

Paper 144-2012

Becoming the Smartest Guys in the Room: An Analysis of the Enron Emails Using an Integration of Text Analytics and Case Management

John York, Doris Wong, Dan Zaratsian, SAS Institute, Cary, NC, USA

ABSTRACT

By integrating SAS® Enterprise Case Management with SAS® Text Analytics, fraud investigators gain an analytical advantage. SAS® Enterprise Case Management expedites fraud investigations by providing an organized environment for managing investigation workflows, documentation, and case notes. Using SAS® Text Analytics, you can index and categorize each document, further enhancing search capabilities, uncovering root causes, and saving time and money during time-sensitive investigations and audits.

The Case Management repository is valuable in itself, not only as an organizational repository, but also as a source of knowledge with alert capabilities that can be investigated with analytics in its own right. The integrated capability of utilizing SAS® Text Analytics with SAS® Enterprise Case Management is illustrated using the Enron email corpus.

INTRODUCTION

As an investigator, gaining an advantage over fraudsters is essential to effectively catching and preventing fraudulent activity. This type of activity takes place every day and it is your job to prevent it. You need access to superior fraud detection and prevention solutions. To fully understand how SAS® is superior in the fraud industry, let's perform a what-if analysis on a previous case of fraudulent activity: the downfall of Enron Corporation.

Enron Corporation was an American energy, commodities and services company based in Houston, Texas. In 2001, they filed for bankruptcy. Before their December 2, 2001 bankruptcy filing, Enron employed approximately 20,000 staff. They were one of the world's leading electricity, natural gas, communications and pulp and paper companies, with claimed revenues of nearly \$101 billion in 2000. Later, it was discovered that many of Enron's recorded assets and profits were inflated or even wholly fraudulent and nonexistent. For example, in 1999, Enron promised to pay back a Merrill Lynch investment with interest in order to show a profit on its books. Debts and losses were placed in offshore accounts that were not included in the firm's financial statements. More sophisticated and mysterious financial transactions between Enron and related companies were used to take unprofitable entities off the company's financial records. These "offshore" entities were limited partnerships between Enron and LJM Cayman L.P. and LJM2 Co-Investment L.P. created to buy Enron's poorly performing stocks to improve its financial statements. These two partnerships received funding of approximately \$390 million from a group of investors. Enron created entities it called "Raptors" and transferred more than \$1.2 billion in assets into Raptor accounts, including millions of shares of Enron common stock, long term rights to purchase millions more shares, plus \$150 million of Enron notes payable as disclosed in the company's financial statement footnotes. They capitalized the Raptors and booked the notes payable issued as assets on its balance sheet while increasing the shareholders' equity for the same amount.

During Enron's rise to the top, they were intertwined with multiple counts of fraudulent activity that could have been detected years before Enron's fall if investigators had the right tools. What if investigators used SAS® Enterprise Case Management integrated with SAS® Text Analytics? Could the fall of Enron have been different? In order to answer this question, you must first understand the hardships behind detecting and preventing fraud and how the integration between SAS® Enterprise Case Management and SAS® Text Analytics can eradicate these hardships.

Understanding the actions and motives of fraudsters is a vital part in predicting fraudulent activity. In order to gain an advantage, you must incorporate analytics in your investigative process. The goal of analytics is to turn data into useful information. Unfortunately, most data is unstructured, making it much harder to analyze effectively.

Unstructured data refers to information that either lacks a pre-defined data model and/or does not fit well into relational tables. SAS has overcome this challenge with SAS® Text Analytics. SAS® Text Analytics products process large volumes of text and provide useful analytics that can support any case under investigation. By integrating SAS® Text Analytics with SAS® Enterprise Case Management, you can run analytics on any unstructured data associated with your investigation.

TEXT ANALYSIS OF OVER 500,000 ENRON EMAILS

Investigations and audits are typically accompanied by large volumes of textual information, coming from multiple sources, and as the investigator, you are challenged with time-sensitive deadlines and an overwhelming amount of information. Perhaps one of the most time-consuming tasks is *finding* relevant information when you need it. SAS® Text Analytics uses natural language processing and pattern matching to facilitate timely fact extraction and data organization, enabling you to gain a holistic understanding of the information pertaining to your case.

Becoming the Smartest Guys in the Room: An Analysis of the Enron Emails Using an Integration of Text Analytics and Case Management, continued

SAS empowers intelligent search and advanced text analytics, supporting key functions such as:

- Multi-source information retrieval and integration.
- Advanced natural language processing to parse information.
- Term stemming and misspelling identification.
- Content categorization based on statistical and linguistic rules.
- Fact and entity extraction based on part-of-speech tagging and pattern recognition.
- Document filtering.

What does this give you? Powerful tools that aid in the organization, exploration and root cause analysis of time-sensitive case evidence.

This case study focuses on the analysis and information extraction of the Enron email archive. As the investigator, the goal is to extract key evidence in support of the investigation of Enron Corporation's suspicious accounting practices. The Enron email archive contains over 500,000 emails from 159 personal email accounts. The task of reading all these emails is unimaginable and impractical, not to mention how difficult it would be to identify patterns among key email accounts. SAS® Text Analytics leverages both statistical and linguistic technology to parse, explore and categorize the email collection.

As an investigator, you are not expected – nor do you have the time – to be an analyst or code sophisticated text analytics algorithms to gain insight. Instead, several behind-the-scenes processes are taking place while documents (in this case, the Enron emails) are imported, filtered and categorized before being sent to SAS® Enterprise Case Management. **Figure 1** represents the high-level process involved in collecting and investigating the Enron email archive.

STEP 1: You have the option to retrieve and aggregate information from disparate sources such as websites, RSS feeds, file directories, databases, and for this case study, the Enron email archive. Next, raw data typically requires pre-processing (i.e. removing formatting tags and converting into a plain text format) so that the unstructured text can be analyzed without the interference of special characters or formatting issues that may distort accurate results.

STEP 2: The first layer of filtering involves removing all records that are *known* to be irrelevant. This step is not always necessary and, in many cases, it is not apparent what records are irrelevant until the data exploration phase uncovers outliers and extraneous themes, or clusters, within the emails.

STEP 3: SAS® Text Miner provides the foundation for text analysis and exploration of unstructured data. In this step, natural language processing is used to break down the email contents into keywords and phrases along with their associated part-of-speech and frequency of occurrence. The software then automatically develops an optimal number of clusters, or themes, across the entire email collection.

STEP 4: Once a set of clusters have been generated, the software uses an additional layer of natural language processing to disambiguate terms and phrases based on their context within the email and their associated part-of-speech. Email categorization, entity extraction (ie. extraction of names, companies, locations, etc), and fact extraction are all supported in this step by leveraging a combination of customizable taxonomies, pattern matching and linguistic rules.

STEP 5: Before making the data available to SAS® Enterprise Case Management, a secondary filter eliminates noise and extracts only key information *relevant* to your case.

Becoming the Smartest Guys in the Room: An Analysis of the Enron Emails Using an Integration of Text Analytics and Case Management, continued

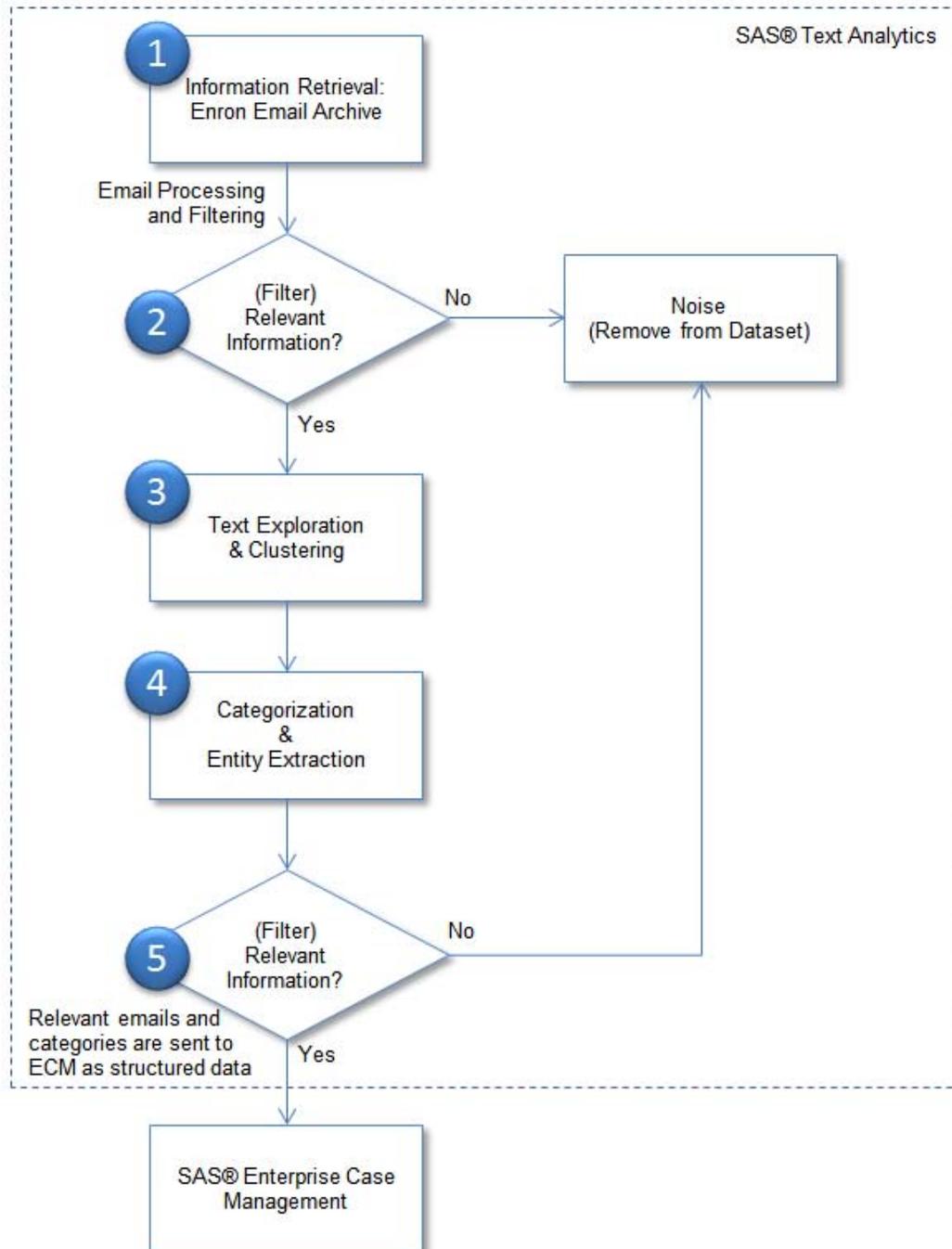


Figure 1: Text analytics process flow for the Enron email analysis

Integrating text analytics with Enterprise Case Management automates the evaluation and comprehension of textual data sources, whether the information is coming from private or public sources. This often results in better models because the text typically represents new information that enriches analytic insight.

This case study merges SAS® Text Analytics with SAS® Enterprise Case Management to deliver organized, relevant information when you need it while providing advanced analytics for clustering and analyzing multiple channels of unstructured data.

Becoming the Smartest Guys in the Room: An Analysis of the Enron Emails Using an Integration of Text Analytics and Case Management, continued

INVESTIGATING ENRON USING SAS® ENTERPRISE CASE MANAGEMENT

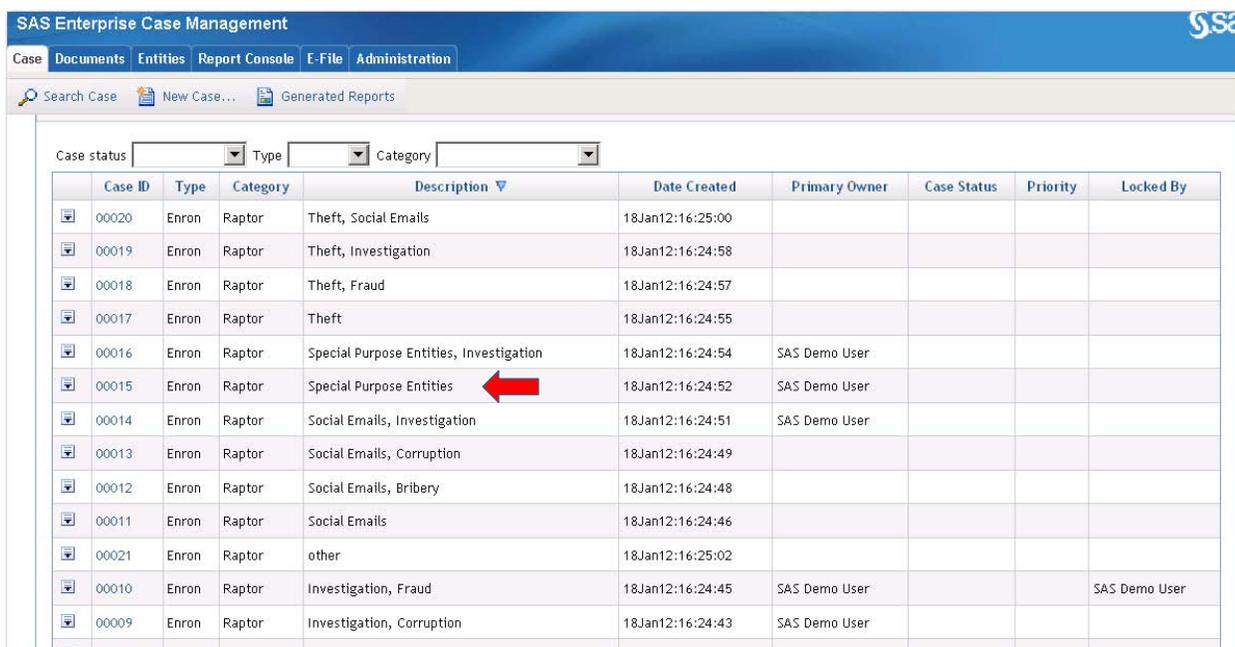
When fraudulent activity is suspected, you, as an investigator create a case. By using SAS® Enterprise Case Management, you could greatly reduce the cost of investigations. SAS® Enterprise Case Management provides a structured environment for managing investigation workflows, attaching comments or documentation and recording financial information such as exposures and losses. You can also adhere to compliance and regulatory requirements by generating and submitting suspicious activity reports electronically. In order to prevent re-keying errors, SAS® Enterprise Case Management pre-populates forms and automatically prepares batch reports. This solution lets you integrate information from siloed monitoring systems used by other lines of business, products or channels. You can import existing records and historical information and store it all within the case. As the textual information becomes organized, it provides structured information that can link to other data sources. For example, internally you can leverage the email addresses to identify involved employees, clients, or business partners, and then link it to the employee's employment record, trading history, expense reports, approved documents and more. Externally, you are able to use keywords mentioned in the email to crawl the internet for text definition or social media discussion. With the power of SAS data integration tools, all the data or documents in various formats can be loaded into SAS® Enterprise Case Management for further investigation. Access links to various source systems can also be added to avoid data duplication.

The benefits of using SAS® Enterprise Case Management over other management systems include:

- Streamlined investigation processes that reduce costs and more effectively prevent fraud.
- Comprehensive reports on everything from global statistics for high-level board reporting to extensive, detailed documentation necessary for supervisor review based on data gathered in the course of the investigation.

The Enron case is the perfect example of what can be accomplished by integrating SAS® Text Analytics and SAS® Enterprise Case Management. By integrating the two, Enron investigators would have had the tools to effectively detect and prevent fraud, create a holistic view of the entire case and prevent future monetary losses. SAS® Enterprise Case Management prevents future monetary losses by creating a holistic view of the customer and enabling communication among investigative departments. Enron's emails contained evidence of a large amount of long-term fraudulent activity. Seeing an all-inclusive, holistic view of their activity in one place may have led to a breakthrough in the investigation for the Enron case auditors and investigators. We were able to view all of the emails mentioning Enron's Raptor offshore entities. Using SAS® Enterprise Case Management, we created an organized view of related topics.

Special purpose entities are the off-shore accounts that Enron used to hide and funnel money so that their financial statements reported stronger earnings. These emails are categorized using linguistic rules to classify as *theft, fraud, special purpose entities, etc*, and can be dissected further to retrieve more information, as illustrated in **Figure 2** and **Figure 3**, below.



The screenshot shows the SAS Enterprise Case Management interface. At the top, there is a navigation bar with tabs for 'Case', 'Documents', 'Entities', 'Report Console', 'E-File', and 'Administration'. Below the navigation bar, there is a search bar and buttons for 'Search Case', 'New Case...', and 'Generated Reports'. The main area displays a table of cases with the following columns: Case ID, Type, Category, Description, Date Created, Primary Owner, Case Status, Priority, and Locked By. The table contains 13 rows of data. A red arrow points to the row with Case ID 00015, which has the description 'Special Purpose Entities'.

Case ID	Type	Category	Description	Date Created	Primary Owner	Case Status	Priority	Locked By
00020	Enron	Raptor	Theft, Social Emails	18Jan12:16:25:00				
00019	Enron	Raptor	Theft, Investigation	18Jan12:16:24:58				
00018	Enron	Raptor	Theft, Fraud	18Jan12:16:24:57				
00017	Enron	Raptor	Theft	18Jan12:16:24:55				
00016	Enron	Raptor	Special Purpose Entities, Investigation	18Jan12:16:24:54	SAS Demo User			
00015	Enron	Raptor	Special Purpose Entities	18Jan12:16:24:52	SAS Demo User			
00014	Enron	Raptor	Social Emails, Investigation	18Jan12:16:24:51	SAS Demo User			
00013	Enron	Raptor	Social Emails, Corruption	18Jan12:16:24:49				
00012	Enron	Raptor	Social Emails, Bribery	18Jan12:16:24:48				
00011	Enron	Raptor	Social Emails	18Jan12:16:24:46				
00021	Enron	Raptor	other	18Jan12:16:25:02				
00010	Enron	Raptor	Investigation, Fraud	18Jan12:16:24:45	SAS Demo User			SAS Demo User
00009	Enron	Raptor	Investigation, Corruption	18Jan12:16:24:43	SAS Demo User			

Becoming the Smartest Guys in the Room: An Analysis of the Enron Emails Using an Integration of Text Analytics and Case Management, continued

Figure 2: SAS® Enterprise Case Management case results and description

Document ID	Type	Category	Subject	Sender	Partial Body
00465	Enron	Email	Brigham	david.delainey@enron.com	Great job guys - keep it coming! Regards Delainey Forwarded by David W Dela
00466	Enron	Email	Catalytica	david.delainey@enron.com	Guys, I suggest that we write it down in order to ensure that we don't get it back as other groups
00469	Enron	Email	Re: Raptor	david.delainey@enron.com	That may work - I don't want to end up with an equity position I just worked hard to eliminate. Fu
00471	Enron	Email	Brigham	david.delainey@enron.com	Great job guys - keep it coming! Regards Delainey Forwarded by David W Dela
00472	Enron	Email	Re: Raptor	david.delainey@enron.com	That may work - I don't want to end up with an equity position I just worked hard to eliminate. Fu
00474	Enron	Email	Catalytica	david.delainey@enron.com	Guys, I suggest that we write it down in order to ensure that we don't get it back as other groups
00477	Enron	Email	Brigham	david.delainey@enron.com	Great job guys - keep it coming! Regards Delainey Forwarded by David W Dela
00480	Enron	Email	Catalytica	david.delainey@enron.com	Guys, I suggest that we write it down in order to ensure that we don't get it back as other groups
00482	Enron	Email	Re: Raptor	david.delainey@enron.com	That may work - I don't want to end up with an equity position I just worked hard to eliminate. Fu

Figure 3: SAS® Enterprise Case Management case information

SAS® Enterprise Case Management gives all types of users, on-demand access to meaningful information through comprehensive reporting capabilities and interactive, visual dashboards that provide critical information. By looking at the case network analysis, you can construct networks of individuals who may have been tracked in current and past investigations, thus providing richer information. You can observe individual suspects and determine how they are connected, again, creating a more holistic view of a case. Figure 4 demonstrates how the Enron emails are linked based upon subjects or entities.

Becoming the Smartest Guys in the Room: An Analysis of the Enron Emails Using an Integration of Text Analytics and Case Management, continued

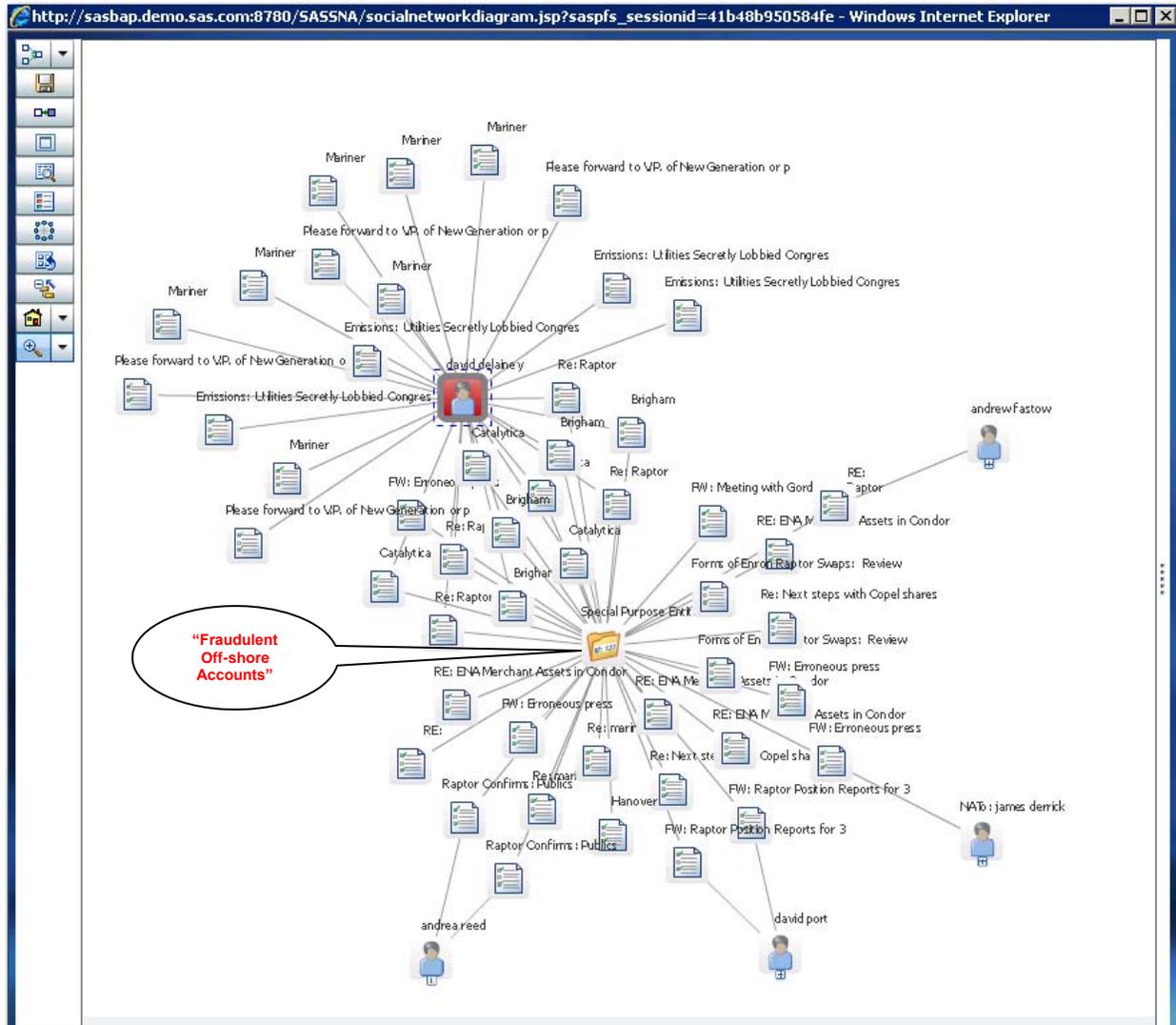


Figure 4: Case network analysis

The wide variety of interactive dashboards can display any information that you would like to visualize.

SAS® Enterprise Case Management eases the burden of audits by providing management, examiners or regulatory agencies with automatically generated audit records that include the identification of the user, a time stamp and the date when actions are performed on any case under investigation. You can also incorporate historical versioning as a structural way of adding more information into a case. Users can review changes made to selected fields at the subject level, incident level and case level. Through versioning, you can add more information to our Enron case such as a new suspect, attachments and additional emails and retain previous versions of the case within the case management system. This eliminates the need to check for differences between versions manually. **Figure 5** is an example of versioning within SAS® Enterprise Case Management.

Becoming the Smartest Guys in the Room: An Analysis of the Enron Emails Using an Integration of Text Analytics and Case Management, continued

Type	Description	Created By	Date Created
Save	Version: 6	SAS Demo User	24Feb12:16:29:07
Save	Version: 5	SAS Demo User	24Feb12:16:28:56
Save	Version: 4	SAS Demo User	24Feb12:16:27:58
Add Comment	More activities are identified	SAS Demo User	24Feb12:16:27:51
Save	Version: 3	SAS Demo User	24Feb12:16:27:00
Save	Version: 2	SAS Demo User	24Feb12:16:26:52
Lock	SAS Demo User	SAS Demo User	24Feb12:16:26:42
Unlock	SAS Demo User	SAS Demo User	24Feb12:16:06:35
Lock	SAS Demo User	SAS Demo User	24Feb12:16:06:20
Unlock	SAS Demo User	SAS Demo User	24Feb12:16:04:58
Add Comment	Off-shore Accounts	SAS Demo User	24Feb12:16:04:41
Lock	SAS Demo User	SAS Demo User	24Feb12:16:04:09
Unlock	SAS Demo User	SAS Demo User	24Feb12:16:02:52
Lock	SAS Demo User	SAS Demo User	24Feb12:16:02:20
Unlock	SAS Demo User	SAS Demo User	24Feb12:16:02:13
Lock	SAS Demo User	SAS Demo User	24Feb12:15:43:49
Unlock	SAS Demo User	SAS Demo User	24Feb12:15:37:16
Lock	SAS Demo User	SAS Demo User	24Feb12:15:28:08

Figure 5: SAS® Enterprise Case Management versioning

CONCLUSION

Using the Enron emails allows us to demonstrate new opportunities for investigators who need to understand large volumes of data as integrated evidence for their cases. Large volumes of information have gone from being a challenge to an advantage for investigators because they are able to use SAS® Text Analytics to understand and categorize these documents. SAS® Enterprise Case Management creates a more comprehensive view of cases. You can gain insight from large amounts of unstructured data integrated from multiple sources, while linking suspicious activity and document trails. With this capability, your auditors and investigators save time and money during the investigative process and prevent future losses.

So, what if the Enron investigators used SAS® Enterprise Case Management integrated with SAS® Text Analytics? How could the outcome have been different? If the Enron investigators used this technology to create a holistic view of their case, they would have saved substantial amounts of time and money.

Becoming the Smartest Guys in the Room: An Analysis of the Enron Emails Using an Integration of Text Analytics and Case Management, continued

REFERENCES

- "Unstructured Data." *Unstructured Data: Wikipedia, The Free Encyclopedia*. Wikipedia, 2 Feb. 2012. Web. 5 Feb. 2012. en.wikipedia.org/wiki/Unstructured_data.
- "SAS® Enterprise Case Management: Efficiently Manage Investigation Workflows." *Case Management, SAS Enterprise Case Management* |SAS. SAS Institute, 2 June 2011. Web. 3 Jan. 2012. sas.com/solutions/fraud/case-management/index.html#section=2.
- "Enron Scandal." *Enron Scandal - Wikipedia, The Free Encyclopedia*. Wikipedia, 27 Jan. 2012. Web. 29 Jan. 2012. en.wikipedia.org/wiki/Enron_scandal.
- "EDRM Enron Email Data Set V2." *EDRM Enron Email Data Set V2 << The Electronic Discovery Reference Model*. EDRM, 10 June 2010. Web. 17 Nov. 2011. edrm.net/resources/data-sets/edrm-enron-email-data-set-v2.

RECOMMENDED READING

- User's Guide for SAS® Content Categorization
- User's Guide for SAS® Enterprise Case Management 3.1

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

John York
SAS Institute Inc.
100 SAS Campus Drive
Cary, NC 27513
Work Phone: (919) 531-8863
E-mail: john.york@sas.com

Doris Wong
SAS Institute Inc.
Address: 100 SAS Campus Drive
City, State ZIP: Cary, NC 27513
Work Phone: (919) 531-1726
E-mail: Doris.Wong@sas.com

Dan Zaratsian
SAS Institute Inc.
100 SAS Campus Drive
Cary, NC 27513
Work Phone: (919) 531-8862
E-mail: Dan.Zaratsian@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.