 1999 and Xiaohui et al. 2010). Intrusion-detection systems can benefit from analysis of social media for those settings where intrusions are discussed, e.g., in forums, etc., gather information about actual and potential attacks, vulnerabilities and compromised systems.

Business Process Management (BPM)

Firms also have used classic monitoring of KPIs (key performance indicators) for management purposes for what is often referred to as "Business Process Management" (e.g., O'Leary 2008). Typically, strategy is mapped into the choice of KPIs, etc., so that the system allows mapping of performance into accomplishment of strategy. In so doing, firms have made use of a range of different types of knowledge discovery, including so-called "process mining" (Van de aalst and Weijters 2004). BPM could also employ information from social media as discussed in this paper to gather informal comments and information about business processes ("this system sucks").

Competitive Intelligence

Firms have been actively using continuous monitoring of competition (e.g., Kahaner 1997), using what has been referred to as "Competitive intelligence" through analyzing a wide range of information on the web, including

- Products,

- Advertising
- Investments
- Recruiting Needs and other issues

A number of researchers have used knowledge discovery approaches to facilitate analysis of competitive intelligence (e.g., Desouza 2001 and Shih et al. 2010). Competitive intelligence typically draws from virtually every available source of Internet information,

Reputation Monitoring and Analysis

Recently firms have begun managing their "reputations" (e.g., O'Leary 2010). In particular, IBM proposed that companies develop and maintain systems referred to as "public image" monitoring and analysis systems. As a result, those systems are designed to facilitate gathering public information on risks related to reputation (e.g., Hootsuite and others). For example, Wal-Mart reportedly has a "war room" where they monitor, analyze and manage their reputation. These systems consider a wide range of different media, e.g., blogs, news media, etc. with a general focus on legal and marketing issues (e.g., Spangler et al. 2007).

IBM's COBRA apparently was designed to monitor specific topics, such as events of interest in marketing, through analysis of information in social media. As an example, in one company the concern was with monitoring expressions of "Outrage," "Boycott," and "Stop Buying." By being aware of the discussion of specific topics the firm was able to negotiate a solution agreeable to both the social media community and the company. Because COBRA relates closely to some of the use of knowledge discovery for CFA I will discuss it in greater detail in the following section and use it as the basis of generating knowledge discovery for CFA using social media.

Continuous Financial Assurance

CFA relates to each of these other types of systems. For example, it is likely that an intruder may target financial systems. As a result, financial transaction processing systems could provide input to intrusion-detection. Business process management (BPM) is likely to employ a number of KPIs, many of which would be financial. Further, BPM is likely to analyze a number of processes, including financial and accounting processes. Accordingly, there is a need for BPM-like information about finance and accounting systems. Competitive intelligence will require bench-marking and much of that activity could include financial information. Finally, reputation monitoring may include a firm's reputation with its partners. As a result, a range of financial policies could influence those relationships. Accordingly, requirements of a reputation monitoring system could be related to CFA.

Summary

Businesses monitor a range of different types of information with different goals. Across each of the areas identified there has been a focus on knowledge discovery. In addition, increasingly, these approaches have included a broad range of information including newly emerging sources, such as blogs, wikis, etc. Two of the recent trends over time in these areas are to embed knowledge discovery and to include analysis of social media and other digitally-available information.

IBM's COBRA

IBM's COBRA (Corporate Brand and Reputation Analysis) was developed with the advent and growth of social media. In particular, COBRA aimed at gathering information from social media and related sources that were important to corporations' understanding of brands, products or even corporate level reputations. The available materials about COBRA are diverse, with contributors from IBM all over the world providing multiple perspectives. Accordingly, I have summarized two of those points of view in figure 1 and in a discussion that captures a consolidated view of the basic processes underlying COBRA. In that consolidated view there are six basic steps in the capture, analysis and discovery of knowledge using COBRA.

Determine Discovery Topics

Initiating COBRA requires determining the overall focus of the monitoring and discovery, initiated with the specific organization and their key concerns. Accordingly, this step is likely to reflect key threats, opportunities and the organization's strategy. For example, as seen in Kreulen (2008), at this stage the analysis would generate particular brands (Kisses or Reese's) and companies (Hershey, Nestles) of specific interest to initiate the process.

Generate Discovery Point of View

The particular point of view is generated in part, based on the specific topics of interest. For example, as also seen in Kreulen (2008), in the case of "Hershey's" "Kisses," the "point of view" could include "choking," "cruelty," "diabetes," and "obesity." Different points of view are then modeled for their potential appearance in social media. For example, "cruelty" was modeled to include "child abuse" and other statements (see figure 2).

Management could provide the points of view based on their concern or strategy. Alternatively, critical points of view could be generated from an analysis of social media, e.g., Twitter Tweets or from blogs etc. Word counts could be used to provide one ranking of importance of particular "points of view." Alternatively, an integrated supervised analysis of social media, based on input and discussions with management could be used.

Monitor and Capture Content

This stage can employ multiple, focused search engines, designed to facilitate capture of content from social media (blogs, message boards and news) in response to queries about joint discovery topics and points of view. In particular, the monitoring and capture process can focus on particular news sources or blogs, using specific search engines. For example, IBM (no date), noted that Boardreader was used as part of COBRA to monitor and find content. Using queries based on the discovery topics and points of view, specific social media would be searched.

Understand Sentiment and Influencers

After the social media has been captured, e.g., specific messages have been identified as important for a particular discovery topic and point of view, there is a need to summarize the content to determine, in general, the sentiment of the message (positive or negative). This is typically done using dictionaries of terms casts as either positive or negative sentiment. As a example, occurrence of the term "love" is typically cast as indicating a positive sentiment about the message. O'Leary (2011) discusses these and other issues in detail. Further, there is concern for who was responsible for the content and who was responsible for the messaging, in the case the information was forwarded on, since they represent the "influencer." That influencer

information also can be captured to study the source and growth of topics and actors involved in the discussion.

Generate Alerts on Tracked Topics

Alerts typically are tracked using a dashboard (e.g., Kreulen 2008). Alerts can take multiple forms. For example, if social media chatter reaches a certain level of activity that could be of interest to a manager. Further, management is likely to be concerned about the issues if there is growing chatter about an issue. In addition, alerts can focus on trends in topics co-occurrence with each other.

Discover Emerging Topics

Discovery of new and emerging topics of potential concern can come from management or from continued analysis of social media. One way of discovering emerging topics is to tract the number of messages/communications that are occurring regarding a particular topic that is not a previously specified point of view. As the occurrence of those non-specified points of views increases, they can be added to specified points of view. If the number of messages/communications of some not specified as a point of view keeps increasing then that could be a signal of a new topic of importance. Similarly, if chatter about a topic decreases below some threshold then that may signal a lack of interest.

Continuous Financial Monitoring and Discovery (CFMD) Using Social Media

The COBRA approach can be used to incorporate an analysis of social media into classic continuous assurance and the resulting continuous financial monitoring and discovery. This section applies the basic approach in the previous section and in figure 1 to CFMD.

Specifically, the analysis of social media can focus on a range of issues, including

- Search for information about fraud and other misbehaviors, e.g., misuse of assets, etc.
- Search for information about financial fraud
- Identify emerging risks, and monitor information about existing risks

Determine Discovery Topics

Topics that belong in CFMD are likely to depend on a number of factors, including focus of the system, company's industry, threats to their financial processes, etc. (e.g., a SWOT-based analysis). In any case, an ontology of such topics can form the basis of monitoring and future discovery. Such topics are chosen based on what it is hoped that the system will find. For example, there may be concern that Twitter messages will provide insight into information leakages about earnings. Another concern could be with "fraud," so in that case, that term or broad-based approaches to accomplishing fraud could be integrated into the ontology. In addition, the analysis might be interested in financial department "productivity" or "system inefficiencies." Discovery topics also can be influenced by the specific type of analysis being undertaken. In CFMD there are a number of potential topics, including:

- Financial Assurance
- Risk Analyzes
- Process Management and Control
- Regulation Audits
- IT Audits

- Fraud Audit
- Internal Audit
- Improper Reporting

Generate Discovery Point of View

The discovery point of view provides a setting for generating greater detail. In the case of fraud, there may be specification of particular methods or signals of fraud. For example, the search might be for signals such as "cash has disappeared." In the case of productivity, there may be a search for content such as "I left work early today." In the case of leakages of earnings information, the search could include an investigation of Twitter messages for information about an "earnings party." Further system inefficiencies might be captured with a search for information such as "our XYZ system sucks." As a final example, there could be an analysis of encompassing terms such as company is on a "downward trend" or "downward spiral." These points of view, summarized in table 2, would map into a tree similar to figure 2.

Monitor and Capture Content

Monitoring and capturing content can include choosing social media generated by employees or other companies. The more narrow the choice to particular people or sources, the more likely that the company is likely to be open to criticism, unless there are specific rationales for their focus. For example, there may be ethical issues at this point if a company decides to focus on messages and news of specific individuals. However, if there is a narrow focus about a specific group then such analysis may generate better results.

Understand Sentiment and Influencers

Understanding sentiment and influencers likely is different when the concern is with occurrence of fraud or understanding issues in productivity. In the setting of fraud it is unlikely that there are positive and negative sentiment issues, but there definitely would be influencers, participating in the fraud. However, information and messages concerning productivity could be positive or negative, and whether the sentiment was positive and negative could be important information. Further, influencers, or those stating whether or not they were being productive would be important information in tracking productivity issues ("I left work early").

Generate Alerts on Tracked Topics

Probably unlike other types of continuous monitoring and analysis, alerts on topics can and should be generated at different organizational levels and functions. In particular, alerts need to be generated at implementation and managerial levels to ensure that the alerts are acted on. If there is fraud then potential different actors being alerted may be involved. For example, in the case of fraud, internal audit would need to be contacted. As a result, by involving multiple actors or multiple levels of actors, there is the opportunity potentially to mitigate problems.

Discover Emerging Topics

Discovery of emerging topics can result from tracking the occurrence or non-occurrence of particular issues. In the case of "occurrence," the emerging co-occurrence of issues with existing issues can lead to finding previously un-identified emerging topics of importance. Further, management may contribute to that emerging list by including (or removing) issues of concern to them.

Knowledge Discovery Using Financial Information in CFA

Knowledge discovery can be extended to other uses in CFA. In particular, knowledge discovery approaches can be used on financial information (e.g., transactions) and on both the transaction information and integrated versions of financial information and social media information, although we will focus on the financial information here.

There are a number of approaches to using knowledge discovery using financial information to try and determine the existence of fraud. Two potential approaches are association mining and decision trees. First, association rule mining is a knowledge discovery approach that examines data for associations and relationships in the financial information. Different rules can be used to capture patterns (e.g., Ceglar and Roddick 2006). Second, knowledge discovery might be done using induction of decision trees (e.g., Quinlan 1979, 1986, and others) on financial information.

To-date, apparently there is a relatively limited set of literature on integrating knowledge discovery (e.g., data mining) to investigate financial fraud. Phua et al. (2010) provide a comprehensive survey of a range of fraud applications and find only a few data mining applications examining financial fraud. Kirkos et al. (2007) used financial statement ratios as input to knowledge discovery approaches. They found a relatively small number of ratios determined how results were categorized. Rosset et al. (1999) used an approach based on Quinlan's work in the telecommunication's industry using financial statement information.

Although there have been other applications, it is apparent that the academic focus has been on financial statement level data. Unfortunately, although most research has been with financial statement level information, perhaps the deepest and most rapidly available insights are likely to occur with transaction information. However, researchers have had limited availability of transaction level information. Accordingly, future research should focus on developing and using knowledge discovery approaches on transaction level information in order to generate knowledge discovery breakthrus.

Additional Emerging Proactive Uses of Web 2.0 Concepts

This paper has suggested using additional digital information related to social media and using knowledge discovery approaches on both that social media data and on classic financial statement and accounting information. Typically, this has meant passively using monitoring and knowledge discovery on digital materials as they are generated.

However, we can also ask the question, "How can we use LinkedIn, Facebook, TheLion.com, etc., "Proactively"?" breaking away from passive monitoring and discovery, to proactive monitoring and discovery.

One approach is to create bogus LinkedIn, Facebook, etc. accounts to try and capture or solicit information, collaboration, etc., using an approach similar to "honeypots." According to Wikipedia, "a honeypot is a trap set to detect, deflect, or in some manner counteract attempts at unauthorized use of information systems." For example, on those social networks, from bogus accounts, there could be a request or search for information. As such, there could be questions asked to the social networks about the firm and its activities, earnings, etc., in an attempt to derive "inappropriate" information. From that activity using potentially fraudulent requests, the organization may be able to actively discover any inappropriate disclosures and uses of

information. Unfortunately, there are likely to be potential information disclosure and ethical issues associated with this approach.

Information Disclosures

As noted above, social media can result in broad based information disclosures. As a result, if the firm is successful at being able to generate inappropriate information using this approach, such as implied earnings information ("earnings party tonight"), the information may be made broadly available, causing potential damage.

Ethical Issues?

Arguably there are potential ethical issues associated with these proactive approaches. For example, firms must address issues, such as, is it ethical to monitor and use knowledge discovery on specific LinkedIn, etc. accounts? Or would it be ethical to create "honey pot" social media accounts. However, as noted in Searcey (2009), "You can't post something on the Internet and claim breech of privacy when someone sees it."

An interesting question is an ethical comparison is, which of reputation monitoring or competitive intelligence or fraud analysis is the "most accepted?" In particular, is it "more acceptable" to look internally, at employee activity, or externally at what others think or are doing? Further, does the reaction vary across countries? As a result, before such approaches are done, it likely is important to ask if there are any "cultural" issues or other concerns with such an approach. For example, would there be any backlash if it was found that a firm was doing these activities?

Summary, Contributions and Extensions

This paper had five different purposes:

- Briefly review continuous financial assurance,
- Relate continuous financial assurance to other types of continuous monitoring and analysis,
- Examine some extensions to classic continuous assurance of financial information by using knowledge discovery on other digital information (e.g., social media),
- Suggest the use of knowledge discovery as a means of analyzing financial transaction data as part of CFA, and
- Examine some emerging "proactive" uses of social media applications, such as "honey pot" social media.

Contributions

This paper has a number of contributions. First, this paper analyzed the verification nature of CFA and suggested a broader-based approach. In particular, it was suggested that a focus on verification limited the potential use, data and capabilities of CFA. Second, this paper has suggested a basic structural change of CFA from "monitoring" to "monitoring and discovery." Third, this paper has noted that CFA is not the only monitoring activity of corporations and that some of those other activities have similar functions, approaches and data. Fourth, the paper has outlined how "discovery" can be embedded into CFA through knowledge discovery analysis of classic financial information and emerging social media sources. Finally, rather than relying only on a passive historical discovery approach, this paper suggests some proactive approaches to discover issues of concern.

Extensions

There are a number of different potential extensions to this research.

CFA/Fraud Ontology. It is apparent that there is a need for an ontology to help drive financial monitoring and discovery. Further, that ontology needs development in multiple areas, such as productivity, system quality and fraud. Such an ontology would likely to be of interest to other functional areas and systems, including business process management. One approach to that issue is given in Leary et al. (2003).

Monitoring and Discovery Information Systems. We have summarized a number of different types of monitoring and discovery in this paper. Unfortunately, there seems to be limited integration across the different functional efforts. In particular, each seems heavily silo'd into independent efforts. CFA is silo'd in finance and accounting, competitive intelligence is silo'd in marketing, etc. However, there are many overlaps between the different areas of business process management, competitive intelligence, intrusion detection, reputation analysis and continuous financial monitoring. Accordingly, it is inevitable that at some point these efforts are integrated or at least made aware of each other. Further, it likely is critical that they examine those situations where independently there is no anomalous or critical result to learn from, but together with the integrated flows of information, there are curious events that need further investigation and knowledge discovery. Such an integration of these kinds of information systems may broadly emerge as "monitoring and discovery information systems" or "monitoring and discovery decision support systems," specifically aimed at an organization's set of information systems designed to monitor and discover.

Hierarchy of CFA Activities. O'Keefe and O'Leary (1993) established a hierarchy of verification and validation of intelligent systems. That hierarchy included verification, validation, credibility, assessment and evaluation. Based on the discussion in this paper, it would appear that CFA is "bigger" than just verification and that a similar hierarchy of activities could be established.

Knowledge Discovery on Integrated Information. In this paper I have focused on using knowledge discovery on social media and on transaction information. In particular, I did not integrate the two sources of information. However, knowledge discovery could be done using both sources of information, attempting to determine the extent to which social media-based exchanges generate responses in the financial data or conversely. For example, humans may note anomalous transactions in comments to others, in social media or other outlets, providing potential insights ("something is fishy").

Make Data Sets Available. In some branches of computer science, test data sets are made available so that alternative approaches can be generated and compared for their effectiveness. Similarly, in the area of CFA the availability of such data sets would be helpful to facilitate further development of the discipline. Unfortunately, access to test data such as transaction data is limited to most academic researchers.

References

Ceglar, A. and J. Roddick, "Association Mining," *ACM Computing Surveys*, Volume 38, No. 2, July 2006, pp. 1-42.

Davis, T., "Cobra," (No date), http://www.almaden.ibm.com/asr/projects/cobra/

Desouza, K., Intelligent Agents for Competitive Intelligence: Survey of Applications, *Competitive Intelligence Review*, Volume 12, Issue 4, pp. 57-63, 2001.

IBM, (No date), Corporate Brand and Reputation Analysis for Consumer Products and Retail, http://www.ibm.com/smarterplanet/us/en/consumer_advocacy/nextsteps/solution/O960604Q71585B06.html

Kirkos, E., Spathis, C. and Manolopoulos, Y., "Data Mining Techniques for the Detection of Fraudulent Financial Statements," *Expert Systems with Applications*, Volume 32, 2007, pp. 995-1003.

Kreulen, J. T., "Reputation and Provenance Workshop Analytic Services," Almaden Services Research, IBM, 2008.

Kahaner, L., *Competitive Intelligence*, Touchstone, New York, NY, 1997.

Lee, W. and Stolfo, S., "Combining Knowledge Discovery and Knowledge Engineering to Build IDSs," www.raid-symposium.org/raid99/PAPERS/Lee_Stolfo.pdf, *1999*

Leary, R., Vandeburghe, W., Zeleznikow, J., "Toward a Financial Fraud Ontology: A Legal Modeling Approach," *ICAIL 2003 Workshop on Legal Ontologies & Web –Based Legal Information Management*, 2003 http://www.leibnizcenter.org/~winkels/LegOnt2003/Leary.pdf

Lerman, A. Individual Investor's Attention to Accounting Information: Message Board Discussions, unpublished paper, New York University (Jan. 2010).

O'Keefe, R.M., and O'Leary, D. E., "Performing and Managing Expert System Validation," Advances in Expert Systems for Management, Volume 1, pp. 141-176, 1993.

O'Leary, D.E., "Knowledge Discovery as a Threat to Database Security," in *Knowledge Discovery in Databases*, AAAI Press/MIT Press, 1991, pp. 507-516.

O'Leary, D.E., "Intrusion-Detection Systems," *Journal of Information Systems* (AAA), Volume 6, No. 1, Spring, 1992, pp. 63-74.

O'Leary, D.E., "Some Privacy Issues in Knowledge Discovery," *IEEE Expert*, Volume 10, Number 2, 1995, pp. 48-52.

O'Leary, D. E., "Monitoring and Managing Data and Process Quality using Data Mining: Business Intelligence for Purchasing and Accounts Payable Processes," in *Data Mining Methods and Applications*, K. Lawrence, R. Klimberg and R. Kudbya, Auerbach-Taylor & Francis, 2008.

O'Leary, D.E., "Using Digital Media to Monitor and Forecast a Firm's Public Image," *Advances in Business and Management Forecasting*, Volume 7, Emerald Publishing, 2010-a, p. 207-219.

O'Leary, D.E., "The Virtual Close and Continuous Monitoring at Cisco." *American Accounting Association*, San Francisco, August 2010-b.

O'Leary, D.E., "Blog Mining – Review and Extensions," *Decision Support Systems*, Volume 51, Number 4, pp. 821-830, 2011.

Phua, C., Lee, V., Smith, K. and Gayler, R., "A Comprehensive Survey of Data Mining-based Fraud Detection Research," http://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf

Quinlan, J.R., "Discovering Rules from large collections of examples: A Case Study," in D. Michie, ed., *Expert Systems in the Micro Electronic Age*, Edinburgh University Press, 1979.

Quinlan, J.R., "Induction of Decision Trees," Machine Learning, Volume 1, pp. 81-106.

Rosset, S., Murad, U., Neumann, E., Idan, Y. & Pinkas, G., (1999). "Discovery of Fraud Rules for Telecommunications - Challenges and Solutions." *Proc. of SIGKDD99*, 409-413.

Sabherwal, S., Sarkar,S., Zhang, Y., "Online Talk: Does it Matter," *Managerial Finance*, 34 (6) (2008) 423–436.

Searcey, D., "Employers watching workers on line spurs privacy debate," *Wall Street Journal*, April 23, 2009. http://online.wsj.com/article/SB124045009224646091.html

Shih, M., Liu, D. and Hsu, M., "Discovering Competitive Intelligence by Mining Changes in Patent Trends," *Expert Systems with Applications*, Volume 37, 2010, pp. 2882-2890.

Spangler, S., Chen, Y., Proctor, L., Lelecu, Ana, Behal, A., He, B., Griffin, T., Liu, A., Wade, B., and Davis, T., "COBRA – mining web for Corporate Brand and Reputation Analysis," *Web Intelligence and Agent Systems*, Volume 7, Number 3, 2009, pp. 243-254.

Tumarkin, R. "Internet Message Board Activity and Market Efficiency: A Case Study of the Internet Service Sector Using RagingBull.com," *Financial Markets, Institutions and Instruments* 11 (4) (2002) 313–335.

Van der Aalst, W. and Weijters, A., "Process Mining: A Research Agenda," *Computers in Industry*, Volume 54, pp. 231-244, 2004.

Vasarhelyi, M.A., M.A. Alles, and A. Kogan. 2004.Principles of Analytic Monitoring for Continuous Assurance, *Journal of Emerging Technologies in Accounting* 1: 1-21.

Xiaohui, C., Beaver, J. and Potok, T., "Swarm-based Knowledge Discovery for Intrusion Behavior Discovering," *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 2010.

Figure 1

COBRA Process

Sources: Kreulen (2008) and Davis (no date)

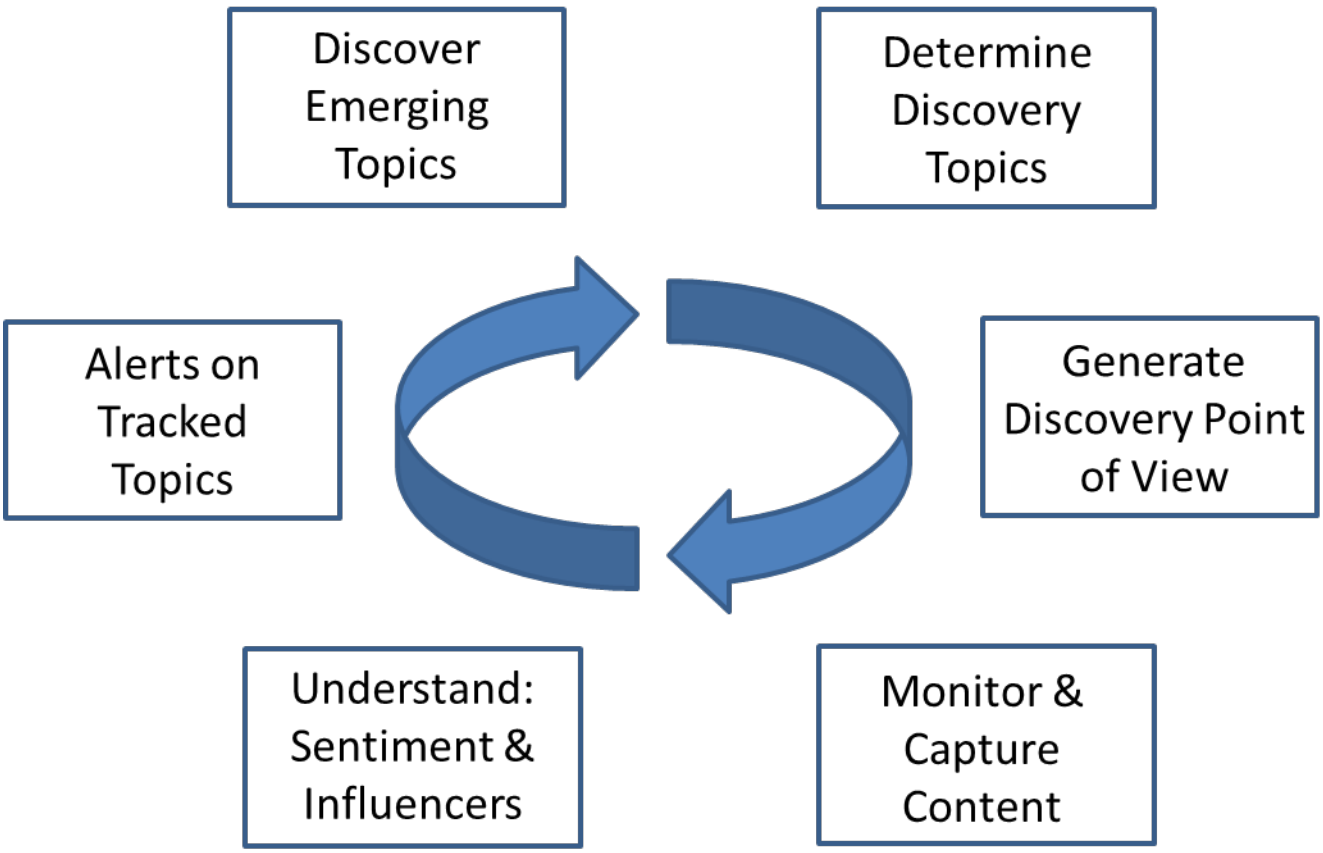

Figure 2

Example of COBRA (Kreulen 2008)

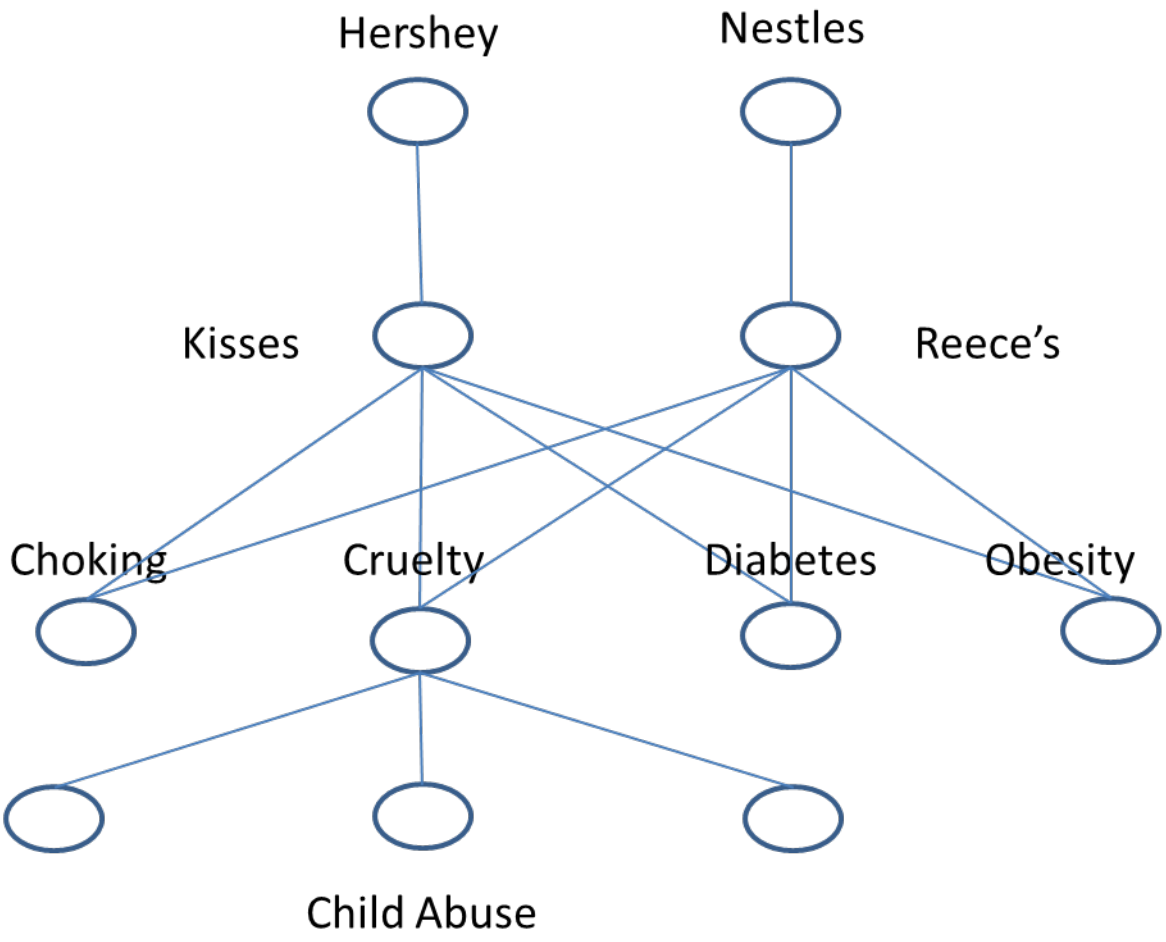

Table 1

"Continuous Monitoring" Web Occurrences

Topic	*# Google Pages*	*# Actual*
Continuous Monitoring of Digital Information	1	0
Continuous Monitoring of Blog	2	0
Continuous Monitoring of Blogs	8	8
Continuous Monitoring of Wiki(s)	0	0
Continuous Monitoring of Tags(s)	5	1
Continuous Monitoring of Twitter	8	8
Continuous Monitoring of Web 2.0 (Information)	0	0
Continuous Monitoring of email	13	13
Continuous Monitoring of Accounting	30	
Continuous Auditing and Monitoring	44,400	
Continuous Auditing	75,500	

July 4, 2011

Table 2

Occurrence of Some Phrases Indicating Potential Issues

Phrase	No. of Google Pages
I left work early	1,680,000
Company in trouble	513,000
Company in a downward spiral	219,000
Cash has disappeared	53,000
Touch of fraud	7,810
I broke into the system	22
Our accounting system sucks	4
Boss has just laid off	3

February 28, 2012